

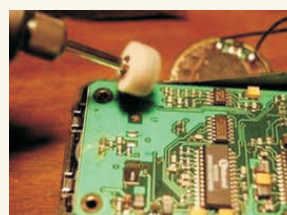
Applied Engineering and Technologies

CONTENTS

VOLUME - 2, ISSUE 1

APRIL - 2013

Topic	Pg. No.
• A Survey on Mixed Operating Mode/Self Synchronization	1
• A Novel Technique for Yawning Based Drowsiness Detection	4
• Operational Transconductance Amplifier Design in 180 nm Technology	7
• A Framework for Face Recognition from Video Sequence Using HMM	11
• Super Railway ticketing using ZIFBEE	16
• Normality and Stationarity analysis of signals using Kolmogorov-Smirnov Test	20
• Hand Gesture Recognition	25
• Overview of Impulse-based UWB transmitter receiver	30
• Study of AES designed on FPGA using Pipelined Partial Rolling to achieve high throughput	34
• Speed Control Mechanism Using Terrain Detection	38
• Route maintenance schemes based on Node life prediction for two reactive routing protocols in Mobile ad Hoc networks	44
• Dynamic Modelling of Arm	49
• Public key cryptographic communication system	54
• Transfer function modelling	61
• 3D Fingerprint Recognition Code	67
• Automation of Raw Mill Drive in Cement Industry using PLC and SCADA	73
• Robotic Arm controlled by potentiometric system	77
• Design and Implementation of PMBLDC Motor for Solar Powered Pumps	80



Published by

Vidyalankar Institute of Technology

Wadala, Mumbai, Maharashtra

www.vit.edu.in

VIT

Vidyalankar Institute Of Technology



Vidyalankar Institute of Technology (VIT) is an 'A' grade Engineering degree college approved by the AICTE, the DTE (Maharashtra), and affiliated to the University of Mumbai. Incepted in the year 1999, the Institute has steadily gained recognition and today attracts top talent.

VIT conducts courses in Electronics, Electronics and Telecommunication, Computers, Information Technology and Biomedical Engineering with an average annual intake of 700 students at the undergraduate level. The Institute also conducts Masters in Engineering programs in three disciplines in addition to Masters in Management Studies.

The Institute's mission is to be a leading academic centre that imparts quality education in the fields of Engineering, Technology and Management. The institute aims to facilitate a research environment that is conducive to evolving knowledge and generation and dissemination of new ideas. Information about VIT can be found on www.vit.edu.in.

EDITORIAL BOARD:

EDITOR IN CHIEF:

Prof. Seema Shah, VIT

CONSULTING EDITORS:

Prof. Dharmendra Shadija, Sheffield Hallam University, UK
Prof. Dr. Subhasis Chaudhuri, IIT Bombay
Prof. Dr. Vivek Agarwal, IIT Bombay
Prof. Dr. Sidhartha Dattagupta, IIT Bombay
Prof. Dr. R.N. Banawar, IIT Bombay
Dr. S.M. Dhume,, NITIE, Mumbai
Dr. Aditya Abhayankar, VIIT, Pune
Mr. Rahul Rane, L&T Automation and Control
Mr. Abhijeet Karkare, Alcatel Lucent
Mr. Nitin Nadkarni , Techmahindra
Dr. Vimal Someswar, K D Ambani Hospital
Mr. Milind Padalkar, HCL Technologies
Prof. Dr. S. A. Patekar, VIT, Mumbai
Dr. Ramprakash Dwivedi, Inha University
Prof. Dr. Uma Rao, SAKEC, Mumbai
Dr. A. J. Parameswaran, TIFR, Mumbai

EDITORIAL TEAM:

Dr. Anjali Deshpande, VIT
Prof. Geetha Naryanan
Prof. Girish Gidaye

Review Committee:

Dr. Varsha Patil, MCERC, Nashik
Dr. R.G. Karandikar, KJSCE, Mumbai
Dr. Deepak Bhoir, FRCRCE, Mumbai
Dr. Manoj Deshpande, ACPCOE, Navi Mumbai
Dr. Shanta Sondur, VSIT, Mumbai
Prof. B.R. Prabhu, KJSCE, Mumbai
Dr. M. D. Patil, RAIT, Mumbai
Dr. S. S. Sane, VJTI, Mumbai
Dr. K. T. V. Reddy, Principal, PVPPCOE, Mumbai

REVIEW COMMITTEE-INTERNAL:

Dr. Anjali Deshpande
Dr. Saurabh Mehta
Prof. Seema Shah
Prof. Varsha Bhosale
Dr. Gajanan Nagare
Dr. Ankush Sharma
Prof. Sachin Bojewar

PUBLICATION AND

WEBSITE DEVELOPMENT:

Prof. Meenakshi Nandula, VIT
Mr. Shashikant Poskar, Vidyalankar Group
Mr. Nilesh Parmar, Vidyalankar Group
Mr. Vivek Venkatram, VIT

Form the desk of Editor

Vidyalankar Institute of Technology's Annual Technical Journal '**Applied Engineering and Technologies**' is a forum for engineering professionals to showcase their research talents and gives them access to the use of contemporary innovations and the developments in various fields of Engineering. Eminent guest editors and the in-house editorial team comprising of experienced faculty ensure that good quality research papers are included in the journal. The rigorous screening process includes critical comments, suggestion and modification cycle. This initiative takes the Institute one more step ahead in its evolution into a research institute that also imparts education, thereby realising the ultimate dream envisaged by its founder Prof. C. S. Deshpande.

This second volume of **Applied Engineering and Technologies** focuses on applications in the areas of Electronics, Electrical, Communication, Computers and Bio-Medical Engineering. The target readers are research students, undergraduate and postgraduate students and practicing and upcoming professionals from industry and academia. The journal has been designed to spark discussions and offer myriad research opportunities to the readers. VIT has been regularly conducting annual National Conferences, ISTE approved Short Term Training Programs and Technical Workshops, which have always generated a buzz, and have witnessed enthusiastic participation from across academia and industry. This year too, an International Conference on "Recent Advances and Challenges in Engineering and Management" (RACEM-2013) was organized and hosted by Vidyalankar Institute of Technology.

"Research is to see what everybody else has seen, and to think what nobody else has thought", rightly asserts Albert Szent-Gyorgyi. Technical journal articles are one of the principal means whereby technologists and scientists share their work with the greater community. Institutes able to access, generate, and apply scientific knowledge have a competitive edge. Writing research papers is an essential aspect of academics for both faculty members and students and the process of writing a research paper in itself can be a very rewarding experience. It is an exercise through which the author learns how to research a topic, applies critical thinking skills and coherently expresses his understanding, analysis and conclusions for a specific audience.

Best wishes to **Applied Engineering and Technologies** and I hope that the articles published in it are important means of distributing research findings and disseminating new knowledge and fresh perspectives.

Seema Shah
Principal (Actg), VIT

A SURVEY ON MIXED OPERATING MODE/SELF SYNCHRONIZATION

Dipak S. Marathe

Assistant Professor in Electronics & Telecommunication Department.

A. C. Patil College of Engineering, Kharghar, Navi-Mumbai-410210.

deepak.acp@gmail.com

ABSTRACT

Mixed Operating Mode (MOM) is a digital hardware system which utilizes the advantages of Synchronous and Asynchronous Sequential circuit. But, an Asynchronous sequential circuit suffers from essential hazard. Earlier there were two design philosophies: by the addition of delay elements to the state output, or by input gate modification. These approaches make the whole design slower. On the other hand, the latter approach is based on the assumption that the gate delays are always higher than any wire delay present in the network, but in VLSI circuits the above assumption may not be true and also GALS (Globally Asynchronous, Locally Synchronous) FPGA model has a major drawback of implementing AI(Asynchronous Interfaces) in FPGA devices. AI design style which is based on asynchronous controllers that provides communication between modules(called ports) subject to essential hazard.

Using the concept of MOM results in not only saving the logic but also uses less power and making the whole design faster and more flexible. It also improves the throughput and reducing the latency, by satisfying the essential signal condition, this method is capable of providing robust ports, i.e. essential-hazard-free.

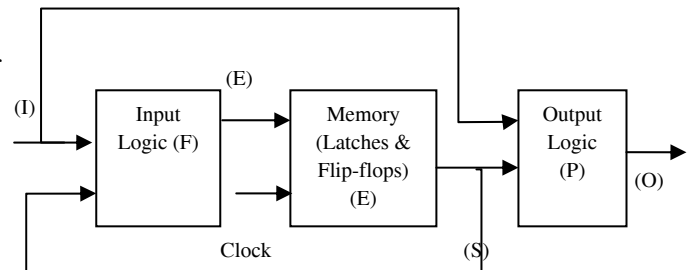
Keywords: VLSI, GALS, FPGA, AI.

I. INTRODUCTION

Synchronous sequential machines changes state in response to a synchronizing clock pulse shown in figure(1.1) and Asynchronous sequential machines do not use a clock shown in figure(1.2). Synchronous machines require that the internal time delays caused by the excitation logic, flip-flops and output logic will transpired prior to the arrival of the next clock pulse. In fact the synchronous machine logic designer must consider the time delays involved to determine the maximum operating speed of the clock. Asynchronous sequential circuits have no clock; the internal or secondary states changes according to the delay times of logic. Because of lack of synchronizing clock asynchronous sequential circuits are more difficult to design and also having problem of critical races, cycles and essential hazard [1].

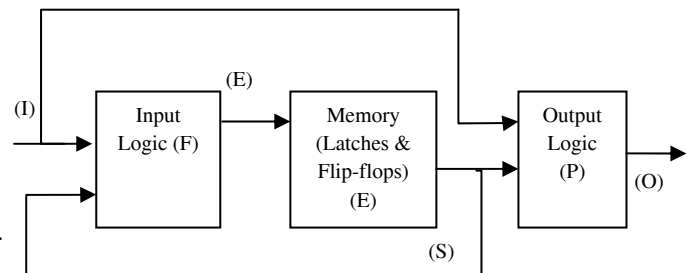
A. Races

Races exist in asynchronous machines when two or more state variables change during a state transition.



State Variable Feedback

Figure (1.1) Block diagram of Synchronous machine model.



State Variable Feedback

Figure (1.2) Block diagram of Asynchronous machine model

B. Noncritical race

A race said to be noncritical when the correct stable next state is eventually reached during a state transition. Transition path depends on whether F2 or F1 changes first. If F2 changes first the race has a transition path

(F2, F1, X, Y) 0000 → 0001 → 1001 → 1101

If F1 changes first the transition path

(F2, F1, X, Y) 0000 → 0001 → 1001 → 1101

Both transition paths lead to total stable state 1101, for this reason race is noncritical [1].

C. Critical race

A critical race occurs in asynchronous circuits when two or more internal state variables change at same time. Assume the machine is in total stable state 0000 and the inputs XY changes from 00 to 11. Three possible results can occur depending on which state variable changes first. In case 1(F2 and F1 change at the same time)

F2 F1 X Y	F2 F1 X Y	F2 F1 X Y
0 0 0 0	→ 0 0 0 1	→ 1 1 0 1

In case 2(F2 changes before F1)

F2 F1 X Y	F2 F1 X Y	F2 F1 X Y
0 0 0 0	→ 0 0 0 1	→ 1 0 0 1

In case 1(F1 changes before F2)

F2 F1 X Y	F2 F1 X Y	F2 F1 X Y
0 0 0 0	→ 0 0 0 1	→ 0 1 0 1

The final total stable state is dependent on which state variable changes first or if the both change at the same time. The result is not predictable and causes a critical race. Critical races are to be avoided [1].

D. Cycle

A cycle occurs when an asynchronous machine makes a transition through a series of unstable states.

E. Essential hazard

An essential hazard is caused by unequal propagation delay paths that originate from the same input. Essential hazards are eliminated only matching the propagation delay paths of a signal. The elimination of essential hazards is an application dependent problem [2]. The essential hazards are found by evaluating each stable total state. The initial state is listed followed by the input value listed in parenthesis. The arrows indicate the state transitions. If the new state is the same after one transition as it is after three transitions, no essential hazard exists; if it different, an essential hazard exists:

1. C (00) → F (01) → F (00) → F (01)

No essential hazard exists.

2. C (00) → B (10) → D (00) → F (10)

Essential hazard exist.

II. MOTIVATION

Significant problems arise in asynchronous circuit design due to critical races and hazards. Critical races and hazards results from unequal propagation delay paths in the combinational logic that realizes the state variables excitation equations. Static and dynamic hazards can be eliminated by providing holding terms in the excitation logic. Such holding terms increases the logic complexity.

A technique called mixed operating mode (MOM) or self synchronization has been proposed as a solution to asynchronous by a timing problems [2]. This approach partitions each state variable so that it has both synchronous and asynchronous inputs. Figure (2.1) shows a typical mixed mode flip-flop.

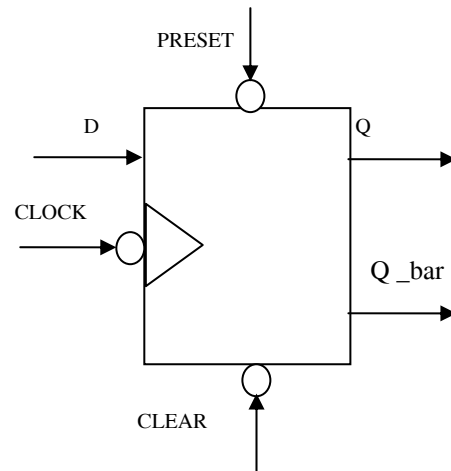


Figure (2.1) Mixed operating mode state variable

III. OBJECTIVE AND SCOPE

Earlier approaches to essential hazard-free design of asynchronous sequential circuits are based on any one of the two design philosophies: by the addition of delay elements to the state output, or by input gate modification. This approach makes the whole design slower. On the other hand, the latter approach is based on the assumption that the gate delays are always higher than any wire delay present in the network, but in VLSI circuits the above assumption may not be true. Mixed Sequential Machine is useful for high speed industrial application using VLSI technique.

IV. DESCRIPTION OF RESEARCH WORK

By using both synchronous and asynchronous inputs of the same state variable, a particularly difficult problem can be solved. If an essential hazards exists for a given state transition, then the synchronous inputs are used. If an essential hazard does not exist, then asynchronous inputs are used. This implies that a clock must be generated whenever essential hazards occur between state transitions. Extra logic is used to sense the existence of an essential hazard. If one exists then a clock pulse is generated and distributed to the specific state variable(s) that needs it. By providing such an internal clock pulse to only those state variables where essential hazards may occur eliminates the critical races without slowing down all the state transitions[3],[4].

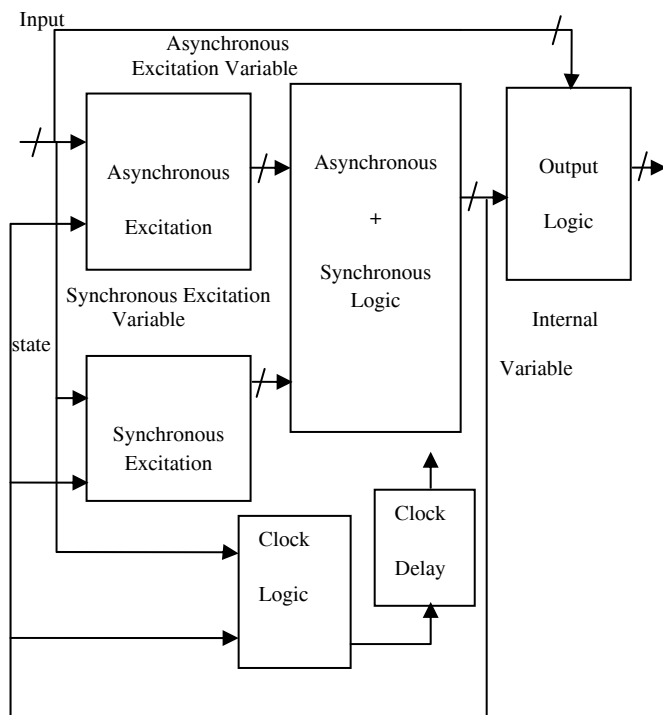


Figure (4.1) Mixed operating mode sequential circuit model.

V. PROPOSED IMPLEMENTATION TOOLS

Proposed system consisting of synchronous and asynchronous modules which can be synthesized and verified using different Frontend VLSI CAD tools like Xilinx, Altera, Synopsis, Petrify tool etc and the proposed implementations is in FPGA.

CONCLUSION

Hazard free design methodology results in saving logic, less power consumption, faster and more flexible design. The Minimum number of self-synchronous transitions making the design faster and Eliminating critical races [8], Concurrent Error Detection (CED) [9].

Concept of selective triggering or controlled excitation results in saving logic and more flexible design [10]. One of the IEEE research paper [11] suggests that Hybrid design not only improves the throughput but also reduces the latency. GALS (Globally Asynchronous, Locally Synchronous) FPGA model has a major drawback is implementing AI(Asynchronous Interfaces) in FPGA devices. AI design style which is based on asynchronous controllers that provides communication between modules(called ports) subject to essential hazard [12].

REFERENCES

- [1] Digital Logic Application and Design by John M.Yarbrough.
- [2] "Synthesis of Asynchronous Machines using Mixed-Operation Mode (MOM) "by O.Yenersoy IEEE Transactions on computers, Volume, C-28, no.4, April, 1979.
- [3] "Theory and Design of Mixed-Mode Sequential Machines "by Bulent I.Dervisoglu & Howard A. Sholl Member of IEEE.
- [4] "Hazard-free Design of Mixed Operating Mode Asynchronous Sequential Circuits" by Jen-Shiun Chiang; Damu Radhakrishnanb international journal of electronics vol.68, 1990.
- [6] "Synthesis of Self-Clocked Asynchronous State Machines" by Faxhad Aghdasi IEEE Region 10 Conference. Tencon 92, 11th - 13th November, 1992 Melbourne, Awealia.
- [7] "Self clocked asynchronous state machine design with PAL221P6" by Farhad Aghdasi and Martin Bolton, Feb-2003.
- [8] "Generation of a Clock Pulse for Asynchronous Sequential Machines to Eliminate Critical Races "by Jon G. Bredeson, Paul T. Hulina- IEEE transaction Aug-2006.
- [9] "Concurrent Error Detection Methods for Asynchronous Burst Machines" IEEE transaction on June 2007.
- [10] Multiple-input change Asynchronous machines using controlled excitation and flip-flops by Chung, Henry Y.H, Das, and Santana -IEEE Current Version July 2008.
- [11] An adaptively pipelined mixed synchronous-asynchronous digital FIR filter chip operating at 1.3 GHz-IEEE transactions on July-2010.
- [12] FPGA Implementation of Robust Asynchronous Wrappers for Globally Asynchronous Systems (GALS) by Durate L.Olivera,Lester A.Faria IEEE-2012.

A NOVEL TECHNIQUE FOR YAWNING BASED DROWSINESS DETECTION

Monali V. Rajput

Lecturer, Information Technology Department
Vidyalankar Institute of Technology,
Mumbai, India
monali.rajput@gmail.com

Mandar Sohani

Assistant Professor, Computer Department
Vidyalankar Institute of Technology,
Mumbai, India
mandar.sohani@vit.edu.in

Prof. (Dr.) J. W. Bakal

Principal, Shivajirao Jondhale College of Engineering,
Mumbai, India
bakaljw@gmail.com

ABSTRACT

Major cause of road accidents is tiredness and lethargy of driver. The main aim of the project is to find out the methods to detect driver drowsiness and alerting them, hence increasing the transportation safety. By using many body and face gesture as a sign of driver fatigue detection including yawning, eye tiredness, eye movement, these condition shows that driver is not in proper driving condition. In this novel technique yawning and eye movement are used. Driver fatigue can increase the chances of car accident, reason for this type of car accidents is due to fact that driver fails to take necessary actions prior to collision occurs. Therefore assisting system which will monitor the behavior of driver and also will give the necessary alerts to the driver is used to prevent the road accidents [1].

Keywords-Yawning,Eyemap,Mouthmap

I. INTRODUCTION

The advance of computing technology has provided the means for building intelligent vehicle systems. Drowsy driver detection system is one of the potential applications of intelligent vehicle systems. Excessive sleepiness results in an increased risk of motor vehicle crashes primarily because the driver either falls asleep while driving or experiences reduced attention to road events and the driving task due to fatigue. Statistics show that 20% of all the traffic accidents are due to drivers with a diminished vigilance level [1]. Furthermore, accidents related to driver hypo-vigilance are more serious than other types of accidents, since sleepy drivers often do not take correct action prior to a collision. For this reason, developing systems for monitoring driver's level of vigilance and alerting the driver, when he is drowsy and not paying adequate attention to the road, is essential to prevent accidents. The prevention of such accidents is a major focus of effort in the field of active safety research.

Driver drowsiness detection is a car or any vehicle safety technology which prevents accidents when the driver is

getting drowsy. 'Drowsy driving', sleep-related vehicle accidents are a common type of road crash. The proposed system is aimed to reduce the road accidents. In this system driver face is recorded by the camera which will be installed near the front mirror. In this system yawning and eye movements are considered. First the yawn is detected, to do so system will detect and track the face using series of frame shots taken by the camera. Then we can detect the location of eyes and mouth in the face. The need of eye movement detection is to verify the location of segmented mouth. By adding this constraint the system will be more robust to false detection. The mouth geometrical features are then used to detect the yawn. Once the system has detected the yawn it will alert driver of his fatigue and improper driving situation [1].

II. LITERATURE REVIEW

Many studies and researches showed that the automobile drivers involved in accidents are highly related to driver's fatigues such as non-stop late night driving behaviors. According to the statistics, more than 1.2 million people die each year on the world's roads, and between 20 and 50 million suffer non-fatal injuries due to road accidents [1]. Driver Hypo vigilance is one of the main reasons for road accidents in the world. The National Highway Traffic Safety Administration (NHTSA) conservatively estimates that 100000 police reports on vehicle crashes are the direct results of driver drowsiness each year. This results in approximately 1550 deaths, 71000 injuries and \$12.5 billion in monetary losses [2]. Automation offers daily life improvement through the integration of various systems into a united, familiar, insightful and non-threatening solution. Due to the urge to live a more comfortable life, it is expected that more and more people will prefer to choose the Smart Car solution for the range of benefits offered.

Considering all these situation drowsiness detection is helpful approach to solve the problem at initial state Drivers yawn can be detected to recognize the driver's state. The work proposed in [3] detects the face image, and then the segmented face region is used to locate the

chin, after that midpoint of nostrils is located. Thus drivers yawn can be calculated by the vertical distance between them. Again [4] makes the use of approach called Gravity center template. In this driver left and right mouth corners are detected by means of grey projection while doing this it also makes use of Gabor wavelet. Finally it applies LDA to classify features vector to detect the yawning state.

Another paper presents the system, where initially face is detected through viola Jones face detection method to make sure that driver is in video frame. Then mouth region extracted from face region, in which lips are captured through spatial fuzzy C-means(S-FCM) clustering. Simultaneously pupils are also detected. Then the eye and mouth information are passed to support vector machine which will analyze the true state of driver [5]. The system in [6] take the benefit of two cameras. One camera is fixed which provides driver head position and another camera which is controllable is used to detect the driver's mouth and extract mouth information. Then they make the use of haar like features to detect driver's mouth. In second camera lastly the yawning is detected by the ration of mouth height and width.

Another existing approach makes the use of Kalman filter to track the face region then the mouth window is located within face region and degree of mouth openness is extracted based on mouth features to determine driver yawning in video .In this system if false detection is happen then the system is reinitialized [7].

III.OBJECTIVE

Fatigue is often the cause of serious automobile crashes, failing concentration can spoil steering behavior and reduce driver alertness. Monotonous driving is particularly exhausting and can lead to loss of concentration. Driver fatigue increases the possibility of car accidents. National Highway Traffic Safety Administration (NHTSA) analysis data indicates that driving while drowsy is a contributing factor to 22 to 24 percent of car crashes. This high accidents rate is due to the fact that sleepy drivers fail to take correct action prior to a collision. An important derision in driver's fatigue is that the driver may be too tired to realize his own level of drowsiness.

Hence the objective of proposed system is to develop a monitoring system which can detect lack of attention and alert the driver of his drowsiness which will ultimately prevent accidents and save lives. The proposed system will then generate the alert in case of drowsiness is detected. Proposed system focuses on vehicle-based operator alertness monitoring technologies. There are several ways by which fatigue can be detected number of criteria available like eye detection, iris detection, yawning detection in proposed system yawning detection is the criteria used to check the drowsiness of the driver.

IV.METHODOLOGY

The driver drowsiness behavior detection using yawning feature system consists of different module to properly analyze changes in the mouth of driver. These modules are categorized as,

- Face Detection
- Eye Movement Detection
- Mouth Movement Detection
- Yawning Detection

In this proposed system database is created where the normal faces are stored in one database and its respective yawn face is stored in other database. Then to detect the yawn, the first step is to detect the face using template matching algorithm. After detecting the face next step is to detect the location of eyes and mouth. For this purpose eye map and mouth map is calculated using standard formula. The location of eyes in the face is used as an approach to verify the location of the segmented mouth. The mouth geometrical features are then used to detect the yawn. The proposed method is explained in figure 1.

V. DESCRIPTION

A] Face Detection

There are different methods to detect the face like Knowledge-based, Feature invariant approach, Template matching, Appearance-based method. From these methods here we make use of template matching method [8]. The basic idea of template matching is to store several standard patterns which

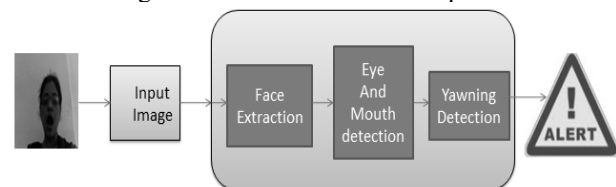


Figure 1.Block Diagram of Proposed System

will describe the face as a whole or the facial features separately. These templates are hand-coded which uses correlation technique to locate the face.

It is important to note that the detection scheme should be invariant to skin type and changes in lighting condition therefore will make use of set of bounding rules for different colour space which will help to improve the detection efficiency. RGB colour space is used to detect skin colour at uniform or lateral daylight illumination and under flashlight illumination [1].

$$\begin{aligned}
 &(R>95) \text{ AND } (G>40) \text{ AND } (B>20) \text{ AND} \\
 &(\max\{R, G, B\} \\
 &- \min\{R, G, B\}>15) \text{ AND } (|R - |>15) \text{ AND} \\
 &(R>G) \text{ AND} \\
 &(R>B) \text{ AND } (R>220) \text{ AND } (B > 170) \text{ AND} \\
 &(R>B) \text{ AND } (G>B) \quad (1)
 \end{aligned}$$

YCbCr Colour space is a strong determination of skin Colour and rules apply to this Colour space is, $(Cr \leq 1.5862 * Cb + 20) \text{ AND } (Cr \geq 0.3448 *$

$$\begin{aligned}
 &Cb + 76.2069) \\
 &AND (Cr \geq -4.5652 * Cb + 234.5652) AND \\
 &(Cr \leq -1.15 * \\
 &Cb + 301.75) AND \\
 &(Cr \leq -2.2857 * Cb + 432.85) \quad (2)
 \end{aligned}$$

Another Colour space is HSV space. Hue values exhibit the most noticeable separation between skin and non-skin regions.

$$H < 25 \text{ and } H > 230 \quad (3)$$

The result of the skin location technique is a black and white image which highlights the skin location by converting the face to white and the background and the other areas around the driver to black. Finally, the template matching method is used to detect the more accurate faces, and neglect the non-faces [1].

B] Eye Movement Detection

Among the various facial features, eyes, mouth are the most prominent features for recognition. The approach for eye movement detection is based on eye map. Eye map means locating eyes based on feature maps derived from the luminance and the chrominance of an image [9]. For detecting eyes, the eye maps based on chrominance component are created. After detecting the eye map image we will then convert this image in to black and white image using proper thresholding. Image after thresholding will include eyes in white while the rest is all black. Using the equation below eye maps based on chrominance component are built this is used further to detect the eyes [1].

$$\text{Eye Map} = \{(C_b)^2 + (C_r)^2 + (C_b / C_r)\} \quad (4)$$

C] Mouth Movement Detection

Going ahead after detecting eye next step is to detect the mouth region and the lips. Mouth area is segmented first here the strong difference between lips color and face color is used. The color of mouth region contains redder component compared to the blue component than other facial regions. Usually mouth area is detected based on color information after detecting the face [9]. After detecting the mouth map again some post processing steps such as black and white conversion, finding the biggest connected component will be done [1].

$$\text{Mouth Map} = C_r^2 \times \{C_r^2 - \eta \cdot (C_r / C_b)\}^2 \quad (5)$$

Where η is,

$$\eta = 0.95 \cdot \frac{\frac{1}{n} \sum_{(x,y) \in FG} C_r(x,y)^2}{\frac{1}{n} \sum_{(x,y) \in FG} C_r(x,y) / C_b(x,y)}$$

D] Yawning Detection

A yawn is a reflex of simultaneous inhalation of air and stretching of the eardrums, followed by exhalation of breath. Yawning is commonly associated with tiredness, stress, overwork, lack of

stimulation and boredom. There are two steps involved while detecting the yawn state. First of all yawn component in face independent of mouth location is detected. A criterion to detect the same is to find the hole in mouth as a result of wide mouth opening. In second state mouth location is used to verify the correctness of detected component. The large hole is selected as a candidate for yawning mouth and this hole is the non-skin area inside the face. The hole can be related to eyes, mouth or open mouth. It is assumed that the open mouth will be the largest among three which can be a yawning state [1]. In this way the detected yawn state is compared with the available face of yawning state to check the drowsiness of a driver.

CONCLUSION

The proposed novel technique is different from the existing system in terms of its robustness as it applies several verification criterions [1]. The algorithm used here minimizes the complexity, increases accuracy and reduces the detection dependency on the face geometry.

REFERENCES

- [1] Shabnam Abtahi, Behnoosh, Driver Drowsiness Monitoring Based on Yawning Detection, Distributed Collaborative Virtual Environment Research Laboratory, University of Ottawa, Canada
- [2] NHTSA, "Drowsy driver detection and warning system for commercial vehicle drivers: Field proportional test design, analysis, and progress", National Highway Traffic Safety Administration, Washington, DC, [Online], Available: <http://www.nhtsa.dot.gov/>.
- [3] S.G. Klauer, T.A. Dingus, Neale, V.L., Sudweeks, J.D., and Ramsey, DJ, "The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data," Virginia Tech Transportation Institute Report #DOT HS 810594
- [4] X. Fan, B. Yin, Y. Fan. "Yawning Detection for Monitoring Driver Fatigue." In: Sixth International Conf. on Machine Learning and Cybernetics, Hong Kong, 2007, pp. 664-668
- [5] T. Azim, M.A. Jaffer, A.M. Mirza. Automatic Fatigue Detection of Drivers through Pupil Detection and Yawning Analysis. In: Proc. Fourth International Conf. on Innovative Computing, Information and Control, 2009, pp. 441-445
- [6] L. Li, Y. Chen, Z. Li. "Yawning Detection for Monitoring Driver Fatigue based on Two Camera. In: Proc. 12th International IEEE conf. on Intelligent Transportation Systems, St. Louis, MO, USA, 2009, pp. 12-17
- [7] T. Wang, P. Shi. "Yawning Detection for Determining Driver Drowsiness." IEEE International Workshop VLBI Design & Video Tech. Suzhou, China, 2005, pp., 373-376.
- [8] Inseong Kim, Joon Hyung Shim, and Jinkyu Yang "Face detection"
- [9] Hsu Rein-Lien, M. Abdel-Mottaleb, and A. K. Jain. "Face Detection in color Images." IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, issue 5. 2002

OPERATIONAL TRANSCONDUCTANCE AMPLIFIER DESIGN IN 180 NM TECHNOLOGY

Sushil Kumar

School of Information and Communication Technology
Gautam Buddha University, Uttar Pradesh-201 310, India
sushilkumar0108@gmail.com

ABSTRACT

The paper presents the design concept of Operational Transconductance Amplifier (OTA). The Operational Transconductance Amplifier (OTA) is designed and simulated in 180 nm CMOS process technology with supply and bias voltage of 1.8 V. Cadence Spectre environment UMC 180 nm technology file is used for the design and simulation of the Operational Transconductance Amplifier (OTA). Open loop gain of about 72 dB which is having Gain Bandwidth (GBW) of 39 KHz. The Operational Transconductance Amplifier (OTA) shows power dissipation of 10 mW and Slew Rate 2.33 V/ μ sec. The Operational Transconductance Amplifier (OTA) shows Common Mode Rejection Ratio (CMRR) of 91 dB and Power Supply Rejection Ratio (PSRR) of 84 dB.

Keywords- OTA, CMOS, Cadence, Open loop gain, GBW, Slew Rate, CMRR, PSRR.

I. INTRODUCTION

Now a days, with the shrinking technology, more and more circuits are integrated on a single chip. This emerging technology reduces the overall size and increases the chip density. As the technology is scaling down, the transistor size is reduced which results that more transistors can be fitted on the same die. The Operational Transconductance Amplifier (OTA) is a basic building block in most of analog circuit with linear input output characteristics. The main difference between Operational Transconductance Amplifier (OTA) and conventional operational Amplifier is that in conventional Operational Amplifier, the output is in the form of voltage and in Operational Transconductance Amplifier (OTA), the output is in the form of current [1]. The Operational Transconductance Amplifier (OTA) is very much similar to conventional Operational Amplifiers in which both do have differential inputs. The Operational Transconductance Amplifier (OTA) is an amplifier takes voltage as differential input and produces current as an output. Therefore, it is also known as a voltage controlled current source (VCCS). It has usually an additional input for a current to control

the amplifier's transconductance. Operational Transconductance Amplifier (OTA) is very much similar to a standard operational amplifier because it has a high impedance differential input stage and that it can be used with negative feedback. The Operational Transconductance Amplifier (OTA) is not as useful as the ordinary operational amplifier because its output is a current [2]. It has number of important electronic applications such as the variable frequency oscillator, variable gain amplifier and filters stages which are more difficult to implement with standard operational amplifier. It uses open loop usually with negative feedback in linear applications due to the magnitude of the resistance attached to its output controls its output voltage. Thus, a choice of resistance can be made that can keep the output going into saturation even in case of high differential input voltages. The amplifier's transconductance is directly proportional to its current which makes it useful for electronic control of amplifier gain and all [3].

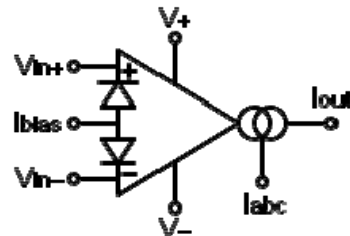


Figure 1: Schematic symbol of OTA

The Operational Transconductance Amplifier (OTA) like the standard operational amplifier, has both non inverting (+) and inverting (-) inputs; power supply lines ($V+$ and $V-$); and a single output. It has two additional biasing inputs, I_{abc} and I_{bias} unlike the traditional operational amplifier [4].

II. DESIGN OF OPERATIONAL TRANSCONDUCTANCE AMPLIFIER (OTA)

Schematic diagram of Operational Transconductance Amplifier (OTA) is shown in figure 1. In this schematic diagram of Operational Transconductance Amplifier

(OTA), supply voltage (VDD) is 1.8V. Also, the transistors M11_1 and M11 act as a constant current source and transistors M1, M2 and M3 work as two current mirror pairs. Transistors M4, M5, M6 and M7 are the Differential amplifier and transistor M8 is an output amplifier stage. Also, the design parameters of the Operational Transconductance Amplifier (OTA) are shown in table 1 given below. A Operational amplifier is normally characterized by various crucial parameters like bandwidth, noise, open loop gain, slew rate etc. The design parameters such as transistor size, bias voltage determine the performance measures [5].

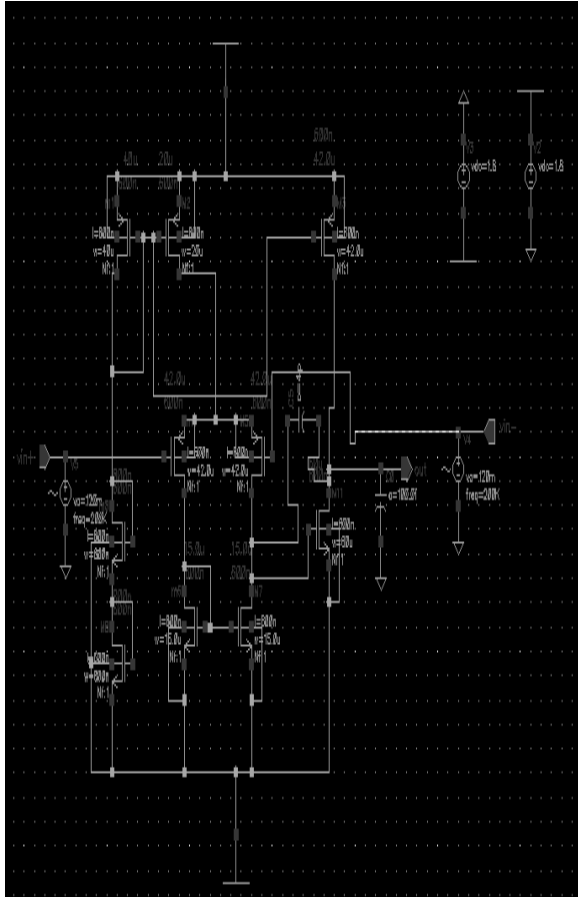


Figure 2: The schematic diagram of Operational Transconductance Amplifier (OTA)

TABLE 1: DESIGN PARAMETERS

Transistor	W/L Ratio (nm)
M1, M2, M3	400/6
M4, M5	200/6
M6, M7, M8, M9	420/6
M8, M9	500/6
M10, M11	600/6
M12, M13	8/6

III.SIMULATION RESULTS AND ANALYSIS

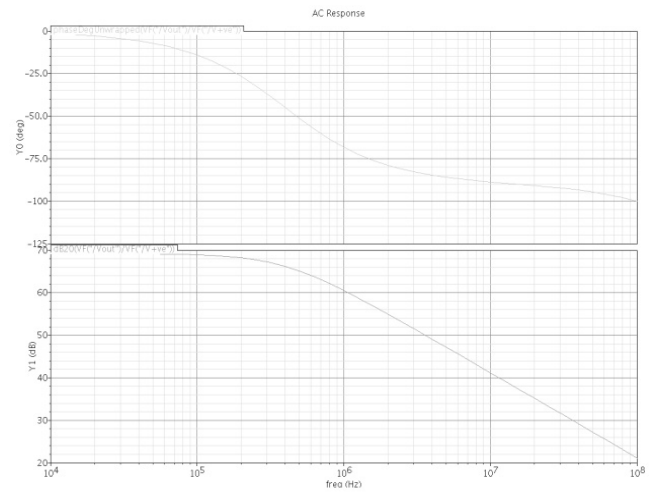


Figure 3: The AC response showing gain and phase change with frequency

The figure 3 shown above, gives the idea how the curves of gain and phase change behave w.r.t. frequency. It is clearly visible in the figure that the numerical value of change in phase is more than the gain at a particular frequency that is 10^6 Hz.

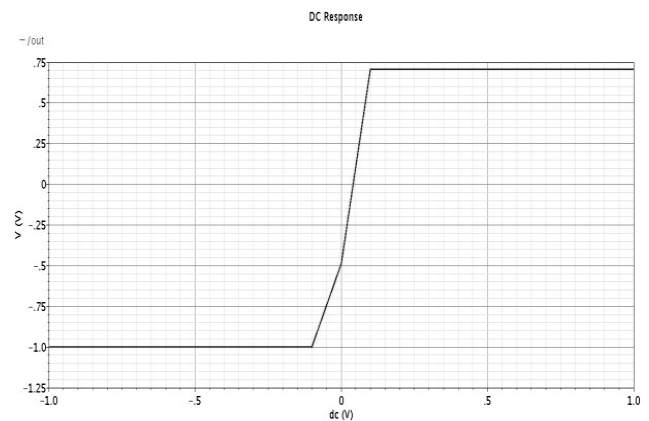


Figure 4: The DC sweep response

In the DC sweep response, the DC voltage is swept as the independent axis [6]. At point 0 V DC voltage, the corresponding value of the voltage on Y-axis comes out to be -0.5 V.

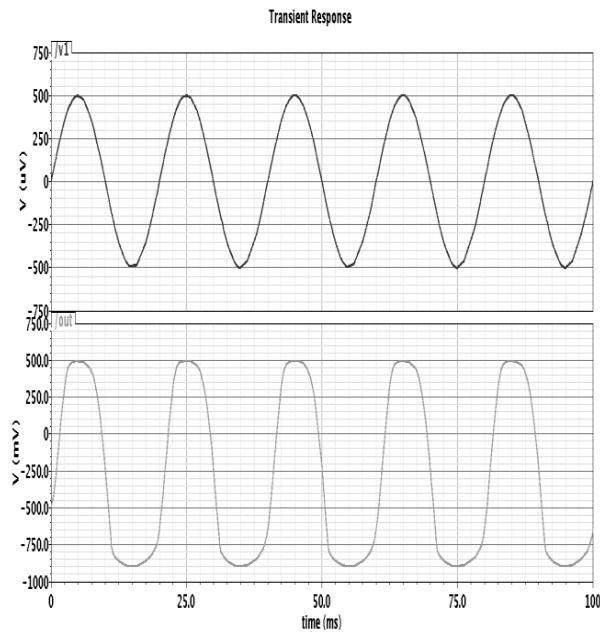


Figure 5: The Transient response with input as pulse

In the figure given above, the Transient analysis is done to voltages or currents as a function of time on the independent axis i.e; x-axis. The crest of the output voltage is more flattened as compared to the input voltage, signifies the stability of the OTA.

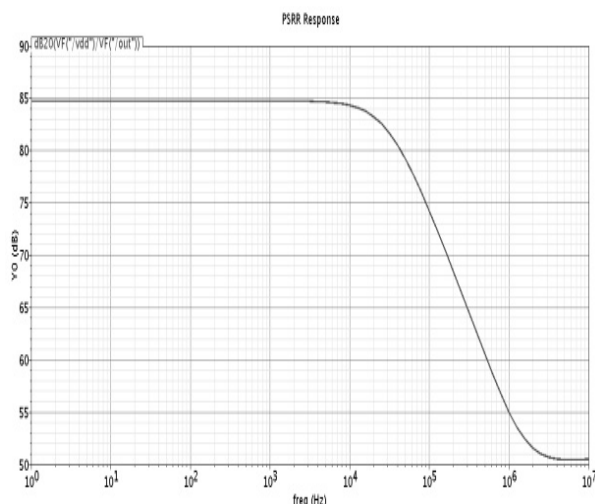


Figure 6: The Power Supply Rejection Ratio (PSRR) response

The Power Supply Rejection Ratio (PSRR) for an operational amplifier describes the amount of noise from a provided power supply that a particular device can reject. The PSRR is also defined as the ratio of change in supply voltage to the differential input voltage, it produces in operational amplifier in decibels. An ideal OTA should have infinite PSRR [8], which is vividly visible in the figure after 10^6 Hz.

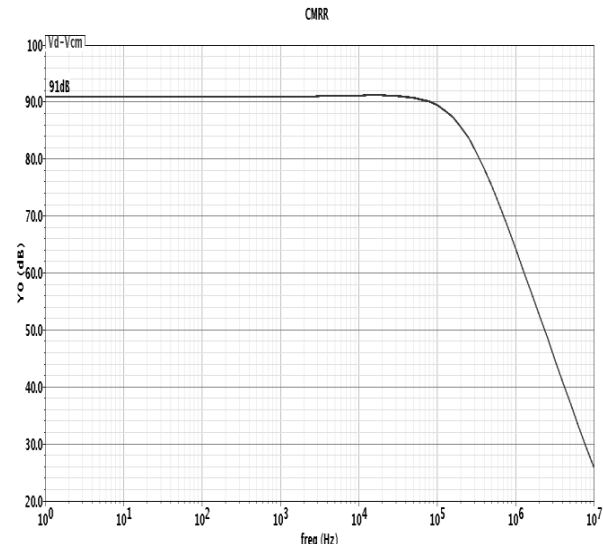


Figure 7: The Common Mode Rejection Ratio (CMRR) response

The Common Mode Rejection Ratio (CMRR) of a normal differential amplifier is the tendency of that particular electronic device to reject the input signals common to both the input ends. A higher CMRR is importantly useful in applications where the signal of interest is by a small voltage fluctuations on a positively large voltage offset or while the relevant information is contained in the two signals' voltage difference. In the above case, CMRR of 91 dB has successfully been achieved. Also, the value of CMRR seldom depends on signal frequency.

TABLE 2: OTA CHARACTERISTICS

Specification	Simulation
CMOS Technology	180 nm
Supply Voltage	1.8V
Bias Voltage	1.8V
Open Loop Gain	72 dB
CMRR	91 dB
PSRR	84 B

CONCLUSION

A simple Operational Transconductance Amplifier (OTA) idea for the low power and low voltage applications has been presented. This Operational Transconductance Amplifier (OTA) can be effectively in low voltage, low power applications. Also, it can be applied to the high time constant applications such as physical transducers, small battery operated devices and process controllers. This piece of work can also be utilized in Analog to Digital (ADC) design, filter design and amplifiers due to its high Common Mode Rejection ratio (CMRR), low power consumption and high gain [9]. Also, one of its principal applications is in implementing electronically controlled applications such as the variable gain amplifier, variable frequency oscillators and filter design [11].

REFERENCES

- [1] A. G. Girardi, F.P. Cortes, L.S. de Paula, S. Bampi Design using Trapezoidal Association of MOS Transistors. Proceedings of IFIP VLSI-SoC 2005, Perth, Australia, October 2005.
- [2] Adel S. Sedra, Kenneth C. Smith "Microelectronic Circuits", Oxford University press, Fourth edition, 2002, pp. 89-91.
- [3] D. A. Johns and K. Martin, Analog Integrated Circuit Design, New York:Wiley, 1997.
- [4] Houda Daoud, Samir Ben Salem, Sonia Zouari, Mourad Loulou, "Folded Cascade OTA Design for Wide Band Applications", Design and Test of Integrated Systems in Nanoscale Technology, 2006.
- [5] Jin Tao Li, Sio Hang Pun, Peng Un Mak and Mang I Vai," Analysis of Op-Amp Power Supply Current Sensing Current Mode Instrumentation Amplifier for Biosignal Acquisition system", IEEE conference, August 2008, pp. 2295-2298.
- [6] M. Banu, J. M. Khoury, and Y. Tsividis, "Fully Differential Operational Amplifier with Accurate Output Balancing ", IEEE Journal of Solid State Circuits, Vol. 23, No. 6, pp. December 1990.
- [7] Mixed signal VLSI Design by Jacob Baker.
- [8] Paul R. Gray, Paul L Hurst, Stephen H. Lewis and Robert G. Mayer, "Analysis and design of analog integrated circuits", Fourth edition, John Wiley and Sons, inc. 2001, pp. 425-439.
- [8] Phillip E. Allen and Douglas R. Holberg "CMOS analog circuit design", second edition, Oxford University press, 2007, pp. 269-274.
- [9] R. Hogervorst, J. P. Tero, R.G.H. Eschauzier, and J. H. Huijsing, "A Compact Power Efficient 3 V CMOS Rail to Rail Input/Output Operational Amplifier for VLSI Cell Libraries", IEEE Journal of Solid State Circuits, Vol. 29, pp. December 1988.
- [10] Y. Tsividis, Operation and Modelling of the MOS Transistor, second edition Boston, MA: Mc Graw Hill, 1998.

A FRAMEWORK FOR FACE RECOGNITION FROM VIDEO SEQUENCES USING HMM

M. R. Baviskar

Dept. of Comp. Science & Engineering,
Govt. Engineering College,
Aurangabad (MS), India.
madhumita_baviskar@yahoo.co.in,

M. B. Nagori

Dept. of Comp. Science & Engineering,
Govt. Engineering College,
Aurangabad (MS), India.
kshirsagarmeghana@gmail.com

Vidhya Ashok

Dept. of Information Technology,
Vidyalankar Institute of Technology,
Mumbai, India.
vidhyashok@gmail.com

ABSTRACT

While traditional face recognition is typically based on still images, face recognition from video sequences has become popular recently. This work describes an ongoing project related to a new method to perform face recognition from video sequences. Faces are detected and tracked in a video sequence using Hidden Markov Model. This process also allows locating and extracting facial feature regions around the eyes, nose and mouth. During the training process, the statistics of training video sequences of each subject, and the temporal dynamics, are learned by an HMM. During the recognition process, the temporal characteristics of the test video sequence are analyzed over time by the HMM corresponding to each subject. Conventional techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), and feature based Elastic Bunch Graph Matching (EBGM) and 2D and 3D face models are well known for face detection and recognition.

Keywords - Face Detection; Face Recognition; Principal Component Analysis (PCA); Hidden Markov Model (HMM)

I. INTRODUCTION

Face detection and recognition has emerged as an active area of research in fields such as security system, videoconferencing and identification. The methods for face recognition can be broadly grouped in two classes, i.e. static and dynamic approaches. The former concerns face recognition in single (static) images, while the latter concentrates on people possibly moving in video sequences. The static approach has developed many interesting and powerful techniques, succeeding in different applications. Nevertheless, as far as recognition in video sequences is concerned, much work still remains to be done.

Face recognition in video sequences often involves four important steps as shown in Figure 1: (1) face detection; (2) feature extraction; (3) feature matching; and (4) recognition. [7]

It should be clear that the large amount of data involved in video sequences represent a challenge for real-time implementation of these four steps. In order to circumvent this problem, not all face regions will be taken into account by the foreseen system. Instead, the system is based on extracting information from important facial feature regions (FFRs) defined around specific landmarks, i.e. the eyes, the nose and the mouth.

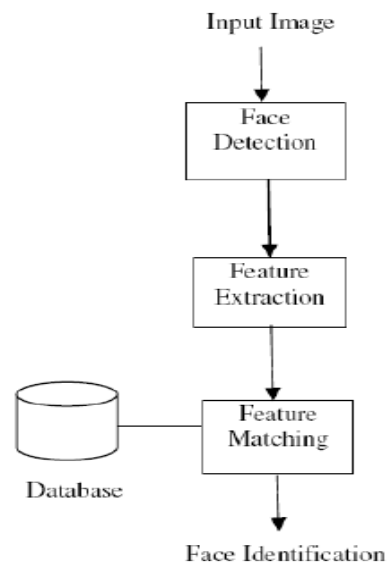


Figure 1. Face Recognition System

II. FACE RECOGNITION – DIFFERENT APPROACHES

Generally, face identification technique can be considered as image based or feature based. The image based methods uses predefined standard face patterns where as feature based techniques concentrate on extracted features such as distance between eyes, skin color, eye socket depth etc. More specifically, face recognition techniques fall in three categories, holistic, feature based model based and hybrid approaches. Figure 2 [7] shows taxonomy of face recognition methods used.

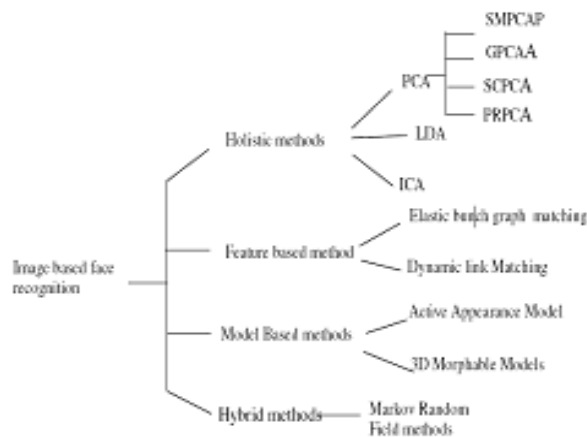


Figure 2. Taxonomy of Face Recognition Approaches

A. Holistic Methods

In holistic or appearance based approach, the whole face region is considered as the input data to the face recognition system. Examples are Eigen faces, fisher faces, probabilistic eigenface etc. These techniques help to lower the dimensions of the dataset without tampering the key characteristics. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) are the most widely used holistic methods.

B. Hybrid Methods

Human perception system recognizes faces using local facial features and the whole face region information. The hybrid method is more akin to human perception system since it is influenced by both feature based and holistic methods. This approach is most effective and efficient in the presence of irrelevant data. The key factors affecting performance depend on the selected features and the techniques used to combine them.

A hybrid method using Markov Random Field (MRF) model is developed fully utilizing the noise immunity feature of MRF models. According to the proposed method, the image is divided into smaller image patches, each having specific Ids. The model can be represented as in the Figure 3[7]. It includes two layers, observable nodes (squares in the graph, representing image patches) and hidden nodes (circles in the graph, representing the patch Ids).

III. METHODOLOGY

C. Principal Component Analysis(PCA)

The Principal Component Analysis (PCA) is one of the most powerful techniques that have been used in image recognition or in compression.[7] The function of PCA is to reduce the large size of the data space (variables) to the smaller intrinsic dimensionality or size of feature space (independent variables), that are needed to describe the data cost efficiently. In PCA, the recognition system is based on the representation of the face images using the so called *eigenfaces*. An illustration of eigen faces is given in Figure 4.

Face Images are projected into a feature space ("Face Space") that best encodes the variation among known face images.

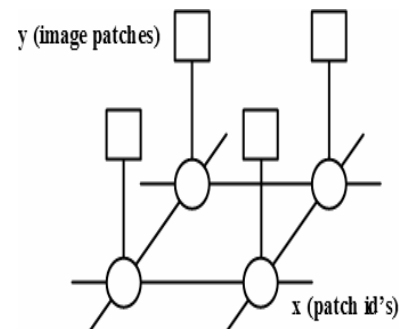


Figure 3. MRF Model



Figure 4. Eigen Faces

The face space is defined by the "eigenfaces", which are the eigenvectors of the set of faces. They can be extracted out of original image data by means of Principle Component Analysis (PCA). A general approach to the PCA is to first solve the characteristic polynomial equation for all eigenvalues and then find their corresponding eigenvectors to produce principal components (PCs).

The basic objective of PCA is to reduce the dimension of the workspace. So principal components are selected in a view that they contain a major quantity of information.

The basic steps of PCA are summarized as follows:

1. Collect x_i of an n dimensional data set $x, i=1,2,3,\dots,m$
2. Calculate mean m_x and subtract it from each data point, $x_i - m_x$
3. Calculate Covariance Matrix $C=(x_i - m_x)(x_i - m_x)^T$
4. Determine eigenvalues and eigenvectors of Holistic methods the matrix c
5. Sort eigen values and corresponding eigenvectors in decreasing order.
6. Select first $d \leq n$ Eigen vectors and generates data set in new representation.
7. Projected test image is compared to every projected training image and result is the training image closest to the test image.

Principal component analysis (PCA) is standard technique used in statistical pattern recognition and signal processing for data reduction and Feature extraction. As the pattern often contains redundant information, mapping it to a feature vector can get rid of this redundancy and yet preserve most of the intrinsic information content of the pattern. These extracted features have great role in distinguishing input patterns. PCA is also known as eigenface method.

The whole approach of PCA face recognition is shown graphically in Figure 5.[2]

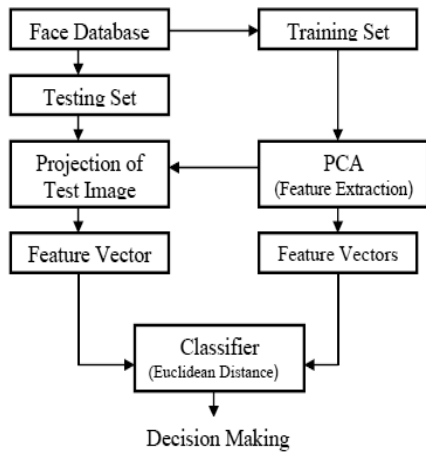


Figure 5. PCA approach for Face Recognition

D. Hidden Markov Model (HMM)

Hidden Markov model (HMM) is a promising method that works well for images with variations in lighting, facial expression, and orientation. The video-based recognition has superior advantages over the image-based recognition.

First, the temporal information of faces can be utilized to facilitate the recognition task. Secondly, more effective representations, such as a 3D face model or super-resolution images, can be obtained from the video sequence and used to improve recognition results. Finally, video-based recognition allows learning or updating the subject model over time. To process images using HMM, the temporal or space sequences are to be considered. In simple terms HMM can be defined as set of finite states with associated probability distributions. Only the outcome is visible to the external user not the states and hence the name Hidden Markov Model.

HMM is defined as the triplets, $\pi = (A, B, \mu)$ [7]

Where

B: Observation symbol probability matrix

A: State transition probability matrix

μ : Initial state distribution

The most significant facial features of a frontal face image include the hair, forehead, eyes, nose and mouth. The image of the face may be modeled using a one-dimensional HMM by assigning each of these regions to a state as illustrated in Figure 6.

In this paper, we apply adaptive HMM temporally to perform the video-based face recognition. As shown in Figure 7, during the training process, the statistics of training sequences, and their temporal dynamics are learned by an HMM [5].

During the recognition process, the temporal characteristics of the test video sequence are analyzed over time by the HMM corresponding to each subject. Our proposed algorithm can learn the dynamic information and improve the recognition performance

compared to the conventional method that simply utilizes the majority voting of image-based recognition results.

In this paper, we apply adaptive HMM temporally to perform the video-based face recognition.[8]

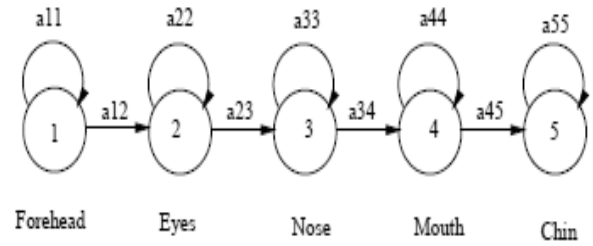


Figure 6. HMM for Face Recognition

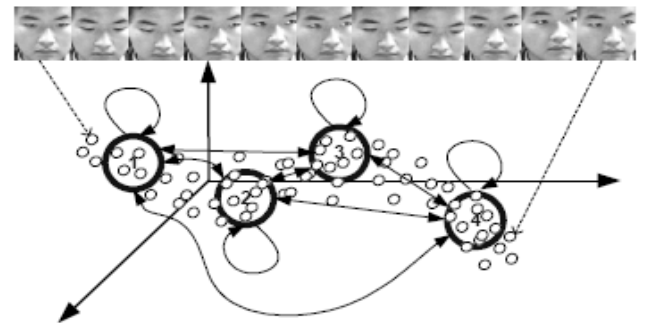


Figure 7. Temporal HMM for modelling face sequences

IV. OUR PROPOSED ALGORITHM

In this section, we describe our proposed algorithms in detail. First, feature extraction, HMM training and HMM testing are presented. Then, we introduce an algorithm to adapt the HMM in order to enhance the modeling over time.

E. Temporal HMM

When applying HMM to face recognition, researchers have proposed to use different features, for example, pixels values, eigen-coefficients and DCT coefficients, as the observation vectors. In our algorithm [5], each frame in the video sequence is considered as one observation. Since PCA gives the optimal representation of the images in terms of the mean square error, all face images are reduced to low-dimensional feature vectors by PCA. Given a face database with L subjects and each subject has a training video sequence containing T images.

$$F_l = \{ f_{l,1}, f_{l,2}, \dots, f_{l,T} \} \quad 1 \leq l \leq L$$

Each image only contains the face portion. By performing eigen-analysis for these $L \times T$ samples, we obtain an eigenspace with a mean vector $\mathbf{m}$ and a few eigenvectors $\{V_1, V_2, \dots, V_d\}$. All the training images are projected into this eigenspace and generate corresponding feature vectors, e_{lt} , $1 \leq l \leq L$, $1 \leq t \leq T$, which will be used as observation vectors in the HMM training. At this stage, we also compute the covariance matrix of all the feature vectors $\mathbf{e}_t$, $\mathbf{C}_e$, which is a diagonal matrix with eigenvalues as the diagonal elements.

Each subject in the database is modeled by a N -state fully connected HMM. The feature vectors e_{it} , $1 \leq t \leq T$ form the observation vectors $\mathbf{O}$ for training the HMM of Subject l . During the training process, when the number of face images assigned to each state is less than the dimension of feature vectors, $\mathbf{U}l_k$ will be a singular matrix.

In the recognition process, given a video sequence containing face images, all frames are projected into the eigenspace and the resulting feature vectors form the observation vectors. Then the likelihood score of the observation vectors given each HMM is computed, where the transition probability and observation pdf are used. The face sequence is recognized as Subject k if:

$$P(\mathbf{O} | \lambda_k) = \max P(\mathbf{O} | \lambda_i)$$

F. Adaptive HMM

In typical speech recognition systems, there is a dichotomy between speaker-independent and speaker dependent systems. While speaker-independent systems are ready to be used without further training, their performance is usually two or three times worse than that of speaker dependent systems. However, the latter requires large amounts of training data from the designated speaker. To address this issue, the concept of speaker adaptation has been introduced, where a small amount of data from the specific speaker are used to modify the speaker independent system and improve its performance.

Motivated by these ideas, we propose to use adaptive HMM for video-based face recognition. That is, during the recognition process, after we recognized one test sequence as one subject, we can use this sequence to update the HMM of that subject, which will learn the new appearance in this sequence and provide an enhanced model of that subject.

In our algorithm [5], we will use the *likelihood difference*, i.e., the difference between the highest likelihood score and the second highest likelihood score, as the feature to make the decision. The reason is that for correct recognition, the likelihood difference tends to be large; while for incorrect recognition, the likelihood difference is usually small. So given a test sequence, we compare its likelihood difference with a predefined threshold, and update the HMM only if the likelihood difference is larger than the threshold. In practice, this pre-defined threshold can be determined by performing experiments on a cross-validation data set.

We use the standard MAP adaptation technique to adapt the HMM. That is, given an existing HMM λ^{old} and observation vectors $\mathbf{O}$ from a test sequence, we estimate a new HMM, $\lambda = (A, B, \pi)$. We use λ^{old} as the initial parameters of λ , and the EM algorithm is used to reestimate λ except that the mean estimation is as follows:

$$\mu_{ik} = (1 - \beta)\mu_{ik}^{old} + \beta \frac{\sum_{t=1}^T \mathbf{O}_t P(q_t = i, m_{qt} = k | \mathbf{O}, \lambda)}{\sum_{t=1}^T P(q_t = i, m_{qt} = k | \mathbf{O}, \lambda)}$$

where μ_{ik}^{old} is the mean vector from the existing HMM λ^{old} and β is a weighting factor that gives the bias between the previous estimate and the current data. In our experiments, we choose β to be 0.3. Also during the iteration of the EM algorithm, we do not update the covariance matrix of the observation pdf, $\mathbf{U}_{ik}$, because from speech research literature, the major discriminative information of an HMM is retained by the mean vectors instead of the covariance matrixes.

The following Figure 8 represents the proposed HMM based face recognition system.

V. APPLICATIONS

The recognition of faces is very important for many applications such as:

- Video Surveillance
- Retrieval of an identity from a data base for criminal investigations
- Forensic Applications
- Security Methods
- Biometrics
- Access Control
- Law Enforcement

CONCLUSION

In this paper, we propose to use adaptive HMM to perform video-based face recognition. During the training process, the statistics of training video sequences of each subject, and their temporal dynamics are learned by an HMM. During the recognition process, the temporal characteristics of the test video sequence are analyzed over time by the HMM corresponding to each subject. The likelihood scores provided by the HMMs are compared, and the highest score provides the identity of the test video sequence.

An automatic face detection and recognition system is being developed. The system is based on face detection and face recognition schemes. The system is intended for video indexing applications and will be able to tell whether or not a specific person is present in a video sequence shot.

REFERENCES

- [1] Teófilo Emídio de Campos, Rogério Schmidt Feris, Roberto Marcondes Cesar "A Framework for Face Recognition from Video Sequences using GWN and Eigenfeature Selection".
- [2] Mr. Udesang K Jaliya, Mr.Keyur Brahmabhatt, Mr.Samip A Patel, Mr.Premal J Patel "A Comparative Study of PCA & LDA Human Face Recognition Methods", Proceedings of National Conference on Recent Trends in Engineering & Technology, 13-14 May, 2011.

- [3] Emiliano Acosta, Luis Torres Alberto Albiol "An Automatic Face Detection and Recognition System for Video Indexing Applications"
- [4] M. A. Turk, A. P. Pentland "Face Recognition using eigenfaces", Proceedings of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 586-591, Maui, Hawaii 1991.
- [5] Xiaoming Liu and Tsuhan Chen "Video-Based Face Recognition Using Adaptive Hidden Markov Models", In proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, Vol. 1, I-340-I-345 .
- [6] Alberto Albiol, Luis Torres, Charles A. Bouman, Edward J. Delp "A simple and efficient Face Detection Algorithm for Video Database Applications".
- [7] S. Sharavanan, M. Azath, "LDA based Face Recognition by using Hidden Markov Model in current trends", Proceedings of International Journal of Engineering and Technology Vol.1(2), 2009, 77-85.
- [8] Ara V. Nefian, Monson H. Hayes III, "Face Recognition using an Embedded HMM".
- [9] Proyecto Fin de Carrera, Ion Marqués "Face Recognition Algorithms", Ph.D Thesis, June 16, 2010.
- [10] Mohammad Abdul Kashem, Md. Nasim Akhter, Shamim Ahmed, and Md. Mahbub Alam, "Face Recognition System Based on Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN)", In proceedings of International Journal of Scientific & Engineering Research Volume 2, Issue 6, June-2011 1 ISSN 2229-5518
- [11] Xiaoou Tang, Zhifeng Li, "Audio-Guided Video-Based Face Recognition", Proceedings of IEEE Transactions on Circuits and System for Video Technology, Vol.19, No. 7, July 2009
- [12] Unsang Park "Face Recognition: Face in video, age invariance, and facial marks", Ph.D Thesis, Dept. of Computer Science, Michigan State University, 2009.

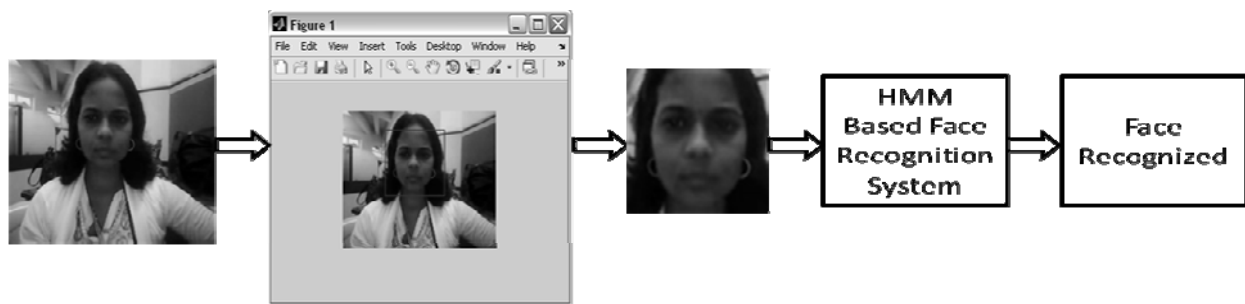


Figure 8. Proposed HMM based Face Recognition

- [13] M.R. Baviskar, V.P. Kshirsagar, M.E. Gaikwad, "Face Recognition using Eigenfaces", in proceedings of IEEE Xplore on Computer Research and Development, Vol. 2, 11-13 March 2011
- [14] http://wikicoursenote.com/wiki/Video-based_face_recognition_using_Adaptive_HMM

SUPER RAILWAY TICKETING USING ZIGBEE

Kunjesh Mehta

Dept. of Computer Engineering,
Rajiv Gandhi Institute of Technology,
Andheri, Mumbai-400 053
mehtakunjesh@yahoo.com,

Jinesh Gandhi

Dept. of Computer Engineering,
Rajiv Gandhi Institute of Technology,
Andheri, Mumbai-400 053
jinesh_smart@yahoo.co.in,

Dwiti Shah

Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology,
Andheri, Mumbai-400 053
shah.dwiti@yahoo.com

ABSTRACT

Smart cards are now-a-days being used in many fields for identification, authentication, and data storage purpose. This paper is mainly concerned on utilities of wireless smart cards in procuring railway tickets. Looking at the history, when the first railway opened, tickets were simply slips of paper with all the details written in by hand. Then, a more efficient system that consisted of small pieces of card, each pre printed with the entire journey details and stored in racks. But, there is always a great rush at ticket counters whether local or long distance trains resulting in an increase in "labor cost". With these two issues in mind, it is the primary object of the present paper to provide an automatic wireless ticket substation wherein the customer will have to just carry a wireless zigbee based card while entering; no need to queue up anywhere.. The move is aimed at no queues for tickets.

Keywords— Zigbee; Smart card; Pay-as-you-go; Railway; Tickets.

1. INTRODUCTION

The Mumbai Suburban Railway (Marathi: मुंबई उपनगरीय रेल्वे) is a mass transit system serving the Mumbai Metropolitan Region. It is part of the public transport system of Mumbai, provided by the state-run Indian Railways' two zonal Western Railways and Central Railways. It has the highest passenger densities of any urban railway system in the world. The trains plying on its routes are commonly referred to as local trains or simply as locals by the general populace. The extensive network of Railways has three lines, called the Western Line, the Central Main Line and the Harbor Line.

This paper deals with automatic wireless railway ticketing system using ZigBee. ZigBee technology is a wireless networking technology built around the IEEE 802.15.4 wireless protocol. It is a low data rate, low power consumption, low cost, wireless networking protocol targeted towards automation and sensor remote control applications.

In this system the passenger will only have to carry a wireless smart card. Ticketing process will

automatically take place at entry/exit points with no passenger involvement. It is similar to prepaid system wherein the card will carry balance and transaction will automatically be performed.

2. EXISTING SYSTEM

Tickets for the suburban trains can be purchased at every train station. The various ways to buy tickets are as follows:

2.1 Traditional counter based ticketing system.

2.1.1 Weakness

- Long Queues.
- Wastage of paper.
- Labor Cost.
- Time Consuming.

2.2 Coupon Validating Machines (CVMs)



Fig. 1 CVM coupons

A Coupon Booklet was introduced where customers can punch them for designated fare at the Coupon Validating Machines (CVMs) at every station. The ticket fares matrix is pasted above the CVM.

2.2.1 Weakness

- Wastage of Paper.
- Faulty Machine.
- Queue.
- Miss use of Coupons

2.3 Automatic Ticket Vending Machine (ATVMs)



Fig. 2 ATVM

Further Smart Cards were introduced that can be topped up (recharged with some amount) and one can use it to print tickets for themselves from an Automatic Ticket Vending Machine (ATVMs).

2.3.1 WEAKNESS

- Faulty Card
- Not user friendly
- Faulty machine.

Although the advancement over the years, long queues still exist due to ever increasing travelling population and faulty machines. Thus there exist a great scope for a more advanced and efficient ticketing system which is conveyed by this paper.

3. PROPOSED SYSTEM

This section describes how Zigbee can improve ticketing operations. We have used different xbee products at every level for Zigbee implementation and will be using Xbee and Zigbee terms alternatively. Xbee is the brand name from Digi International for a family of form factor compatible with Zigbee.

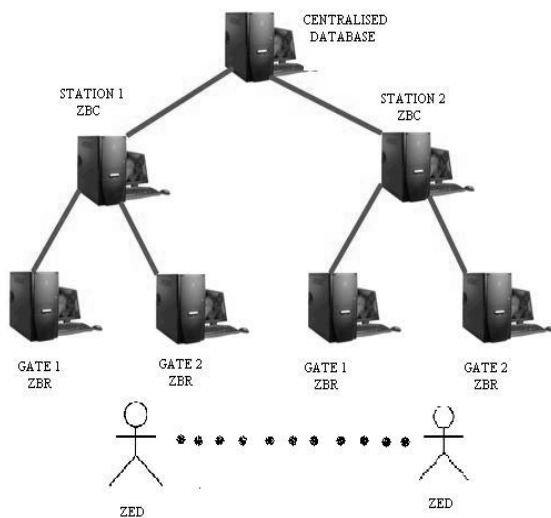


Fig. 3 Architecture

Passengers will be carrying Zigbee end device (ZED) card each. The system will support many passengers at a time by a hierarchy of xbee modules as shown in the diagram. Each station may have more than one entry/exit gate, each gate will have Zigbee based router (ZBR). Each Zigbee based router can accommodate

several xbee clients. Further these routers are connected to Zigbee based co-coordinator (ZBC) where database for each station is maintained. All these co-coordinators are further connected to centralized database through internet or Private MAN..

3.1 Protocol Specification

ZigBee is a specification for a suite of high level communication protocols using small, low-power digital radios based on an IEEE 802 standard for personal area networks. ZigBee devices are often used in mesh network form to transmit data over longer distances, passing data through intermediate devices to reach more distant ones. This allows ZigBee networks to be formed ad-hoc.

ZigBee is targeted at applications that require a low data rate, long battery life, and secure networking. ZigBee has a defined rate of 250 kbit/s, best suited for periodic or intermittent data or a single signal transmission from a sensor or input device. The technology defined by the ZigBee specification is intended to be simpler and less expensive than other WPANs, such as Bluetooth.

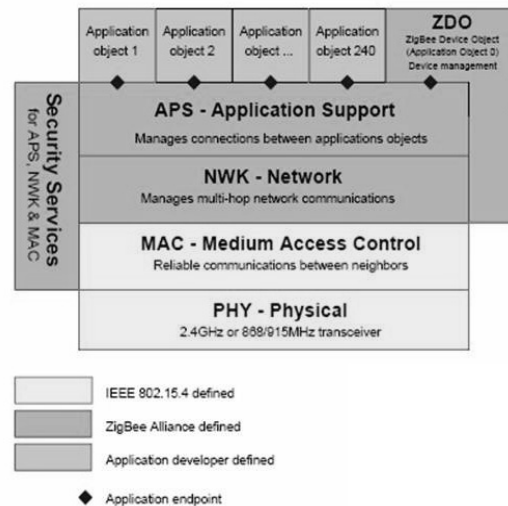


Fig. 4 Protocol stack

ZigBee builds upon the physical layer and medium access control defined in IEEE standard 802.15.4 (2003 version) for low-rate WPANs. The specification goes on to complete the standard by adding four main components: network layer, application layer, ZigBee device objects (ZDOs) and manufacturer-defined application objects which allow for customization and favor total integration.

Besides adding two high-level network layers to the underlying structure, the most significant improvement is the introduction of ZDOs. These are responsible for a number of tasks, which include keeping of device roles, management of requests to join a network, device discovery and security.

3.3 Project Flow

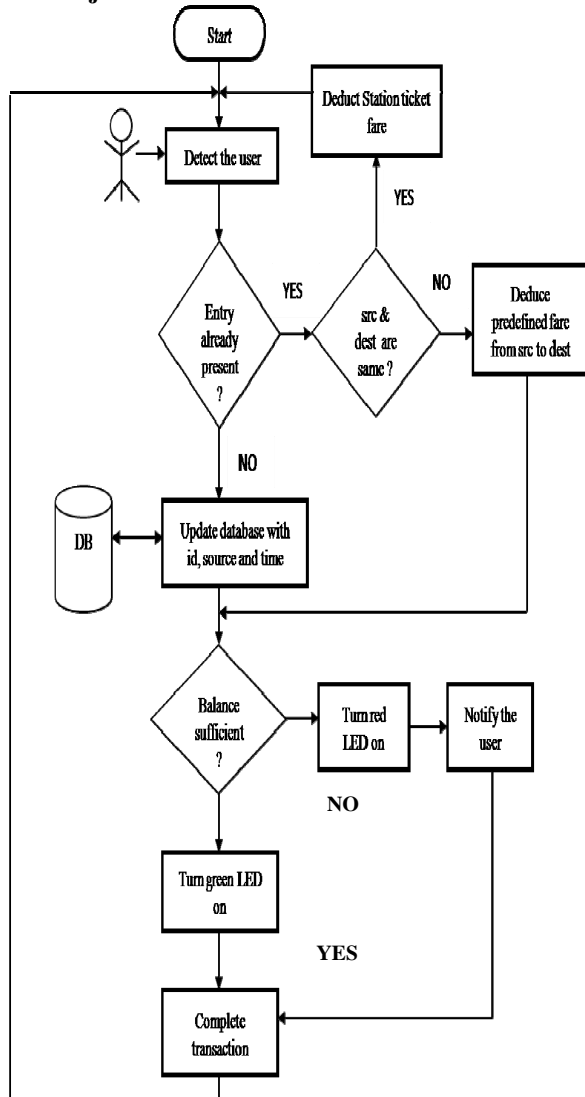


Fig. 5 Flow diagram

The system processes ticket as follows:

- The ZBC at the station continuously senses for the passenger i.e the card (ZED).
- Once detected it checks if the entry is already present for that card.
- If the entry doesn't exist, ZBC will put one and will acknowledge ZED and will pass information to the centralised database.
- If entry is present ZBC will check its source name from centralised database.
- According to the source it will either deduct the platform ticket or the predefined fare between source and destination.
- Since balance is subjected to the card, it may not have sufficient balance.
- So the card will consist of two lights Red and Green. Red indication no balance and Green indicating very low balance. This will help the ticket checker for any fraud.
- The process is repeated 24*7 at every gate of every station.

3.4 System components

- Customer End Device-Zigbee End Device(ZED)
- Entry/Exit Point-Zigbee Router(ZBR)
- Station Coordinator-Zigbee Coordinator(ZBC)
- Centralised Database

3.4.1 Customer End Device-Zigbee End Device (ZED)

The Xbee shield allows an Arduino board to communicate wirelessly using Zigbee. It is based on the Xbee module from MaxStream. The module can communicate up to 100 feet indoors or 300 feet outdoors (with line-of-sight). It can be used as a serial/usb replacement or you can put it into a command mode and configure it for a variety of broadcast and mesh networking options. Here we use Xbee shield in command mode so we can configure it. X-CTU is a free License software which is used to configure the Xbee module.

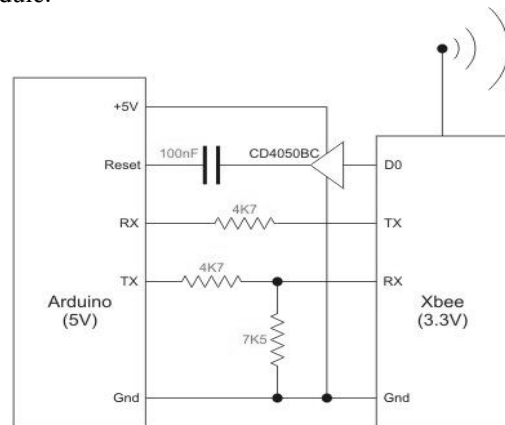


Fig. 6 ZigBee End Device

3.4.2 Entry/Exit Point-Zigbee Router (ZBR)

Same Xbee shield is used but configured differently. There are multiple parameters that need to be configured correctly for two modules to talk to each other (although with the default settings, all modules should be able to talk to each other). They need to be on the same network, as set by the ID parameter. The modules need to be on the same channel, as set by the CH parameter. Finally, a module's destination address (DH and DL parameters) determine which modules on its network and channel will receive the data it transmits. Here we configure Xbee similar to third point i.e. fixed address for the receiving and destination module.

3.4.3 Station Coordinator-Zigbee Coordinator (ZBC)

ZBC is directly connected to pc, so we just require an usb adapter or an explorer compatible with Xbee. In the proposed system we are using DFRobot Xbee USB explorer. It is used for the XBee module configuration parameters to facilitate or to use as a wireless data transmission. It can be easily connected to a PC via mini USB cable. Further, it has XBee-setting support software X-CTU. It can also be used as a USB-TTL adapter.

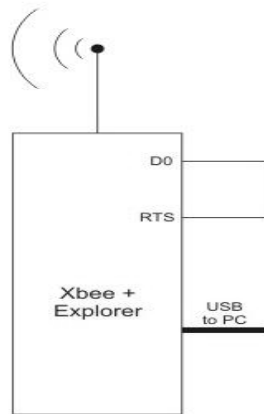


Fig. 7 ZigBee Based Router

3.4.4 Centralised Database

All station servers are connected to main central database. Database contains all the details about the Xbee card details such as id, balance, time, history, date, entry, exit, remark and other details. The system uses Microsoft SQL 2005 for distributed database.

Linked servers provide SQL Server the ability to access data from remote data sources. Using these mechanisms, you can issue queries, perform data modifications and execute remote procedures.

Fare calculation is performed as follows:

- Database will contain a separate table consisting of predefined rates.
- if entry='A', the information is transmitted from ZBR to ZBC to central database.
- Now when the card holder exits from a station the station name is transferred to central database e.g exit='B'.
- Now these entry and exit gates are compared with predefined table rates i.e fare for 'A' to 'B' is Rs.x.
- These rates are automatically deducted from balance.

4. ADVANTAGES

- Reducing operating costs, as a result of reducing the number of disposable tickets and mechanical parts in readers and vending systems
- Increasing ease of use by automating ticket vending and payment handling
- Increasing the throughput of existing ticketing systems; one way of accomplishing this is to migrate from magnetic stripe-based technology to wireless smart cards.
- Maximizing revenue, through the use of differentiated fare policies that enable automatic fare modification.
- Making it easier to replace lost or stolen cards.
- Increasing the control over a ticket's life cycle, thus reducing fraud

5. ACKNOWLEDGEMENT

After we give you a brief insight into the project work of "SMART RAILWAY TICKETING USING ZIGBEE", we would like to pause and remember the people who made the project possible and the people who made the project work a divine experience. My deepest thanks to Mr. Swapnil Gharat, the guide of our project for guiding and for helping us and giving us innovative ideas for improving our work. We would also thank our institution and faculty members without whom this work would have been a distant reality

CONCLUSION

Unless and until one supplements his theoretical knowledge with practically implementing all the concepts, he will not be able to retain the knowledge gained through books for long. So, we are very grateful to our college that we have been given a chance to make this project. This project was a very good learning experience. During the implementing process we could learn working of various components and thus the way in which we can use them in the making of a system.

REFERENCES

- [1] <http://www.zigbee.org/en/about/faq.asp>
- [2] 'Home Networking with Zigbee' by Mikhail Galeev
- [3] http://www.wisegEEK.com/what-is-zigbee.htm?referrer=adwords_campaign=zigbee_ad=013761&_content_kw=what%20is%20zigbee
- [4] 'Multipath Characterization of Airframe Environments for Wireless Networks' by Christopher
- [5] 'Time Synchronization for ZigBee Networks' by Dennis Cox, Emil Jovanov, Aleksandar Milenkovic.
- [6] 'Reliable Broadcasting in ZigBee Networks' by Zafer Sahinoglu, Philip Orlik, Jinyun Zhang and Bharat Bhargava.
- [7] <http://en.wikipedia.org/wiki/ZigBee>
- [8] www.dfrobot.com
- [9] www.arduino.cc
- [10] 'ZigBee/IEEE 802.15.4 Summary' by Sinem Coleri Ergen
- [11] www.digi.com/xbee.com
- [12] James Kurose & Keith W. Ross, "Computer Networks", Fourth Edition, Pearson Publication Limited, 2004, Pp 49-98.
- [13] Jiang, Y., Cao, J., & Du, Y. (2006). Unmanned air vehicle landing based on Zigbee and vision guidance. WCICA 2006, 2, 10310 - 10314.
- [14] ZigBee Specification v1.0, ZigBee Alliance, December 14th, 2004

NORMALITY AND STATIONARITY ANALYSIS OF SIGNALS USING KOLMOGOROV-SMIRNOV TEST

Mr. Shrikrishna Kolhar
VPCOE, Baramati.
shrikolhar88@gmail.com

Mrs. Aparna Shastri
VPCOE, Baramati.
aparnadjoshi@yahoo.com

Mr. Rajveer Shastri
VPCOE, Baramati
rajveer_shastri@yahoo.com

ABSTRACT

In most of the signal processing operations various assumptions are made about the signals or data to be processed. Some assumptions about signals (or stochastic processes for that matter) are so widely and commonly used that they are almost considered as absolute truths. Most of the time signals are assumed to be stationary or the noise is Additive White Gaussian Noise (AWGN) to make the calculations easier. But any departure from above assumptions might affect on level of reliability of estimates. Hence in this paper Normality of signals is quantified using Kolmogorov-Smirnov (K-S) one sample test and stationarity of signals is analyzed using K-S two sample test. Stationarity of signals is analyzed from local stationarity and stationarity length point of view. First computer generated random data are considered and then heart signals are analyzed using K-S tests and under-water acoustic signals are to be analyzed in future.

I. INTRODUCTION

The Kolmogorov-Smirnov test (KS-test) tries to determine if two datasets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data. Technically speaking it is non parametric and distribution free. One of the advantages of the KS-test is that it leads to a graphical presentation of the data.

The tests described and applied here can be described as hypothesis testing. Testing the hypothesis is inferring from a sample whether or not a statement about the population is true, such as sonar data is stationary or ambient noise is Gaussian. In hypothesis testing there are two hypothesis involved, the null hypothesis H_0 and alternative hypothesis H_1 . For the example above sonar data is stationary the null hypothesis is

H_0 : Sonar data is stationary
and the alternative hypothesis is
 H_1 : Sonar data is not stationary

Once the hypothesis is postulated, a test statistic is chosen, and a decision rule is made. The decision rule is to find the possible values that the test statistic can take. Then the test statistic is calculated from a sample from the population, and a decision is made on accepting or rejecting the null hypothesis.

There are two ways to make errors in this kind of hypothesis testing. The first being, rejecting a true null hypothesis, the second is accepting a false null hypothesis. Rejecting a true null hypothesis is called a type I error, and accepting a false null hypothesis a type II error. The critical region of a test statistic is also

called a rejection region, and it is defined as the set of all points in the sample space where the null hypothesis is rejected. The set of all points in the sample space where the null hypothesis is accepted is called an acceptance region. If the null hypothesis is false there is a possibility to accept it with a probability Q . The Power of a test is the ability to reject a false null hypothesis, and it is denoted as $1-\beta$.

TABLE I - TYPE OF ERRORS

H_0	Accept H_0	Reject H_0
H_0 is true	Correct decision probability= $1-\alpha$	Type I error probability= α , level of significance
H_0 is false	Type II error probability= β	Correct decision probability= $1-\beta$ power

Another part of hypothesis testing is the p-value. The p-values is the smallest significance level where the null hypothesis is rejected for a given observation, meaning that, this is the value that the computed test statistic is compared to [1].

II. BASIC DEFINITIONS

A. Stationarity

Consider a stochastic process $x(t)$ with two distinct sets of samples x_{t1} and x_{t1+T} , where the second sample is merely displaced in time by an arbitrary constant T . The two sets are characterized by their joint probability density functions,

$$P(x_{t1}, x_{t2}, x_{t3}, \dots) \quad (1)$$

And

$$P(x_{t1+T}, x_{t2+T}, x_{t3+T}, \dots) \quad (2)$$

These two pdfs may or may not be identical. If they are equal the process is stationary in the strict sense (or simply, stationary). However, there are a variety of different degrees of stationarity and nonstationarity.

B. Non-stationarity

When the two pdfs are not equal, as can be seen below

$$P(x_{t1}, x_{t2}, x_{t3}, \dots) \neq P(x_{t1+T}, x_{t2+T}, x_{t3+T}, \dots) \quad (3)$$

the process is said to be non-stationary.

C. Wide sense stationarity

Wide sense stationarity or weak stationarity is the least strict degree of stationarity and it is defined as follows. The first requirement is that the mean of the signal is constant, For discrete time series the mean is calculated as follows

$$\frac{1}{n} \sum_{n=1}^n x(n) = m \quad (4)$$

This is rarely a limitation in practice since the mean of the signal is often subtracted prior to other estimation or analysis.

The second requirement is that the ACF (Auto Correlation Function) only depends on the time lag T, and not time itself, i.e. it is time independent. The ACF for discrete time $R_x(k)$ is defined,

$$R_x(k) = \sum_n x(n)x(n-k) = f(k) \quad (5)$$

where $f(k)$ is a function of the lag k . Stationarity in the strict sense implies wide sense stationarity, but the reverse is not true., except for those processes that are fully described by the first- and second-order moments, mean and variance, such as a **Gaussian process**. The higher than second-order moments of odd order for a Gaussian process are zero, and the even-order moments higher than second-order are described by the variance.

D. Nth- order stationarity

A more strict class of signals are the ones where the nth-order moment is time independent. The nth-order moment is defined as follows

$$E\{x(t)^n\} = \int_{-\infty}^{\infty} x(t)^n dF(x(t)) \quad (6)$$

where $F(X)$ is the distribution function for $x(t)$. For a process to be labelled n:th order stationary all the moments of the process up to order n have to be time independent. The moments about the mean are defined as follows

$$\mu_n = \int_{-\infty}^{\infty} (x(t) - m)^n dF(x(t)) \quad (7)$$

which for a zero mean process reverts back to the previous form, [1] Chapter 15. The moments about the mean are also called central moments, from here on when talking about moments the central moment, μ_n is assumed, again there being no difference if the signal or process is zero mean.

E. Gaussianity

Gaussianity implies Gaussian process with probability distribution function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

where, parameter μ is the mean & σ^2 is the variance. The distribution with $\mu = 0$ & $\sigma^2 = 1$ is called a standard normal distribution. The higher than second-order moments of odd order for a Gaussian process are zero, and the even-order moments higher than second-order are described by the variance.

III. KOLMOGOROV-SMIRNOV TESTS

A. K-S one sample test (Goodness-of-fit test)

A test for goodness of fit usually involves examining a random sample from some unknown distribution in order to test the null hypothesis that the unknown distribution function is in fact a known, specified function. A random sample $X_1, X_2, \dots, X_n$ is drawn from

some population $F(x)$ and is compared with $F^*(x)$ in some way to see if it is reasonable to say that $F^*(x)$ is the true distribution function of the random sample. The hypotheses are given by

H0: $F(x) = F^*(x)$ for all x from $-\infty$ to ∞

H1: $F(x) \neq F^*(x)$ for at least one value of x

One logical way of comparing the random sample with $F^*(x)$ is by means of the empirical distribution function $S(x)$. Let $X_1, X_2, \dots, X_n$ be a random sample then empirical distribution function $S(x)$ can be defined as a function of x , which equals the fraction of X_i 's that are less than or equal to x for each $x, -\infty < x < \infty$

$$S(x) = \frac{1}{n} \sum_{i=1}^n I_{(X_i \leq x)} \quad (9)$$

The empirical distribution function $S(x)$ is useful as an estimator of $F(x)$, the unknown distribution function of X_i 's. This empirical distribution function $S(x)$ is compared with hypothesized distribution function $F^*(x)$ to see if there is good agreement. One of the simplest measures is the largest distance between the two functions $S(x)$ and $F^*(x)$, measured in a vertical direction. This is the statistic suggested by Kolmogorov in 1933 called as k-s test statistic T which is given by,

$$T = \sup_x |F^*(x) - S(x)| \quad (10)$$

where "sup" stands for supremum i.e. maximum.

Here the decision to reject the null hypothesis is based on comparing the p-value with α , not by comparing the test statistic with the critical value CV. CV is computed separately using approximate formulae or by interpolation in a table. For large n the asymptotic formulae are given by

TABLE II- FORMULAE FOR P-VALUE

Significance level(α)	0.20	0.10	0.05	0.01
Formulae	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Since CV is approximate, a comparison of K-S statistic with CV may occasionally lead to a different conclusion than a comparison of P with α [2], [3].

B. K-S two sample test

K-S two sample test compares two cdfs of two samples from same distribution and finds whether or not there is significant difference between them. If they are approximately equal, according to test statistic, then the process is stationary. The K-S two sample test uses two estimated cumulative distribution functions (CDF) to produce the test statistic. It is defined as the maximum absolute distance between the two unknown CDFs, and the formal definition follows

$$T_{KS} = \sup_x |CDF_1(x) - CDF_2(x)| \quad (11)$$

The assumptions made on the underlying data samples are:

1. The samples are random samples
2. The two samples are mutually independent

3. The measurement scale is at least ordinal

4. For the test to be exact the variables are assumed to be continuous.

If the variables are discrete the test is valid but becomes conservative, meaning that the risk of making a type I error is less than stated, [1].

Here the two sided K-S test has been applied with a significance level of 0.05.

The approximation for large samples,

$$p = \frac{1.36}{\sqrt{n}} \quad (12)$$

is used since the number of points, n is larger than 40. The p -value is the threshold that the test statistic is compared to [2], [4], [5].

IV. DATA ANALYSIS

A. K-S one sample test for Normality checking

Here K-S one sample test is used to assess the normality of signals using a 1-second sliding window with 90% overlap[3]. The simulated signal used here is a combination of randomly generated normal signals with different μ and σ values with different durations given in following table.

TABLE III- COMPUTER GENERATED SIGNAL DESCRIPTION.

Signal	μ	Σ	Duration (no. of samples)
N1	0	1	2 lakh
N2	1	2	3lakh
N3	2	3	2lakh
N4	3	4	3lakh

Here the null hypothesis is that the signal is standard normal. Fig.1(b) shows the result of k-s one sample test where we can see that p -value is greater than k -s statistic for the duration of 2 to 4 lakh and 10 to 12 lakh samples indicating standard normal nature of the signal whereas elsewhere p -value is less than k -s statistic indicating elsewhere signal is not standard normal.

Same method is used to check normality of heart signal shown in Fig.2(a) with sample rate of 22255 samples/second. Signal is not standard normal as p -value is less than k -s statistic (k) at all the times as shown in Fig.2(b).

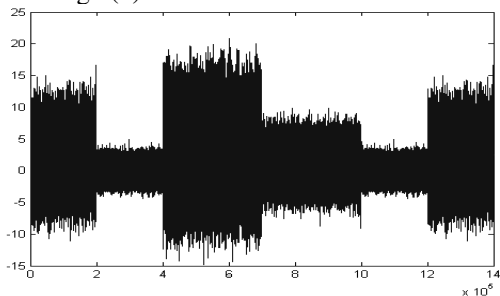


Fig.1(a) Computer generated signal by combining randomly generated normal signals with different mean and standard deviation values shown in TABLE III (N3+N1+N4+N2+N1+N3)

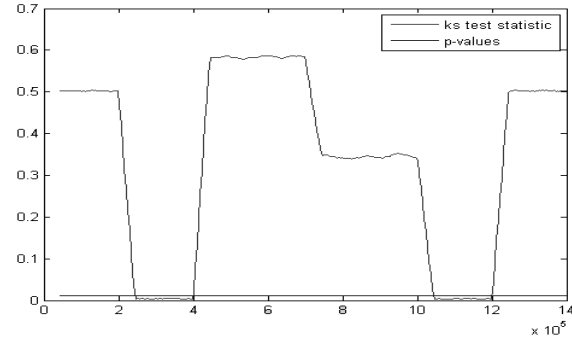


Fig.1(b) Result of K-S one sample test applied on signal in Fig.1(a).

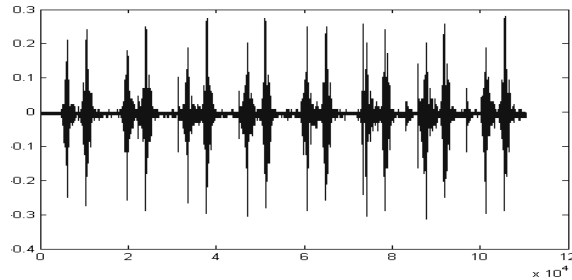


Fig.2(a) Heart signal with sample rate of 22255 samples/second.

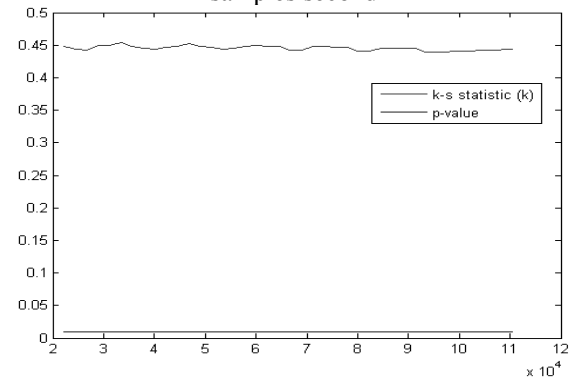


Fig.2(b) Result of K-S one sample test applied on heart signal shown in Fig.2(b).

B. K-S two sample test for stationarity analysis

Method I

In order to test degree of stationarity, the signal was divided into number of non-overlapping, equal length segments and K-S test was applied to consecutive pair of segments. Length of segments is taken to be 100ms for signals with higher sampling frequencies and 10ms for signals with low sampling frequency to see the transitions more accurately. Non-stationary signal with sampling rate of 44100samples/second as shown in Fig.3(a) is created by concatenation of randomly generated normally distributed signals having different mean and standard deviation values as shown in table. Signal was divided into 100ms non-overlapping equal length segments and k -s test applied on it [2]

TABLE IV-MATLAB GENERATED SIGNAL DESCRIPTION

Signal	μ	σ	Duration (no. of samples)
N1	0	1	2lakh
N2	10	2	3lakh
N3	20	3	2lakh
N4	30	4	3lakh

As it can be seen in Fig.3(b) p-values shown by blue color are always greater than K-S statistics (k) shown by red almost at all instants of time except at transitions between two concatenated signals with different means and standard deviations i.e. at $x=3$ lakh, 5 lakh and 8 lakh. At these transitions p-value is less than K-S statistic (k) due to variation in moments of signals with respect to time thus indicating non-stationarity of the signal at these transition point.

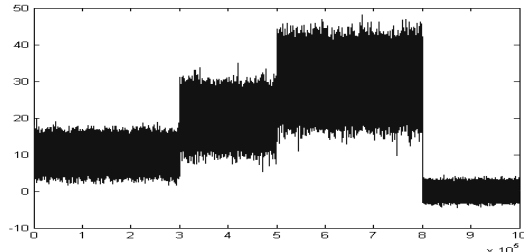


Fig.3(a) Computer generated signal by combining randomly generated normal signals with different mean and standard deviation values shown in TABLE IV ($N_2+N_3+N_4+N_1$)

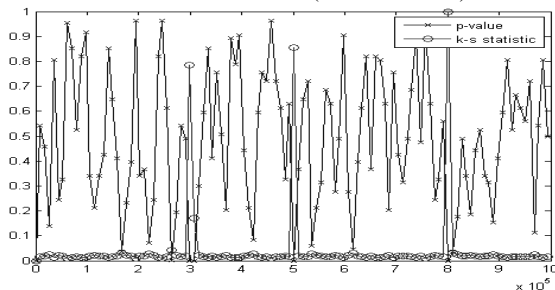


Fig.3(b) Result of K-S two sample test applied on signal shown in Fig.3(a).

Same method is used to find the degree of stationarity of heart signal with sampling rate of 22255 samples/second as shown in Fig.4(a). As it can be seen in Fig.4(b) p-value is greater than k-s statistic (k) for 12.6050% of the time for the signal shown in Fig.4(a) and hence the degree of stationarity of the signal is 12.6050% and signal is non-stationary elsewhere. For another heart signal shown in Fig.5(a) with sampling rate of 11025 samples/second the degree of stationarity is 76.9841% as it can be seen in Fig.5(b).

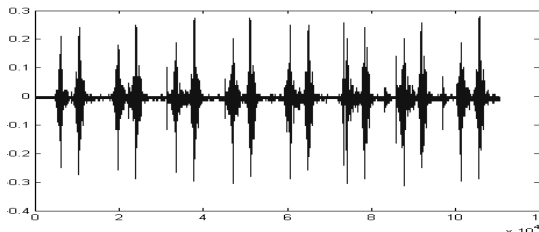


Fig.4(a) Heart signal with sample rate of 22255 samples/second.

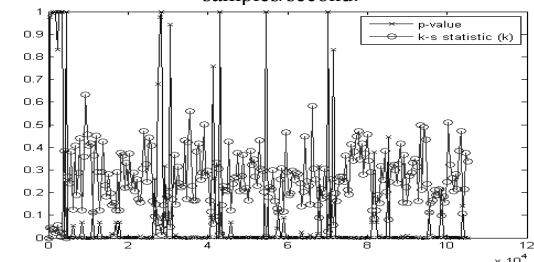


Fig.4(b) Result of K-S two sample test on signal in Fig.4(a)

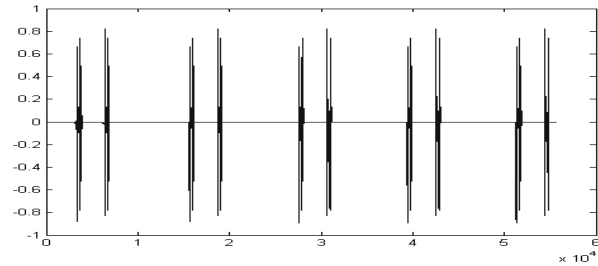


Fig.5(a) Heart signal with sample rate of 11025 samples/second.

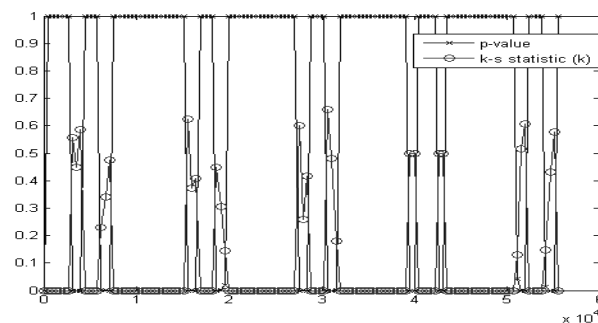


Fig.5(b) Result of K-S two sample test on signal in Fig.5(a)

Method II

In order to test the stationarity length of the data the K-S test was applied to data segments of different lengths. The approach was to divide the available data segments into shorter ones, and then applying the K-S test to each pair of CDFs from the segments. Same signal is used as that was used in Method-1 to check the stationarity length.

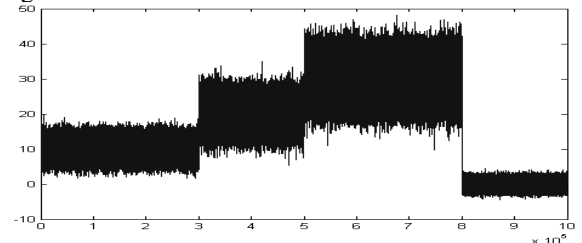


Fig.6(a) Computer generated signal by combining randomly generated normal signals with different mean and standard deviation values shown in TABLE IV ($N_2+N_3+N_4+N_1$)

As we can see in Fig.6(b) p-value (blue line) is greater than mean of k values (pink line) for 3 lakh samples indicating signal is stationary for length of 6.80 seconds. After that p-value is less than mean of k indicating that signal is non-stationary for remaining samples.

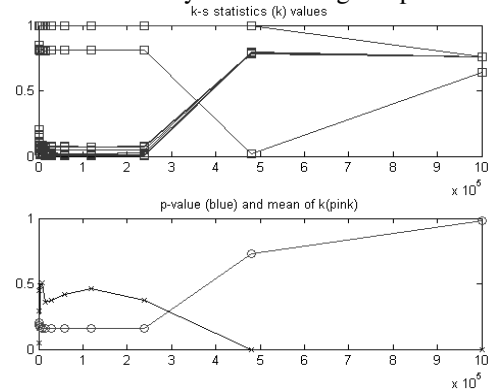


Fig.6(b) Result of K-S two sample test on signal in Fig.6(a)

Stationarity length of heart signal with sampling rate of 22255 samples/second as shown in fig.7(a) is 0.22 second as p-value is greater than mean of k for initial 4900 samples. Similarly, the heart signal with sample rate 11025 samples/second as shown in fig. is stationary for 4.535 seconds where effect of transients for small duration of time was not seen on stationarity length.

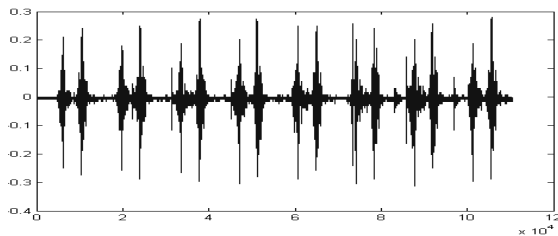


Fig.7(a) Heart signal with sample rate of 22255 samples/second.

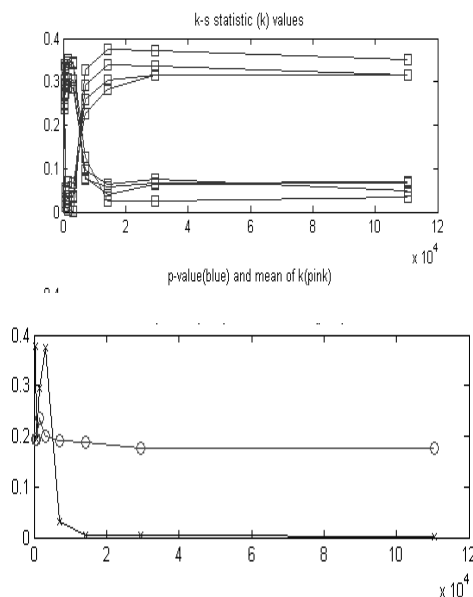


Fig.7(b) Result of K-S two sample test on signal in Fig.7(a)

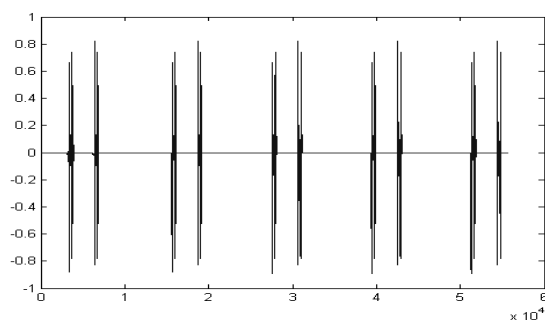


Fig.8(a) Heart signal with sample rate of 11025 samples/second.

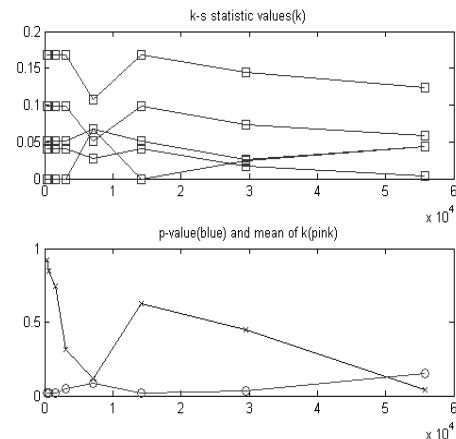


Fig.8(b) Result of K-S two sample test on signal in Fig.8(a)

CONCLUSION

Normality of signals was examined first by using K-S one sample test. Normality of the signals is estimated where the k-s test statistic is less than the threshold given as p-value according to Eq.12. Then the stationarity of signals was examined using K-S two sample test. Degree of stationarity of signals was obtained first in terms of percentage of local stationarity and then in terms of stationarity length. Local stationarity is estimated where the p-value is greater than K-S test statistic and stationarity length is estimated where p-value is greater than mean of K-S test statistics for different pairs of segments. For heart signals it was found that effect of transients of very small durations was not taken into account while estimating stationarity length.

REFERENCES

- [1] W. J. Conover, Practical nonparametric statistics. John Wiley & Sons, Inc., third ed., 1999.
- [2] M. J. Levonen, R. K. Lennartsson, L. Persson, S. McLaughlin, "Stationarity analysis of ambient noise in Baltic sea," published by IEEE in 2006.
- [3] Lisa A. Pflug, Pam M. Jackson, Juliette W. Joup and George E. Joup, "Moment analysis of ambient noise dominated by local shipping," published by IEEE in 1996.
- [4] M.J. Wilmut, N.R. Chapman and S. Anderson, "Statistical properties of horizontal, vertical and omnidirectional underwater noise fields at low frequency," Proceedings of OCEANS'93, vol. 3, pp. 269-273, 1993.
- [5] Lisa A. Pflug, Pamela Jackson, George E. Joup and Juliette W. Joup, "Variability in higher order statistics of measured shallow-water shipping noise," Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistic, pp. 400-404, July 1997.

HAND GESTURE RECOGNITION

D. K. Chitre

(Head of Comp. Dept.)
Terna Engineering College Nerul,
Navi Mumbai, Maharashtra
dvchitre@rediffmail.com

Rakhi A. Kalantri

(Head of Comp. Dept.)
Terna Engineering College Nerul,
Navi Mumbai, Maharashtra
rakhi_Kalantri@rediffmail.com

ABSTRACT

Computer recognition of sign language is an important research problem for enabling communication with hearing impaired people. Many ways of communications are used between human and computer, such as hands, arms, face, head, and/ or body. Hand gesture is one of the typical methods of non-verbal communication for human beings and we naturally use various gestures to express our own intentions in everyday life. Gesture recognizers are supposed to capture and analyze the information transmitted by the hands of a person who communicates in sign language. Many applications of hand gestures are such as medical, communication, learning system, Forensic analysis etc

“HanGeR” is a hand gesture consisting of gesture acquisition, segmentation, filtering, representation, and robust classification. The recognition is based on skin color and would require only a single webcam. To convey a sentence, the user would make the corresponding gestures for the letters, composing the sentence, to a webcam. The output would be the conveyed sentence in text form. The system is possible to designed (A to Z) capital static alphabets as specified in ISL (International Sign Language) i.e. the entire displayed sentence would be in upper case with neither punctuations nor spaces. The system is possible to implemented using Java as frontend and MATLAB as backend for the Microsoft Windows environment allowing real time gesture control of new and existing applications.

The simple interface of “HanGeR” makes it efficient for communication with deaf people without having the need to completely learn the sign language.

Keywords: *Pentium Processor P4, 512 MB RAM, Webcam – 30fps, Operating System: Windows XP, Java Net Beans 6.7, MATLAB.*

I. INTRODUCTION

HUMAN COMPUTER INTERACTION (HCI) is the study of interaction between people (users) and computers. Interaction between users and computers occurs at the user interface (or simply interface), which includes both software and hardware. An important facet of HCI is the securing of user satisfaction. Because human-computer interaction studies a human and a machine in conjunction, it draws from supporting knowledge on both the machine and the human side. On the machine side, techniques in computer graphics, operating systems, programming languages, and development environments are relevant. On the human side, communication theory, graphic and industrial design disciplines, linguistics, social sciences, cognitive psychology, and human factors are relevant. Engineering and design methods are also relevant. Due to the multidisciplinary nature of HCI, people with different backgrounds contribute to its success. HCI is also sometimes referred to as MAN MACHINE INTERACTION (MMI) or computer human interaction.

A. Background

Digital cameras are now integrated into personal computers, mobile cellular devices and hand held computers. These devices usually include a powerful microprocessor, capable of performing millions of computations per second. As microprocessor and digital camera technologies advance, it is now possible to use these resources in order to create new human computer interfaces that are based on recognition of users' gestures. Gesture recognition interfaces can be used as a natural communication channel between humans and machines.

A gesture recognition system is basically an HCI system. It should be based on existing and common hardware such as low-resolution web cams or mobile-integrated cameras. It is also desired that the system will be able to operate under non-uniform background, lighting and noise conditions. Another requirement for the gesture recognition system is to be computationally non-intensive in order to be suitable for real-time classification. To date, most work on sign language

recognition has employed expensive wired “data-gloves” which the user must wear. Glove-based gestural interfaces typically require the user to wear a cumbersome device and carry a load of cables connecting the device to a computer. This hinders the ease and naturalness of the user’s interaction with the computer. Also, these systems limit the user’s freedom of movement.

B. HanGer

“HanGeR” won’t have such considerations as the recognition would be purely based on skin color and would require only a single low-resolution webcam and an algorithm that processes the acquired images and classifies the gestures presented.

Suppose the user wants to convey a sentence using the system, the former would make the corresponding gestures for the letters, composing the sentence, to the webcam. The latter would then display the conveyed sentence. Given an acquired image that contains a hand gesture, the hand gesture recognition algorithm performs two stages. The first is the pre-processing phase, where the hand shape and other distinguishable features are extracted from the image using noise reduction, filtering and several other image processing techniques. The second stage involves the classification of the features set to the appropriate gesture. It would consist of a sign language tutor wherein the user would be able to view the gesture for a particular alphabet.

HanGeR is robust and operates well on several different backgrounds, lighting and moderate noise conditions and it is suitable for both real-time and offline classification.

C. Objectives of HanGer

To develop an efficient and robust hand gesture recognition system that can translate static hand gestures as defined in International Sign Language (ISL) to their corresponding alphabets and display the result in text form.

II. EXISTING SYSTEM

In the existing scenario, a hand gesture recognition system based on Neural network classification has been developed which has a modular architecture consisting of a nearest-neighbor classifier to recognize International Sign Language hand gestures. The system consists of a training window, a practice window and a synthesis window. Fig 2.1 shows a snapshot of the existing system [11].

A. System Design

1. Image Acquisition : The image is acquired using a camera. Two gloves are used which assist in the segmentation process.
2. Image Segmentation : Segmentation is done in the HSV space for its robustness.
3. Feature extraction : Feature vectors are developed for the segmented images.
4. Image Classification : Classification is done using neural networks and a nearest neighbor classifier.

The recognition rate was 94% for trained signers.

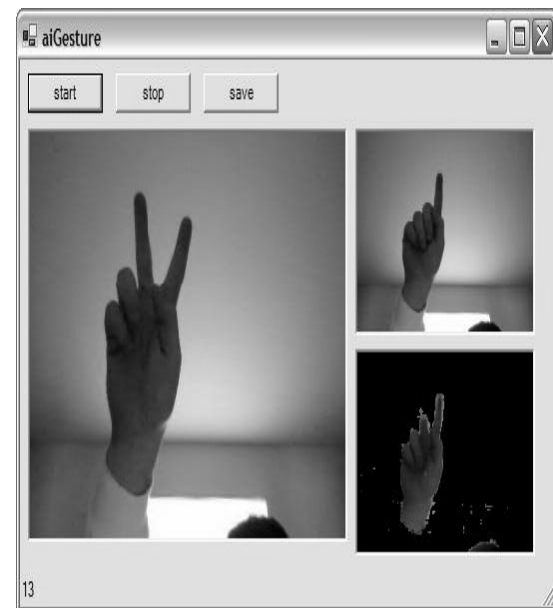


Fig 2.1: Snapshot of existing system

B. SLARTI

SLARTI sign language recognition system, developed at the University of Tasmania consists of multiple feature-recognition neural networks and a nearest-neighbour classifier to recognise Australian sign language (Auslan) hand gestures.

Input Hardware: The specific input devices used in developing SLARTI were a Cyber Glove, and a Polhemus IsoTrak. The Cyber Glove measures the degree of flexing of the various joints of the hand and wrist. The Polhemus allows tracking of the spatial position and orientation of the hand with respect to a fixed electro-magnetic source.

- **Classification:** The task of transforming a stream of input data directly into a classification of the sign being performed by SLARTI is an extremely difficult one. Instead the approach taken within SLARTI was to initially process the input data so as to produce a description of this sequence in terms of the four features.

- The SLARTI system consists of four separate feature-extraction neural networks each trained specifically to recognise one of the features of the sign. The feature vector produced by these networks is then used to perform the overall classification of the input sequence.

III. PROPOSED SYSTEM

Issues Resolved By HanGeR

- No need of cumbersome data gloves, cybergloves; HanGeR does not employ either data gloves or cybergloves which pose an overhead by impeding extra expenses for the device; additional wiring has to be handled and makes the system cumbersome.
- Skin based segmentation: HanGeR employs skin based segmentation which is very efficient as compared to other segmentation approaches such as difference based segmentation and marker based segmentation. These segmentation approaches are rather based on static environments and require a clear background image be obtained before segmentation can begin. Also, these techniques are not robust to changes in image illumination.
- Hidden Markov Model: HanGeR employs HMM for classification as neural networks pose many disadvantages like rigorous training to operate, requires high processing time. The property of HMM to compensate time and amplitude variances of signals has been proven for speech and character recognition.
- Provision of a tutor: Proposed system provides user with a tutor which will help the user to make the gestures. Also a demo video is provided which guides the user to use the system.

The proposed methodology will be developed under the following assumptions:

- All gestures are static postures.
- Each gesture cannot represent more than one command, and each command must be expressed by exactly one gesture.
- Intuitiveness will be based on a small empirical experiment.
- The system shall be developed for gestures recognized in ISL (International Sign Language), as shown in Figure 3.1[12]

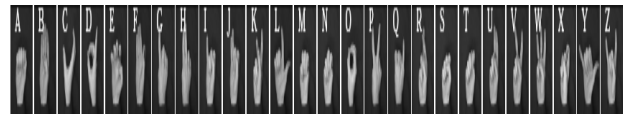


Fig .3.1: The gesture set to be recognized

A. New Features added by HanGeR

- The system has an inbuilt tutor that demonstrates the use of the system. The tutor has dual functionality. First, it shows a video which shall demonstrate how to present a gesture to the installed camera. Secondly, the various gestured defined in ISL are also included to assist the user.
- The system will consist of two modes:
 - Novice mode: wherein the user will just present the gesture and its output will be displayed.
 - Expert mode: wherein the user shall also be shown the histogram
 - The acquired and the segmented image for comparative study.

Pictorial representation of the entire process Figure 3.2

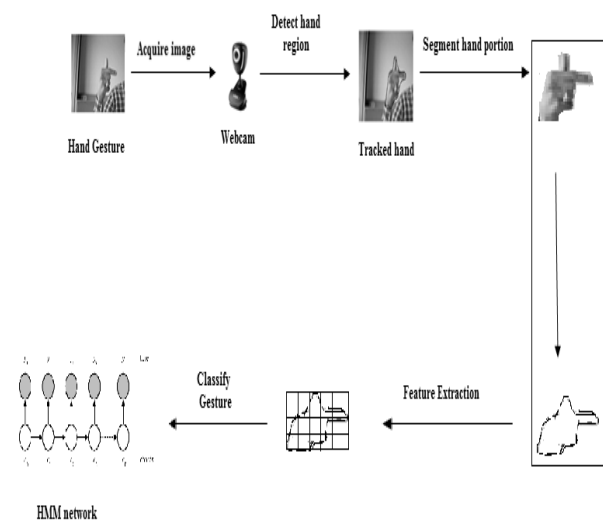


Fig 3.2: Pictorial representation of the proposed system

IV. SYSTEM DESIGN

Hardware Requirements

- Pentium Processor P4; compatible with newer versions.
- 512 MB RAM.
- Webcam – 30fps.

Software Requirements:

- Operating System: Windows XP, Windows XP SP2, Windows XP SP3, Windows Vista or Windows 7.
- Visual Studio 2008 Express Edition.
- OpenCV 2.1 – Image processing library

V. ARCHITECTURE OF HANGER

The generic architecture of HanGeR (Figure 5.1) could be explained as below:

- Image Acquisition**
 This module performs the preliminary work of image acquisition. The module starts with first tracking the hand. Upon tracking and finding the hand in a still position, an image is acquired for the input gesture. The image, which may be either colour or gray scale, is taken from a single camera.
- Image Preprocessing**
 The captured image is first smoothened and filtered to remove any noise that may have been incorporated in to the image during the acquisition phase. The filtered image, which is in the RGB domain, is then mapped into its corresponding Black and White representation.
- Image Segmentation**
 The preprocessed image is then put to a segmentation process wherein only the part of the image comprising of the hand i.e. the fingers, the palm and the fist is extracted and the remaining background is eliminated from the image. Skin color segmentation is employed. The next phase will have the segmented image as it input.
- Feature Extraction**
 After hand region is segmented out, its feature needs to be understood. The feature here corresponds to the position of the fingers, the orientation of the palm and the side of the hand facing the camera. To do this, feature extraction algorithm can be used. The shape of each gesture can easily be detected using edges. All the features, once detected will be used to represent feature representation vector into feature space. Each distinguishable gesture will be recognized by its unique feature vector.
- Classification System**
 The final module corresponds to the phase wherein the gesture gets classified into one of the 27 classes (26 classes for the 26 alphabets and one class for space). Hidden Markov Model (HMM) is used for classification purpose. The training set for the model consists of the numerous images taken from different individuals for the same gestures. Based upon the training, any relevant gesture that is provided to the system gets uniquely classified into the character corresponding for that gesture.

- Display of results**
 The display of output would be the textual representation in the window pane. As the letters get recognized, the corresponding alphabet is displayed. The final output is the conveyed sentence.

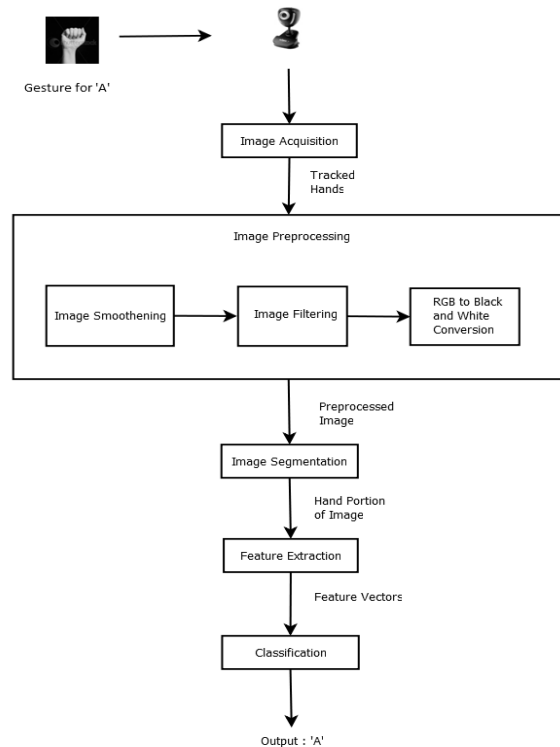


Fig 5.1: Architecture of HanGeR.

VI. USER MENU SCREENSHOTS

The project can be divided into four parts:

- The user tests for gestures.
- The user records a video to be viewed and analyzed later on.
- The user views a pre-recorded video to obtain the gestures in the video.
- A tutor



Figure 6.1: Main user window



Figure 6.2: Start Capture from File menu

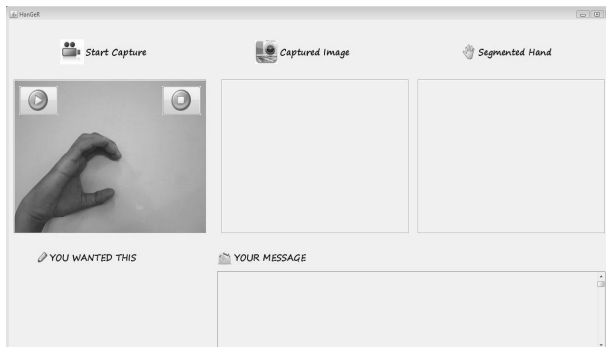


Figure 6.3: Before pressing 'play' button

CONCLUSION

We proposed HanGeR- a hand gesture recognition system for static gestures mentioned in International Sign Language (ISL) using hand position and vector for the features. The various related work in the field of Sign Language Recognition, an arena encompassing hand gesture recognition, were discussed and some of the limitations in the existing system, which we intend to rectify using HanGeR, were highlighted. The hardware and software requirements for this software run on a normal desktop PC is minimal.

The system was successfully implemented for capturing gestures and the additional feature of extracting gestures from video.

REFERENCES

- [1] S. Ahmad, "A usable real-time 3d hand tracker," presented in 'Signals, Systems and Computers', 1994.
- [2] Y. Azoz, L. Devi, and R. Sharma, "Reliable tracking of human arm dynamics by multiple cue integration and constraint fusion", presented in 'Computer Vision and Pattern Recognition', 1998.
- [3] R. Cipolla and N. J. Hollinghurst, "A human-robot interface using pointing with uncalibrated stereo vision," presented in 'Cipolla and Pentland', 1995.

[4] R. Cutler and M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," presented in 'Automatic Face and Gesture Recognition Conference', 1998.

[5] J. Davis and M. Shah, "Visual gesture recognition," presented in 'IEEE Proceedings - Vision, Image and Signal Processing', vol. 141, no. 2, 1994.

[6] Chen Q Georganas, , and Petriu, E. "Real-time vision-based hand gesture recognition using haar-like features" presented in 'Instrumentation and Measurement Technology Conference Proceedings', 2007.

[7] Perng, J., Fisher, B., Hollar, S., and Pister, K. "Acceleration sensing glove (asg)" in 'Wearable Computers', 1999.

[8] Yamato, J., Ohya, J., and Ishii, K. "Recognizing human action in time-sequential images using hidden markov model" presented at 'Computer Vision and Pattern Recognition Proceedings CVPR '92'.

[10] Ana Kuzmanic and Vlasta Zanchi. "Hand shape classification using DTW and LCSS as similarity measures for vision-based gesture".

[11] Lee, C., and Xu, Y. "Online, interactive learning of gestures for human/robot interfaces" presented at 'IEEE International Conference on Robotics and Automation ', 1996.

[12] www.moveinstitute.org/~kolsch/HandVu/HandVu.html

OVERVIEW OF IMPULSE-BASED UWB TRANSMITTER RECEIVER

Archana

M.E Student , Dept of Extc,
K.J.Somaiya College of Engineering
Mumbai
archana3008@gmail.com

Swati S. Mahajan

Dept. of Extc
K.J.Somaiya College of Engineering
Mumbai
swatimahajan@engg.somaiya.edu

ABSTRACT

Ultra-wideband (UWB) radio is an emerging technology in WPAN wireless systems that have attracted a great deal of interest from academia, industries, and global standardization bodies. The IEEE 802.15.3a (TG3a) and IEEE 802.15.4a (TG4a) are two task groups (TGs) within 802.15 working group (WG) that develop their standards based on UWB technology. This paper gives an overview of Impulse –based UWB transmitter and receiver. It encompasses the adaptability of a reconfigurable digital receiver. To complete the analysis we introduce a transmitter and receiver of pulse based ultra wideband modulation. Direct sequence spread spectrum has become the modulation method of choice for wireless local area networks, and personal communication systems, because of its numerous advantages.

Keywords- *Ultra-wideband (UWB), direct sequence spread spectrum (DSSS), PN sequence.*

I. INTRODUCTION

UWB technology has been around since 1960, when it was mainly used for radar and military applications. One of the most promising commercial application areas for UWB technology is the very high data rate wireless connectivity of different home electronic devices at low cost and low power consumption. Ultra-wideband technology offers a solution for sharing the bandwidth resource and physical size requirements of next generation consumer electronic devices. In addition, UWB promises low susceptibility to multipath fading, high transmission security and simple design.

The approval of UWB technology was made by the Federal Communications Commission (FCC) of the United States in 2002, which reserves the unlicensed frequency band between 3.1 to 10.6 GHz (7.5 GHz) for indoor UWB wireless communication systems. Industrial standards such as IEEE 802.15.3a (high-data rate) and IEEE 802.15.4a (very low-data rate) have been introduced based on UWB technology.

UWB communications are fundamentally different from all other communication techniques because it employs extremely narrow radio frequency (RF) pulses, which are generated from the UWB pulse generator, to communicate between transmitters and receivers. Utilizing short-duration pulses as the building blocks for communications directly generates a very wide bandwidth.

The data signal, being transmitted is spread onto a much larger range of frequencies (RF bandwidth) using a specific encoding Scheme. This encoding scheme is known as a Pseudo-noise Sequence or PN sequence. Direct sequence Spread spectrum has become the modulation method of choice for wireless Local area networks and personal communication systems. Direct-sequence spread-spectrum transmissions multiply the data being transmitted by PN signal.

In hardware implementations for data demodulation purpose

a sampling bridge circuit and a matched filter are frequently used. The primary disadvantage of the sampling bridge circuit is the loss of data rate, as it under samples the incoming pulse with a sliding sampling position. Therefore, multiple pulses per bit are required to transmit data, limiting the maximum data rate. To perform data demodulation using a matched filter in hardware, an estimated received pulse must be generated at the receiver and correlated with the incoming waveform. A system with this architecture must have a highly stable oscillator and precise timing accuracy [1].

Recently there has been a burst of research about UWB; hence more and more papers are being published. However, many papers have been found on the transceiver circuit description for UWB with different technology but here we propose a system model of Impulse based UWB transmitter and receiver with Direct Sequence Spread Spectrum technology.

The rest of the paper is organized as follows. In Section II we study the signals and system model. Section III deals; Impulse Based UWB transmitter. In

section IV Impulse based UWB receiver is discussed. Section V deals with the readings for the proposed model. Section VI concludes the paper.

II. SIGNALS AND SYSTEM MODEL

For a single user system, the continuous transmitted data stream is written as

$$s(t) = \sum_{k=-\infty}^{\infty} d(k) \cdot p(t - k \cdot T_s)$$

Where $d(k)$ are stationary uncorrelated BPSK data and T_s is the symbol duration. The UWB pulse $p(t)$ has duration T_{UWB} ($T_{\text{UWB}} < T_s$). The models used in this paper are the models proposed by IEEE 802.15.3a study group [2]. The impulse response can be written as

$$h(t) = \sum_{p=0}^M h_p \delta(t - \tau_p)$$

where the parameter M is the total number of paths in the channel.

III. IMPULSE BASED UWB TRANSMITTER

Transmitter consists of PN Sequence generator which generates a sequence of pseudo random binary numbers using a linear feedback shift registers. XOR is used for mixing data. Unit delay is used for delayed data and oscilloscopes are placed along the path for display purposes. PN Sequence generator is used both as generating message sequence and a sequence for spreading process. When differentially encoding an incoming message, each input data bit must be delayed until the next one arrives. The delayed data bit is then mixed with the next incoming data bit. The output of the mixer gives the difference of the incoming data bit and the delayed data bit. The differentially encoded data is then spread by a high-speed pseudo noise sequence (PN). This spreading process assigns each data bit its own unique code, allowing only a receiver with the same spreading to disperse the encoded data. The 63-bit pseudo noise sequences (PN) used in this paper are generated by a 6th order maximal length sequence shown in equation [3]

$$g(x) = x^6 + x^2 + 1$$

The following table indicates two sets of parameter values

that correspond to a generator polynomial of the 63-bit

pseudo noise sequences (PN) used in this paper for generating 6th order maximal length sequence

Quantity	Example 1	Example 2
Generator polynomial	[1100001]	[6 5 0]
Degree of generator polynomial	6	6
Initial states	[110000]	[110000]

The Shift parameter shifts the starting point of the output sequence. The parameter is described in greater detail below.

It can be possible to shift the starting point of the PN sequence with the Shift parameter. Specification of parameter in either of two ways

- An integer representing the length of the shift.
- A binary vector, called the mask vector, whose length is equal to the degree of the generator polynomial

The maximal length spreading sequence uses a much wider bandwidth than the encoded data bitstream, which causes the spread sequence to have a much lower power spectral density [4]. The transmitted signal is given by

$$x(t) = m(t) \cdot c(t)$$

Where $m(t)$ is the differentially encoded data, and $c(t)$ is the 63 chip PN spreading code.

For recovering of message sequence, we XOR the modulated signal with same type of 63-bit pseudo noise sequences (PN). Here we also use a unit delay to find the original signal. The signal recovering process is successfully done with some propagation delay which was obvious because of some noise & losses. Fig 1 shows the impulse based UWB transmitter.

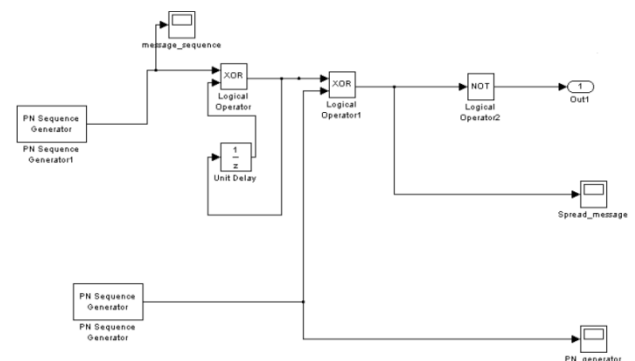


Fig 1. Impulse based UWB Transmitter

IV. IMPULSE BASED UWB RECEIVER

The receiving signal is modulated by Bi-phase modulation technique. Then signal is split into two parallel paths and fed into two identical matched filters with the input to one having a delay of 126

chips. The probability of bit error for Bi – phase modulation assuming matched filter reception is given by [3]

$$P_e = Q\left(\sqrt{\frac{2Eb}{N_0}}\right)$$

The B PSK modulation technique is mathematically described as

$$w(t) = \sum_{j=-\infty}^{\infty} b_j s(t - jT_f)$$

The receiver is shown in figure 2.

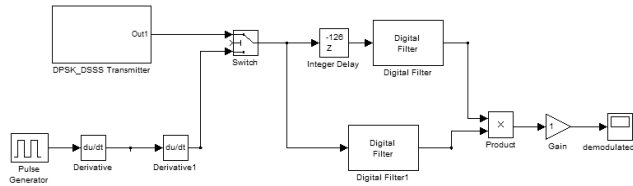


Fig 2: Simulink model of UWB Receiver

One of the benefits of Bi-phase modulation is its ability to eliminate spectral lines due to the changes in pulse polarity.

This aspect minimizes the amount of interference of the conventional radio systems. A decrease in the overall transmitted power could also be attained using Bi-phase modulation, making this a popular technique in UWB systems.

The received waveform is sampled to allow for multiple samples per pulse period. Over sampling gives a more accurate representation of the pulse shape, which then produces better results using a digital matched filter.

Correlation processing, another form of matched filtering, is often used in the digital domain. The method for calculating the correlation output is given by:

$$g(k) = \sum_{r=1}^{N-1} r(t) \cdot h(t)$$

Where,

$g(k)$ = the resulting correlation value

k = k^{th} pulse period

N = the number of samples in one pulse width

$r(t)$ = the received sampled waveform

$h(t)$ = the known pulse waveform

One of the primary drawbacks of the matched filter receiver topology is the lack of knowledge of the pulse shape at the receiver due to distortion in the channel. There are number of ways to correct this problem, including an adaptive digital equalizer or matching a template by storing multiple pulse shapes at the receiver. A more accurate approach is to estimate the pulse shape from the pilot pulses, which will experience the same channel distortion as the

data pulses. This estimation technique is a promising solution to UWB pulse distortion. The outputs of the two matched filters are denoted by x_1 and x_2 are given by

$$X_1 = d(t - t_o) R_c(t)$$

$$X_2 = d(t - t_o - T_b) R_c(t - T_b)$$

Where

T_b = data bit period..

$R_c(t)$ = is the autocorrelation function of the 63-chip pseudorandom sequence.

The two outputs of the matched filters are then mixed and then low pass filtered and the original message is recovered.

The simulations are performed using matlab [4].

Fig. 3 shows the original input message sequence that is generated from a PN sequence generator

V. READINGS FOR PROPOSED MODEL

Following the analytical approach presented in section III and IV, we evaluate the simulation result of impulse radio UWB technology.

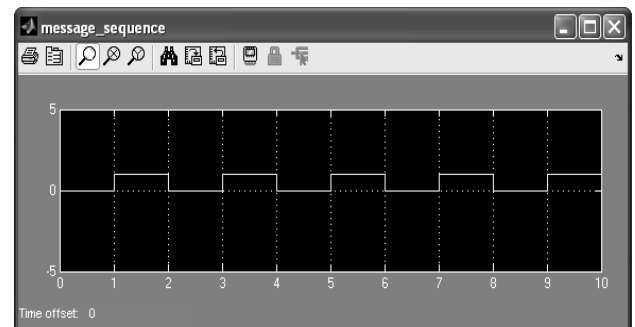


Fig 3. Original transmitted message signal.

The incoming messages are differentially encoded by using mixer and unit delay. Eventually the mixer will give the difference of the incoming data bit and the delayed data bit.

The differentially encoded data is then spread by a high-speed 63-bit pseudo noise (PN) Sequence generator which is generated by a 6th order maximal length sequence. Here, 63 bit pseudo noise sequences (PN) have been used for generating both the message and high frequency carrier signal. This spreading process assigns each data bit its own unique code which is shown in Fig. 4, allowing only a receiver with the same spreading to disperse the encoded data.

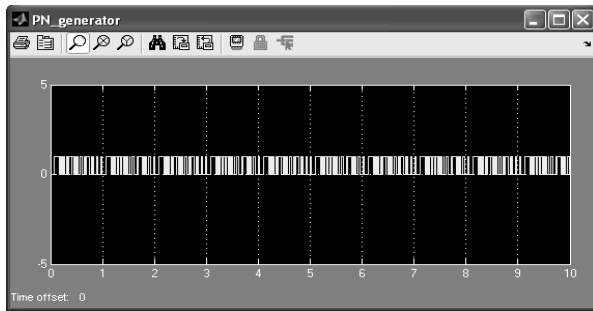


Fig 4 Output waveforms of Simulink UWB Transmitter

In the receiving part of UWB , the modulated signal has been dispread using same type of 63-bit pseudo noise sequence.

Before dispreading, the received signal is split into two parallel paths and fed into two identical matched filters. Among the two split signal, one is the spreading message signal and another is the Bi-phase modulated signal.

Fig.5 shows the received signal after dispreading and Fig. 6 shows the recovered output signal.

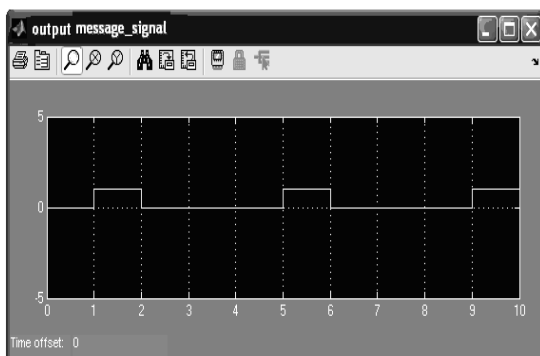


Fig 5: Received Signal into UWB Receiver after Dispsreading

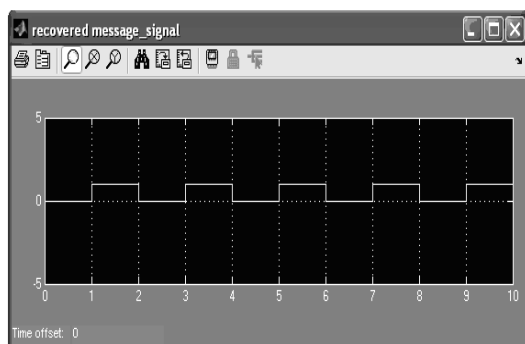


Fig 6: Original recovered output signal

CONCLUSION

The main key facts of impulse based ultra wide band transmitter is, it consists PN Sequence generator which generates a sequence of pseudo random binary numbers using a linear feedback shift registers and the matched filter. The 63-bit pseudo noise sequences (PN) used in this paper is generated by a 6th order maximal length sequence has been used for generating both the message and high frequency carrier signal. The results of the system simulations are performed using MATLAB which gives a solid foundation for the design as a whole.

REFERENCES

- [1] Weihua Zhuang¹, Xuemin (Sherman) Shen¹ and Qi B "Ultra-wideband wireless communications" Wirel. Commun. Mob. Comput. 2003.
- [2] Bikramaditya Das and Susmita Das, "Suppression of Narrowband Interference in impulse radio based high data rate UWB WPAN communication system using NLOS Channel Model " World Academy of Science, Engineering and Technology, 2011.
- [3] Bobby Barua, "Analysis the Performance of Transceiver Circuits for Pulse Based Ultra Wideband" *IACSIT International Journal of Engineering and Technology*, February 2012.
- [4] S.M. Nabritt, M.Qahwash, M.A. Belkerdid, "Simulink Simulation of a Direct Sequence Spread Spectrum Differential Phase Shift Keying SAW Correlator", Electrical and Comp. Engr. Dept, University of Central Florida, Orlando FL 32816, Wireless Personal Communications, The Kluwer International Series in Engineering and Computer Science, 2000, Volume 536, VI, 239-249
- [5] Matlab, Version 11 , The Mathworks, Inc., Natick, MA.
- [6] L. W. Couch II, Digital and Analog Communication Systems, 6th Edition, New Jersey: Prentice H

STUDY OF AES DESIGNED ON FPGA USING PIPELINED PARTIAL ROLLING TO ACHIEVE HIGH THROUGHPUT

Mukulika Aniket Hardas

Department of Electronics and
Telecommunication VIT,
Wadala, Mumbai, Maharashtra, India
mlikabdg@gmail.com

Ranjana Gite

Department of Electronics and
Telecommunication
VIT, Wadala, Mumbai,
Maharashtra, India

Dattatray Bade

Department of Electronics and
Telecommunication
VIT, Wadala, Mumbai,
Maharashtra, India

ABSTRACT

Considering the increasing importance of the data security in communication, this paper addresses design, FPGA implementation and performance of AES algorithm. An optimized code for the Rijndael algorithm with 128-bit keys has been developed. [1] The area and throughput are carefully traded off to make it suitable for wireless military communication and mobile telephony where emphasis is on the speed as well as on area of implementation.

In this paper, a pipelined partial rolling (PPR) architecture for the AES encryption for FPGA is discussed. Using the proposed architecture on the Altera Stratix EP1S20F780C5 FPGA, significant throughputs are achieved using designs of S-Modules M4Ks.[2] Architecture presented uses memory modules (i.e., Dual-Port RAMs) of Field-Programmable Gate Array (FPGAs) for storing all the results of the fixed operations (i.e., Look-Up Table).[3].

Paper is divided in two sections. First section introduces Rijndael algorithm and FPGA. Section two introduces architecture of Pipelined Partial Rolling (PPR).

Keywords- AES, FPGA, Encryption, Decryption, DES, PPR.

SECTION – I

1. INTRODUCTION

Several techniques have been developed to keep data secure, private, and protected. However, the need for secure transactions in ecommerce, private networks, and secure messaging is a challenge to the encryption system.

Advanced encryption standard (AES) was issued as Federal Information Processing Standards (FIPS) by National Institute of Standards and Technology (NIST) to overcome drawbacks of DES algorithm. The hardware based implementation of AES Rijndael Algorithm is required because it is secure and consumes much less power than a software implementation. [4]

As the physical development is done by various methods depending on where it is being used, inform of computer software such as C++, java, or in micro controller using an assembly. FPGA, in form of hardware to work in real time, can be used in various applications.

Among all the study reports of the finalist of the AES algorithm, Rijndael FPGA study reports that the implementation of both encryption and decryption takes about 60% more space than the implementation of encryption alone.

A. The AES Algorithm

The Rijndael algorithm (AES) is based on arithmetic in a finite Galois field $GF(2^8)$ and considers only operations using a 128 bit cipher key and 128-bit data blocks. [5]. The algorithm can encrypt the data to an unintelligible form called cipher-text and decrypt the same information back into its original form, which is called plaintext. [6].

This is a symmetric block cipher algorithm i.e. encryption and decryption being done by same key and faster to execute than asymmetric key algorithms. AES may have different key lengths and number of rounds as shown in table 1.

Key Length (in bits)	Key Length (in words)	Number of Rounds
AES-128	4	10
AES-192	6	12
AES-256	8	14

Table 1: Key length and the number of rounds

Algorithm is parted in three parts as Encryption, Key scheduling and Decryption.

Encryption and decryption are series of well-defined steps that can be followed as a procedure, as shown in figure 1.

1. Encryption and Decryption

Encryption process has two inputs viz. state and cipher key. A data block to be encrypted by AES is split into an array of bytes called as state and the cipher key is a sequence of 128,192 or 256 bits. The Encryption or

Decryption operations are conducted on the State array. The input to the encryption (and decryption) algorithm is a single 128-bit block.

It consists of a number of different transformations applied consecutively over the data block bits, in a fixed number of iterations, called rounds. The number of rounds depends on the length of the key used for the encryption process as referred in table 1. Figure 1 represents an overview of the Encryption process.

Steps in Encryption are SubByte, ShiftRows, MixColumns and AddRoundKey

- SubBytes** - The bytes substitution transformation is a non-linear substitution of bytes that operates independently on each byte of the State using a substitution table (Sbox) to perform a byte-by-byte substitution of the block.
- ShiftRows** - This is a simple permutation. Bytes in the last three rows of the State are cyclically shifted over different numbers of bytes (offsets). The first row is not shifted.
- MixColumns** - A mixing operation which operates on the columns of the state, combining the four bytes in each column using a linear transformation. This transformation makes use of arithmetic over GF (2^8).
- AddRoundKey** - Each byte of the state is combined with the round key; each round key is derived from the cipher key using a key schedule.

The decryption structure has exactly the same sequence of transformation as that in the encryption structure but in reverse direction. Hence the last round values of both the data and key are first round inputs for the Decryption process and follows in decreasing order. AddRoundKey is the same for both encryption and decryption. However the three other functions have inverses used in the decryption process: Inverse SubBytes, Inverse ShiftRows, and Inverse MixColumns.

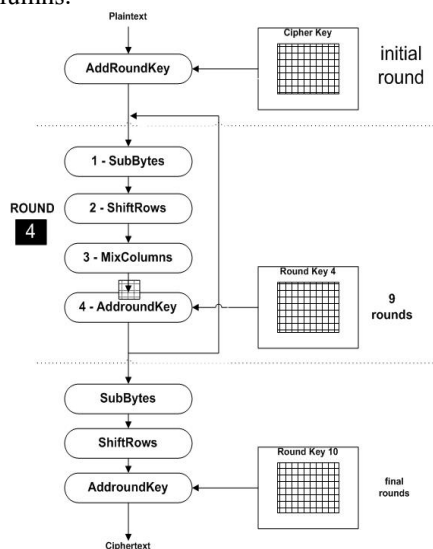


Figure 1 Encryption Process

2. Key Scheduling

Key scheduling is the expansion of given cipher key into 11 partial keys, used in the encryption system i.e. in the initial round, further 9 main rounds and the final rounds as shown in figure 1.

Operations performed here are Rotation, substitution and EX-ORing.

The rotate operation takes a 32-bit word like (in hexadecimal) 1D 2C 3A 4F and rotates it eight bits to the left such that the high eight bits "wrap around" and become the low eight bits of the result 2C 3A 4F 1D. Rcon is the exponentiation of "2" to a user-specified value as referred from Rijndael documentation calls.

Block from the plaintext message which is to be encrypted is given as an input to the encryption process. This along with the cipher key is further passed through nine rounds of transformation. Finally to the last round which outputs as cipher text.

B. FPGA

Field programmable gate arrays (FPGAs) are digital integrated circuits (ICs) that contain configurable (programmable) blocks of logic along with configurable interconnects between these blocks.

PLDs (programmable logic devices) contain a relatively limited number of logic gates, and the functions they can be used to implement are much smaller and simpler.

ASICs (*application-specific integrated circuits*) offers limitations in size, complexity, and performance; also designing and building one is an extremely time-consuming and expensive process, with the added disadvantage that the final design is "frozen in silicon" and cannot be modified without creating a new version of the device.

On the other hand, FPGAs functionality can be customized in the field like PLDs, but they can contain millions of logic gates and be used to implement extremely large and complex functions that previously could be realized only using ASICs. And therefore it is a midway solution for PLDs and ASICs. Various AES implementations using ASICs or FPGAs have been reported since 2001. The work is focussed on the small chip area or on high throughput by using the rolled or unrolled architecture. [7]

SECTION II

A pipelined partial rolling (PPR) architecture is to achieve a high throughput with small area. It performs partial-rolling in the round while adopting the pipeline technique.

A. AES Architecture

Among the various architectures, the rolling architecture and the unrolling architecture are shown in Figure 2(a) and (b) respectively.

The **rolling architecture** shown in Figure 2(a) uses a feedback structure where the data are iteratively transformed by the round functions. This approach occupies small area, but achieves low throughput. Existing rolling implementations have the throughput of approximately 1 to 1.4 G-bit/s, and the size of the memory for the core is just 32 K bits.

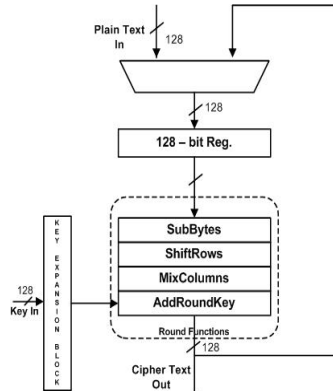


Figure 2a: Rolling architecture of AES

In the **unrolling architecture** shown in Figure 2(b), the round blocks are pipelined and the inserted pipeline registers allows simultaneous operation of all 11 round blocks. Due to the pipeline, this approach achieves a high throughput, but requires large area. Existing unrolling implementations have the throughput of approximately 10 to 23 G-bit/s, and the size of the memory for the core is up to 320 K bits.

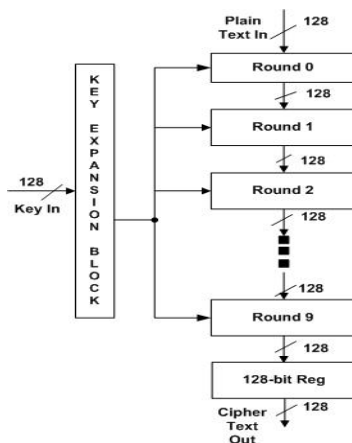


Figure 2b: Unrolling architecture of AES

B. Pipelined Partial Rolling (PPR)

PPR is a memory-based architecture in which ROM is used to realise S-box. In AES algorithm, among the four transformations Subbyte is the most complicated transformation as it uses 16 S-Boxes for rolling and unrolling implementation. Also it requires high area and latency. Using PPR number of S-boxes to be used in the AES rounds reduces.

Matrix multiplication here uses X-OR operation and because Mixcolumns and AddRound key includes matrix multiplication, we can easily implemented XOR operations in the above two transformation.

Architecture:

Figure 2b represents the architecture of PPR. It has two parts viz. Key expansion block and ten round functions. The key expansion block generates the key for each round which is 128 bit data, further processed in the round functions.

Each round function has following functional blocks, as shown in figure 3.

1. Rolling part
2. Pipeline Register
3. Mix columns
4. 128-bit XOR

Rolling part performs shift rows and substitution of bytes transformation. It is the most important block in the round functions. These transformations can be performed in two ways as using 4S-Module and 8 S-Module. S-Modules of the 4SM are used four times, while the S-Modules of the 8SM are used twice. Four 4SM and two 8SM are used in each round function.

4SM has 16 byte cyclic shifter which has 16 byte input and 4-byte output. 4-bit output signals from the control block are also given as an input to the shifter. Output of the cyclic shifter is given to 4 S Modules. Each S-Module consists of a 2K-bit ROM and a 32-bit feedback register, where the ROM stores the table for the S-Box, and the feed-back register stores the outputs of the S-Box.

8SM has 16 byte cyclic shifter which has 16 byte input and 8-byte output. Instead of 16 byte cyclic shifter, blocks of permutation and selector are used to arrange the input data in the required sequence. But the use of 16 byte cyclic shifter produces higher memory efficiency. Table 2 show the permutation, where both inputs and outputs are 16-byte data. As in 4SM, output signals from the control block decides the selection of upper bytes or lower bytes of the selector as shown in table 3.

Controller Output	Input to the selector
0	Upper bytes of the selector
1	lower bytes of the selector

Table 3: Selection of input bytes of selector in 8SM

128 bit output from the rolling part is stored in the pipeline register. States of each round can be obtained from the register. Further Mixcolumns and add round key is performed with the data. Round key is the output of key expansion block as discussed above.

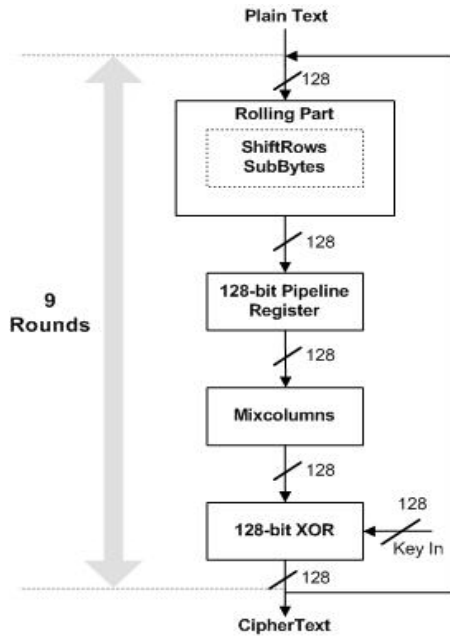


Figure 3: Round Function in PPR

4. FPGA Implementation

We use Altera Stratix, which is the first-generation of Altera which offers M4K. They are special RAM blocks ideal for applications requiring medium-sized memory. M4K can be configured between (4096×1) : (256×16) . These blocks can have dual port functionality and suitable for synchronous ROMs.

Inputs	X ₀	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
Outputs	X ₀	X ₁₀	X ₄	X ₁₄	X ₈	X ₂	X ₁₂	X ₆	X ₅	X ₁₅	X ₉	X ₃	X ₁₃	X ₇	X ₁	X ₁₁

Table 2: Relationship between the inputs and the outputs of the permutation network in 8SM

REFERENCES

- [1] An Area-Throughput Efficient FPGA implementation of Block Cipher AES algorithm by Banraplang Jyrwa ECE, Department, NIT Jalandhar, India and Roy Paily ECE, Department, IIT Guwahati, India.
- [2] An FPGA Design of AES Encryption Circuit with 128-bit Keys by Hui QIN, Tsutomu SASAO and Yukihiro IGUCHI
- [3] AES Encryption Algorithm Hardware Implementation Architecture: Resource and Execution Time Optimization, by Samir El Adib, Naoufal Raissouni
- [4] W. Diffie and H. Hellman, "Privacy and authentication: An Introduction to cryptography", Proceedings of IEEE, pp 397-427, vol 67, 1979.
Mukulika Aniket Hardas, Prof. Ranjana Gite and Prof. Dattatray Bade, "Review of AES for Data Security on Various Development Platforms", International Journal of Information and Computation Technology, pp 287-290, Volume 2, Number 3 (2012).
- [5] FPGA Implementation of AES Algorithm by Mr. Atul M. Borkar, Dr. R. V. Kshirsagar and Mrs. M. V. Vyawahare
- [6] Wikipedia "Rijndael Key Schedule", http://en.wikipedia.org/wiki/Rijndael_key_schedule
- [7] "The Design Warrior's Guide to FPGAs" by Clive "Max" Maxfield.

To implement two S-boxes, M4K can be configured as a dual-port synchronous (256×8) bit words ROM and therefore 8 copies of the M4K are sufficient for each SubByte. Values in LUT (Look Up Table) are loaded into M4K at the configuration time.

In round function, MixColumns and AddRound key are realized as network of EX-OR which requires Logic Element (LE). Also in rolling part, 16-bit cyclic shifter for the 4SM and selector for 8SM are also implemented by LEs.

CONCLUSION

In this paper, an efficient FPGA implementation of 128 bit block is presented with the pipelined partial rolling (PPR) architecture for an AES encryption processor. We implemented two different designs: AES-4SM and AES-8SM in rolling part in the architecture on Altera Stratix EP1S20F780C5 FPGA. The AES-4SM achieves a throughput of 5.61 Gbps by using 20 M4Ks, and the AES-8SM achieves a throughput of 10.49 Gbps by using 40 M4Ks. Compared with the unrolling implementation that achieves a throughput of 20.48 Gbps by using 80 M4Ks on the same FPGA, the AES-4SM and the AES-8SM improve the memory efficiency by 9.6% and 2.4%, respectively, and reduce the amount of the memory by 75% and 50%, respectively. The PPR architecture fills the gap between unrolling and rolling architectures, and fits on less expensive FPGAs.

SPEED CONTROL MECHANISM USING TERRAIN DETECTION

Ankit Nanavaty

Dept. of Computer,
MCT's Rajiv Gandhi Inst. of Tech.,
Mumbai, India
ankit.n24@gmail.com

Ankit Kandoi

Dept. of Computer,
MCT's Rajiv Gandhi Inst. of Tech.,
Mumbai, India
ankitkandoi@gmail.com

Ruchi Patel

Dept. of Computer,
MCT's Rajiv Gandhi Inst. of Tech.,
Mumbai, India
ruchi_spatel@yahoo.com

Swapnil Gharat

Dept. of Computer,
MCT's Rajiv Gandhi Inst. of Tech.,
Mumbai, India
swapnilgharat2000@gmail.com

ABSTRACT

Autonomous navigation in an outdoor environment requires the ability to discriminate among different types of terrain. A non-trivial problem is to manage the robot's speed based on terrain roughness. This paper presents a speed control system for a robotic platform travelling over natural terrain. This will ensure safe navigation of the robot on an unknown terrain.

This phenomenon is widely used in the area of planetary robots where the surface texture is completely unknown. A video camera will be mounted on the robot, with a similar perspective to the driver, which takes the video of the road, with different classes of textures when the car is in motion. These textures (loose stones, grass, ground, concrete, asphalt, slopes) will then be processed using Image Processing Algorithms. Based on the results obtained after applying the algorithms, the velocity estimations are done and the speed of the robot changes accordingly.

Keywords- *wavelet transform; openness; DEM*

I. INTRODUCTION

Navigation with vehicles in unknown environments is an easy task for humans, but very difficult for robots. For autonomous navigation in outdoor terrains, the robot must be endowed with the ability to interpret data acquired from its environment so as to plan and follow trajectories from its current location to the desired location, considering the features of the terrains.

Most autonomous navigation works have focused on the problem of recognition and avoidance of obstacles; however, so far, almost no attention has been placed on the recognition of terrains' textures and irregularities, and how they accept the performance and safety of robots during navigation. Autonomous off-road robots will be employed in military operations, and also in civilian applications such as wide-area environment monitoring, disaster recovering, search-and-rescue activities, as well as planetary exploration. [1]. There is a strong push in the Army to move towards autonomous and semi-autonomous vehicles to perform tasks which

may be too cumbersome or dangerous for human driven vehicles to perform. Unmanned ground vehicles (UGVs) can potentially be used to find and disarm improvised explosive devices (IEDs) without risking the lives of soldiers [2].

The right navigating velocity regarding the terrain features is a requisite to keep robots away from risks of falling and sliding due to surface slopes or smash textures, the vehicle velocity must be updated according to the terrain features in order to guarantee robots safe navigation. For wheeled-robots navigation on different terrains, it is necessary to have required information about the surface features. One of those features on which this work focuses is the surface roughness on which the vehicle is moving. The terrain is the principle source of chassis excitation in off-road vehicles and the control of the vehicle is dependent on effectively characterizing the terrain slope, roughness, and surface condition [2]. The robot's velocity during real navigation depends on the terrain roughness; moreover, the robot integrity depends on the right velocity of the robot while navigating through terrain.

The information of environment needs to be quickly and accurately processed by the robot's navigation systems for a right displacing. The velocity setting must be performed remotely, i.e. the robot must detect and classify soft irregularities and textures before the robot passes on them, so that the robot disposes enough time to react and set its velocity according to the terrains' characteristics [3]. On the other hand, by extending the robots' abilities to recognize the terrains' features and then, accordingly, update the velocity while navigating, the usage of the robots' energy and computing resources would improve as well as the autonomy would be strengthened.

During navigation, the robot can move on irregularities under the following criteria; if irregularities are soft (slopes with inclination less than or equal to 15 degrees), then the robot can move over them; otherwise, the non-soft irregularities (slopes with inclination bigger than 15 degrees) are considered as slopes, slope detection algorithms need to be applied and velocity of the robot must be adjusted accordingly.

II. PROPOSED SPEED CONTROL SYSTEM

A. Flow Chart

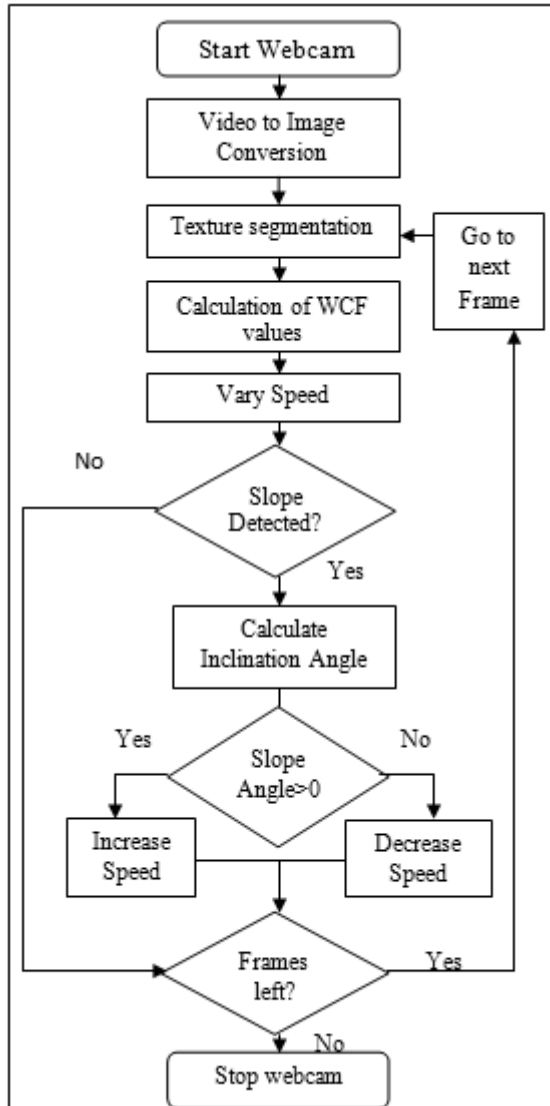


Figure 1. Flow of the Speed Control System

B. Algorithm

- A wireless camera is mounted on the robot (vehicle) which captures the video and transmits it the Processing end (PC).
- This video is converted into image frames for further processing.
- The first frame is taken and the Texture Segmentation Algorithm (Texture Segmentation using Wavelet Transform) is applied. Thus we obtain various segments of the processed image.
- The Wavelet Co-occurrence Feature (WCF) values of each segment in the image, is calculated. WCF values give the features of the segments such as contrast, cluster, shade and cluster prominence and the relationship between them [5]. According to these WCF values, the speed of the robot is adjusted.
- If the WCF value is greater than the maximum threshold, it implies maximum roughness, and thus we vary the robot speed to the minimum value for safe navigation. If the WCF value is less than the minimum threshold, it implies minimum roughness, and thus we vary the robot speed to the maximum value for safe navigation.
- Along with surface roughness, any slope if exists, in the path then it is also detected. If a slope is detected we calculate the angle of inclination of the slope.
- If the inclination angle is negative, implies that the inclination is downward, and thus speed of the robot is decreased. If the inclination angle is positive, implies that the inclination is upward, and thus robot speed is increased.
- Hence, the robot motion takes place with the calculated speed, where the change in speed is calculated considering both, the surface texture and the slope.
- This process is continued till all the frames are processed and the speed of the robot is detected for the whole path.

III. DISCRETE WAVELET TRANSFORM

A. Definition

Discrete Wavelet Transform (DWT) is performed in the frequency domain, where the input image is decomposed to different frequency levels using the Discrete Wavelet Frames (DWF). Few statistical methods have been proposed in the past for texture analysis. Inherent disadvantages with those approaches, such as increased computational cost and irreversibility, can be eliminated using the wavelet transform [4].

Wavelets are functions generated from one single function W by dilations and translations. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of wavelets [5]. Any such superposition decomposes the given function into different scale levels where each level is further decomposed with a resolution adapted to that level. The discrete wavelet transform (DWT) is identical to a hierarchical sub band system where the sub-bands are logarithmically spaced in frequency and represent octave-band decomposition. By applying DWT, the image is actually divided i.e., decomposed into four sub-bands and critically sub-sampled [5]. These four sub-bands arise from separate applications of vertical and horizontal filters. The sub-bands labeled LH1, HL1 and HH1 represent the finest scale wavelet coefficients i.e., detail images while the sub-band LL1 corresponds to coarse level coefficients i.e., approximation image. To obtain the next coarse level of wavelet coefficients, the sub-band LL1 alone is further decomposed and critically sampled. This results in two-level wavelet decomposition. Similarly, to obtain further

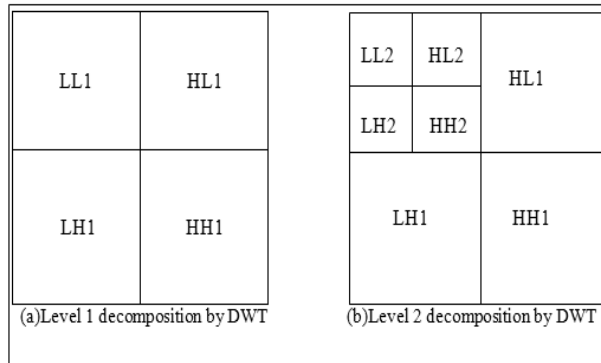


Figure 2. Discrete Wavelet Transform

decomposition, LL2 will be used. This process continues until some final scale is reached. The values or transformed coefficients in approximation and detail images (sub-band images) are the essential features, which are as useful for texture discrimination and segmentation. Since textures, either micro or macro, have non-uniform gray level variations, they are statistically characterized by the values in the DWT transformed sub band images or the features derived from these sub-band images or their combinations. In other words, the features derived from these approximation and detail sub-band images uniquely characterize a texture. The features obtained from these DWT transformed images give the segmented images.

B. Advantages over Fourier Transform

Although the Fourier transform has been the mainstay of transform-based image processing since the late 1950s, a more recent transformation, called the wavelet transform, is now making it even easier to compress, transmit, and analyze many images. Unlike the Fourier transform, whose basis functions are sinusoids, wavelet transforms are based on small waves, called wavelets, of varying frequency and limited duration [6]. A wavelet is a waveform that is bounded in both frequency and duration [10]. This allows them to provide the equivalent of a musical score for an image, revealing not only what notes (or frequencies) to play but also when to play them. Conventional Fourier transforms, on the other hand, provide only notes or frequency information; temporal information is lost in the transformation process. Most real world signals (such as music or images) have a finite duration and abrupt changes in frequency. This makes wavelet transform more efficient [10].

IV. TEXTURE SEGMENTATION

The study of terrain vehicle interaction has been a topic which has been researched for quite some time. There is a vast amount of literature addressing various aspects of modeling the interface between the tire and the terrain. The term terrain segmentation has been used rather loosely in the literature. Researchers have used it to refer to understanding the mechanical properties of the terrain called terra-mechanics. Determining how the terrain properties affect a vehicle's ability to traverse it, which

has also been referred to as traffic ability. It has also been used to describe the process of classifying the type of terrain (i.e. sand, dirt, gravel). One will also see the term terrain segmentation used for larger scales than relevant for vehicle dynamic studies, such as in geological surveys or aerial vehicle mapping. Many of these works use Digital Elevation Models (DEM) to represent the terrain. The focus of this section will be to present the literature as it relates to the development and testing of ground vehicle systems. Thus, the literature presented here will include terrain segmentation as it relates to mechanical properties, small ground robots, and larger ground vehicles.

Here, texture mosaic images of size $N \times N$ are considered [4]. The analysis is carried out by considering sub images (i.e., block) of size 32×32 . Each 32×32 sub-image, taken from top left corner of the original image, is decomposed using one level DWT and co-occurrence matrices (C) are derived for sub-image. Fig. 2 (a) shows the level 1 decomposition (i.e., LL1, LL2, LL3, LL4) and fig. 2 (b) shows the level 2 decomposition with detail sub-bands (i.e., LL2, LH2, HL2 & HH2 sub-bands) of wavelet decomposed sub-image. Then, from these co-occurrence matrices (C), significant WCFs, such as contrast, cluster shade and cluster prominence, are computed using formulae given, as texture features [3]. In our implementation, the contrast feature values, calculated over all the blocks, are subjected to linear normalization in the scale of 0–255, while the cluster shade and cluster prominence features, which found to have very large dynamic ranges, are subjected to logarithmic normalization in the scale of 0–255 for computation [5].

A. Segmentation Algorithm

Input : Texture mosaic image of size $N \times N$.

Output: Texture segmented image.

- Step 1.** Read the texture mosaic image.
- Step 2.** Obtain 32×32 sub-image blocks, starting from the top left corner.
- Step 3.** Decompose sub-image blocks using 2-D DWT.
- Step 4.** Derive co-occurrence matrices (C) for original image, and detail sub-bands of DWT decomposed sub image blocks.
- Step 5.** Calculate WCFs such as contrast, cluster shade and cluster prominence from co-occurrence matrices.
- Step 6.** Calculate the difference between sums of WCFs of adjacent sub-image blocks. This results in segmentation band.
- Step 7.** Apply disk filtering and thresholding techniques to remove noise like artifacts if any in the segmentation band.
- Step 8.** Apply skeleton extraction algorithm to get thinned or segmented line of one pixel thickness.

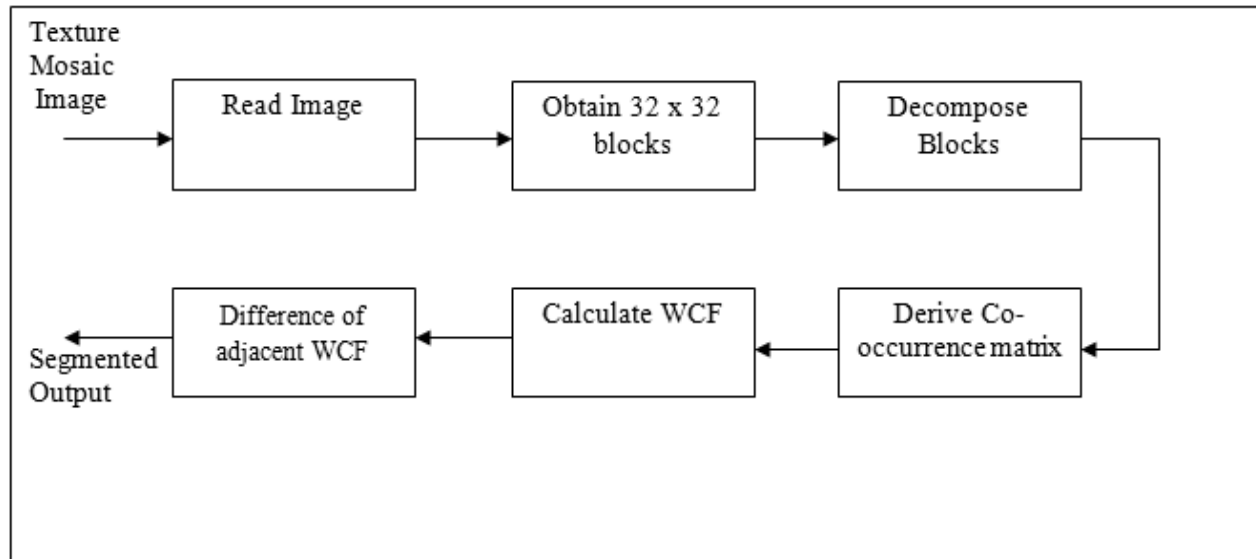


Figure 3. Flow Chart of Texture Segmentation Algorithm

Fig. 3 shows the Flowchart of the Texture Segmentation Algorithm. The texture segmentation is carried out by comparing the normalized co-occurrence features of discrete wavelet transformed adjacent but overlapping 32×32 sub-image blocks, both horizontally and vertically [4]. Each successive block is differ from the previous one in its spatial location by one column or one row, depending on whether the successive block is taken in horizontal or vertical direction, respectively. Here, the sum of the above normalized features of one block is compared with the corresponding sum of features derived from the next block. The difference in feature values is less when successive blocks belong to the same texture region and it increases in the texture border region while it is high when the successive blocks are from two different texture regions.

By carrying out the above block by block feature comparison both in horizontal and vertical directions; a segmentation band is formed across the texture boundaries [11]. When the difference in feature values within the same texture region is high, noise like artifacts or spurious spots appear in the segmented image. This spurious spots are removed by applying disk filtering and thresholding techniques (i.e., post processing). Then, the post processed segmented band is thinned using skeleton extraction algorithm to get segmented line of one pixel thickness [11]. The thinned result gives the line of demarcation among the different textures present in the image i.e., thinned lines are exactly aligned with texture boundaries.

B. Advantages over Other Algorithms

More recently, methods based on multi-resolution or multi-channel analysis, such as Gabor filters and wavelet transform, have received a lot of attention. But, the outputs of Gabor filter banks are not mutually

orthogonal, which may result in a significant correlation between texture features [5]. Finally, these transformations are usually not reversible, which limits their applicability for texture synthesis.

Most of these problems can be avoided if one uses the wavelet transform, which provides a precise and unifying frame work for the analysis and characterization of a signal at different scales (e.g., Unser, 1995). Another advantage of wavelet transform over Gabor filter is that the low pass and high pass filters used in the wavelet transform remain the same between two consecutive scales while the Gabor approach requires filters of different parameters. In other words, Gabor filters require proper tuning of filter parameters at different scales.

One advantage of the wavelet transform over the histogram processing is that in many cases, a large number of the detail coefficients turn out to be very small in magnitude. Truncating these small coefficients from the representation introduces only small errors in the reconstructed signal. We can approximate the original data distribution effectively by keeping only the most significant coefficients [7].

V. SLOPE DETECTION

Autonomous navigation in cross-country environments presents many new challenges with respect to more traditional, urban environments. The lack of highly structured components in the scene complicates the design of even basic functionalities such as slope detection. In addition to the terrain typing, geometric description of the scene is also an important component of the perceptual system.

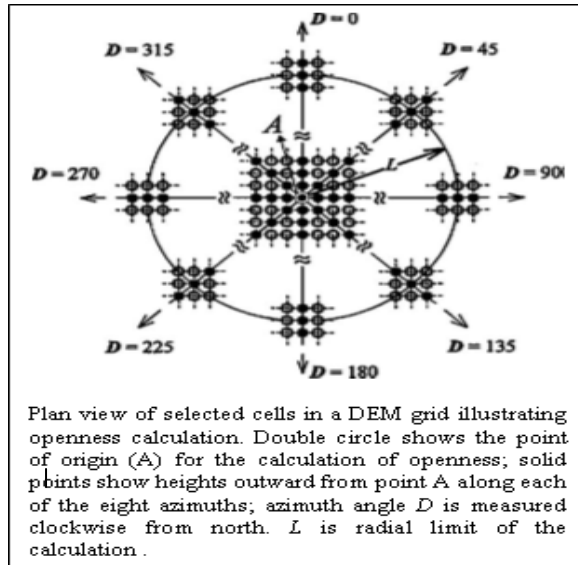


Figure 4. Plan view of selected cells in a DEM grid

A. Openness

The new image-processing technique named *openness*, actually calculates the surface openness from any DEM (digital elevation model) file and projects it into map form to be used to accurately visualize landscape variation [8]. These product maps appear as simple digital images of shaded relief and slope angles, but in actuality, they emphasize dominate surface concavities and convexities of the landscape. For angles less than 90 degrees, openness is equivalent to the internal angle of a cone, its apex at a DEM location, constrained by neighboring elevations within a specified radial distance [9]. Terrain Openness is essentially an angular measure of the relation between surface relief and horizontal distance. It incorporates the terrain line-of-sight principle but does not require light source and is calculated from zenith and nadir angles along eight azimuths. Openness has two viewer perspectives. Positive values, expressing openness above the surface, whereas negative values describe this attribute below the surface. Openness values are mapped by gray-scale tones [9].

The calculation of openness at point A in DEM grid with spatial scale L is illustrated in Figs. 4-5 and involves the following steps :

- 1) For each azimuth direction D ($D = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, and 315°), calculate the elevation angles along the profile from point A (double circle in Fig. 4 and big black dot in Fig. 5) out to length scale L . In Figure 5, there is an elevation angle for each grid point along the profile represented by the small black dot. The elevation angle is positive if the distant point is higher than A; negative if distant point is lower than A. These angles form a set DSL for each azimuth direction D .

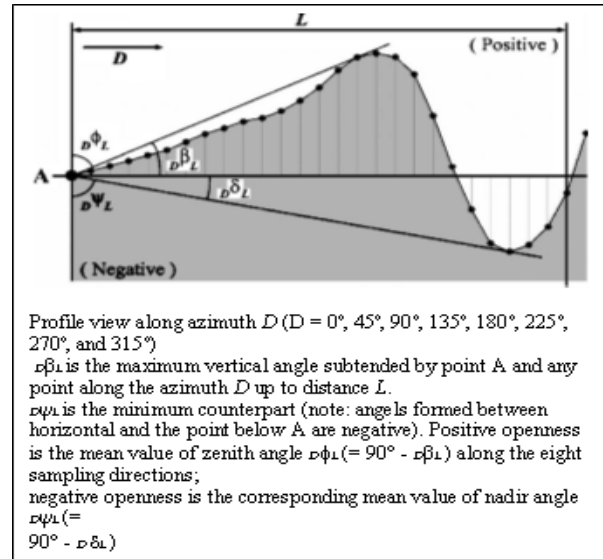


Figure 5. Profile view along azimuth D

- 2) calculate maximum elevation angle:
 $D\beta L = \max(DSL)$;
- 3) calculate minimum elevation angle:
 $D\delta L = \min(DSL)$;
- 4) calculate zenith angle: $D\phi L = 90^\circ - D\beta L$;
- 5) calculate nadir angle: $D\psi L = 90^\circ - D\delta L$;
- 6) obtain positive openness:
 $\phi L = (0\phi L + 45\phi L + \dots + 315\phi L)/8$
- 7) obtain negative openness:
 $\psi L = (0\psi L + 45\psi L + \dots + 315\psi L)/8$

VI. VELOCITY ESTIMATION

It is proposed to imitate the human perception by employing a neuro-fuzzy approach for velocity updating. It is explained how a neuro-fuzzy approach can be a useful tool for the velocity updating purpose [2]. Hence, to imitate the human experience for terrain recognition and the corresponding velocity adjustments when driving, the choice of an adequate method for recognizing the surface average appearance, without lost in unnecessary terrain details, is clever for outdoors navigation. The ABV method, having a low of detailed texture recognition efficacy is used for texture segmentation.

In Earth exploration missions, where human life could be in danger, autonomous rovers are required for explosive landmines search, deep sea exploration, or to determine the eruption risk when exploring active volcano craters, as well others [2]. In these dangerous circumstances, the high autonomy of robots strengthens the robotic support to human safety. On the other hand, by extending the robots' abilities to recognize the terrains' features and then, accordingly, update the velocity while navigating, the usage of the robots' energy and computing resources would improve as well as the autonomy would be strengthened.

VII. CONCLUSIONS

- In this project, for texture segmentation, the concept of discrete wavelet transform is presented for applying to textured images for decomposing them into detail and approximation regions.
- Co-occurrence features, computed out of the wavelet decomposed images, are used for texture segmentation.
- The idea behind this proposed method is to exhibit the usage of co-occurrence features computed from discrete wavelet transformed images (i.e., WCF) for texture segmentation.
- The features are approximately the same when the windows or sub images considered are from the same texture and different if they are from different textures.
- Varieties of textures, collected from standard album, are stitched to form target images which are used for experimentation and it is found that the proposed method yield better results than the texture spectrum method, a single resolution technique
- Appearance-Based recognition algorithms successfully classify terrain textures by regarding the average appearance.
- The algorithms are computationally inexpensive and easy implementation.
- The low computational cost to process the information acquired from the environment allows establishing that a car, at a certain speed, can have enough time to react to the roughness changes of the terrain.
- The wavelet based algorithm for wheeled-robots allows a meta-classification of textures and soft irregularities used for the terrains' roughness recognition.
- The wavelet based algorithm allows velocity updating by outdoor terrains navigation.
- Using slope detection, the slope of the surface is recognized and accordingly the speed of the robot is adjusted to enable safe movement of robot on the inclined surfaces.
- The average appearance of the terrains' textures, according to experimental results, is the requisite for velocity updating purpose on outdoor terrains navigation.
- By applying robot velocity updating by regarding the terrains' roughness, the experimental results show the precision improvement.
- Whenever the velocity is updated by regarding the terrains features, the precision gets improved to achieve the goal position. In addition, the risk that the robot has an accident also decreases.
- Thus, by using the wavelet based texture recognition method and slope detection, we can ensure safer navigation of the robots on an unknown terrain.

ACKNOWLEDGMENT

We are very grateful towards Mr. Swapnil Gharat, our guide and mentor, for his valuable suggestions and input towards the betterment of this paper. We would like to express our gratitude towards Prof. S.B. Wankhade, HoD, Computer Engineering for his assistance and guidance. We would also like to thank Prof. Dr. Udhav Bhosle, Principle, Rajiv Gandhi Institute of Technology, for providing us with the wonderful opportunity.

REFERENCES

- [1] "Reactive speed control system based on terrain roughness detection" by Mattia Castelnovi, Laboratorio Dist., University of Genova; Ronald Arkin, College of Computing Georgia Institute of Technology; Thomas R Collins, School of ECE/GTRI, Georgia Institute of Technology.
- [2] "Terrain characterization and roughness estimation for simulation and control of unmanned ground vehicles" by Jeremy James Dawkins -December 12, 2011.
- [3] "Wheeled-robot's velocity updating and odometry based localization by navigating on outdoor terrains" by M. en C. Farid García Lamont. Thesis advisor: José Matías Alvarado Mentado-November 2010.
- [4] "Unsupervised texture segmentation using discrete wavelet frames" by S. Liapis, N. Alvertos, and G. Tziritas Institute of Computer Science - FORTH, and, Department of Computer Science, University of Crete, P.O. Box 1470, Heraklion, Greece.
- [5] "Texture segmentation using wavelet transform" by S. Arivazhagan, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Amathur (P.O.), Sivakasi 626 005, Tamil Nadu, India; L. Ganesan, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Amathur (P.O.), Sivakasi 626 005, Tamil Nadu, India, 4 February 2003; received in revised form 31 July 2003.
- [6] "Digital image processing", Second Edition by Rafael C. Gonzalez and Richard E. Woods.
- [7] "Wavelet-based histograms for selectivity estimation" by Yossi Matias, Department of Computer Science Tel Aviv University, Israel; Jerey Scott Vittery Department of Computer Science, Duke University; Min Wangz Department of Computer Science, Duke University.
- [8] "Optimizing surface openness of dissection patterns at various digital elevation model resolutions", URAP Program: Fall 2010 by Josh Page, Dr. Wei Luo.
- [9] "Visualizing topography by openness: a new application of image processing to digital elevation" by Models Ryuzo Yokoyama, Michio Shirasawa, and Richard J. Pike.
- [10] "Design of feature extraction in content based image retrieval (CBIR) using color and texture", Swati V. Sakhare & Vrushali G. Nasre Dept. of Electronics Engg., Bapurao Deshmukh College of Engg., Sevagram, Wardha (India).
- [11] "Wavelet based image segmentation", Andrea Gavlasov'a, Ale's Proch'azka, and Martina Mudrov'a Institute of Chemical Technology, Department of Computing and Control Engineering.

ROUTE MAINTENANCE SCHEMES BASED ON NODE LIFE PREDICTION FOR TWO REACTIVE ROUTING PROTOCOLS IN MOBILE AD HOC NETWORKS

Ravindra Eklarkar

Department of Electronics and
Communication Engineering
Gurunanak Dev Engn. College, Bidar
reklarker@gmail.com

Vinay Dutta. V. Kohir

Department of Electronics and
Communication Engineering
P.D.A College of Engineering, Gulbarga
vvhahir@yahoo.com

V. D. Mythri

Principal
Guru Nanak Dev Engineering
College, Bidar
vdmytri@yahoo.com

ABSTRACT

In mobile Ad Hoc networks, a multi hop route breaks mainly because of the failure of a link or adjacent links caused by relative movement of one or more nodes, or malfunction and power depletion of one node. When this happens route repair is typically performed to establish a new route. Many route repair algorithms are developed to solve the problem. Such route repair mechanisms affects network performance which reduces packet delivery ratio and increases packet delay. In this paper, a new idea on Route maintenance mechanism based on Node life time prediction is proposed which increases the network Lifetime and hence increases packet delivery ratio. Simulation results show that our mechanism leads to significant performance improvement.

I. INTRODUCTION

A Mobile Ad-Hoc Network(MANET) is an infrastructure-less collection of mobile nodes that can arbitrarily change their geographic locations such that these networks have dynamic topologies which are composed of bandwidth constrained wireless links [1] Ad-Hoc network has certain characteristics, which imposes new demands on the routing protocol. The most important characteristic is the dynamic topology, which is a consequence of node mobility. Nodes can change position quite frequently, which means that we need a routing protocol that quickly adapts to topology changes. Many routing protocols that can be divided into table-driven and on-demand have been proposed for ad-hoc networks. Table driven routing protocols [2] attempt to maintain consistent, up-to-date routing information from each node to every other node in the network, even if nodes have no packet to send. Examples of table driven protocols include OLSR [3] , DSDV [4] . In contrast, in the on-demand routing protocol is [5] , each node builds and maintains only the route needed for a source destination pair without periodic route table update exchanges. Examples of on-

Demand protocols include AODV [6], DSR [7] and TORA [8] . On-Demand routing protocols are more suitable for MANET than table-driven because of its characteristics of dynamic topology, limited bandwidth and power constrain. on-demand routing protocols consists of two phases: route discovery and route maintenance. Node mobility, malfunction, and power depletion may lead to breakage of existing routes. A route consists of a sequence of links. Even if only one link in the sequence fails, the route is no longer works. Once a route fails, consequent route maintenance is triggered. The initiator of the route update process attempts to use flooding-search to obtain other path to the destination. This will result in the decrease of the network throughput as well as long delays. The problem worsens when mobility is high. With the increase of the network size, wireless bandwidth wastes and delay grows substantially because of the longer way between the initiator and the destination. In the recent years, researches on routing protocols of ad-hoc networks are mainly on the strategy of route discovery and those on route maintenance are relatively lower. Traditional route maintenance schemes do not give importance to the uncertainty of node status. In real network, nodes can enter or leave the network freely. It may result from intentional power off or energy exhaustion. It may be due to software or hardware failures, too. The changes of the node status are the major reasons why the link breaks. However all the study and simulations simply assume that all nodes without exception, keep working from beginning to end without stoppage. Researchers use cross-layer mechanisms, backup routes or prediction of link breakages to enhance route reliability [9] [10], to make routing protocols more robust against link breakages. Performance of routing protocol is improved by optimizing the algorithm on the routing layer is given in [11] [12] Generally when the link breaks between the node and the next hop node, the next one hop node or next two hop node still has an active route to the destination. The next one hop and two hop node is closer to the repairing node than the destination node By broadcasting RREQ on a small scale, we can establish a route from the repairing node to its next one hop or two

hop node. This is equal to the successful route setup from the repairing node to the destination node. So the improvements could reduce route setup time and control overhead. In Local repair of [11], a link repair packet is broadcasted from a downstream node where the link breaks to an upstream node. This leaves out the process of the upstream node broadcasting RREQ and achieves the purposes of reducing time for route setup and reducing control overhead. The local repair scheme of [12] is based on keeping the next two-hop node. When the node discovers a link breakage in an active route, it established the route by broadcasting RREQ with the next two-hop node as its route destinations and bypassing the next hop node. This reduces the route setup time and the overhead of control packets to some extent. The disadvantage is that if the failure of software and hardware or power-off of a downstream node results in the link breakage, the upstream nodes cannot receive link repair packet. In this case, the repair efficiency is rather low and it will be longer to set up a route. One of the main design Constraints in mobile ad-hoc networks is that they are power constrained. Hence every effort is to be channeled towards reducing power. More precisely, network life time is a key design metric in MANETs. Since every node has to perform the functions of router, if some nodes die early due to lack of energy, it will not be possible for other nodes to communicate with each other. Hence the network will get disconnected and network lifetime will be adversely affected. In this paper we propose a new idea on Route maintenance mechanism based on life time prediction which improves the network performance. Simulation results show that our mechanism leads to significant performance improvement.

II. NODE LIFE TIME PREDICTION ROUTING

If there are two nodes that have the same energy level, an active node that is used in many data forwarding paths consumes energy more quickly, and thus, it has shorter lifetime than the remaining inactive node. We evaluate the node lifetime that is based on its current energy and its past activity. We proposed two new Route maintenance Schemes Node lifetime prediction Routing in Dynamic Source Routing protocol(NLPRDSR) and Node Life time Prediction Routing in Ad Hoc on Demand Distance vector Routing protocol(NLPRAODV). The two schemes uses battery lifetime prediction. The objective of these routing protocols is to extend the service life of MANET with dynamic topology. This protocol favors the path whose lifetime is maximum. We represent our objective function as follows Where Each node tries to estimate its battery lifetime based on its past activity using Simple Moving Average (SMA) predictor by keeping track of the last N values of residual energy and the corresponding time instances for the last N packets received/relayed by each mobile node. This information

is recorded and stored in each node. We have carefully compared the predicted lifetimes based on the SMA approach to the actual lifetimes for different values of N. our motivation in using lifetime prediction is that mobility introduces different dynamics into the network. Since mobility in MANET can change the traffic patterns through the node, which thereby affects the rate of depletion of its battery. We chose to employ lifetime prediction as recent history is a good indicator of the traffic through the node. This approach avoids the power congested nodes and chooses paths that are lightly loaded. NLPRAoDV is an on Demand Distance vector Routing Protocol.

III. DSR BACKGROUND

A. Route Discovery

In DSR, activity begins with the source node flooding the network with RREQ packets when it has data to send. An intermediate node broadcasts the RREQ unless it gets the path to the destination from its cache, or it has previously broadcasts the same RREQ packet. Consequently, intermediate nodes forward only the first received RREQ packet. The destination node only replies to the first arrived RREQ since that packet tends to take the shortest path. In NLPRDSR, all nodes except the destination calculate their predicted life time, Tl and replace the min lifetime in the header with Tl if Tl is lower than the existing min lifetime value in the header.

$$T_i(t) = \frac{E_r, i(t)}{\frac{1}{N-1} \sum_{k=i-N+1}^i R_k(t)} \quad (1)$$

where:

$E_r, i(t)$: remaining energy at the ith packet is being sent or relayed through the current node

$R_k(t)$: rate of energy depletion of the current node when the Kth packet was sent and is calculated by the ratio of the difference between residual energies of the nodes for packets k-i and k and the difference between arrival times of these two packets.

N : length of the history used for calculating the SMA When an intermediate node receives a RREQ packet, it starts a timer(Tr) and keeps the min lifetime in the header of that packet as Min-Lifetime. If additional RREQs arrives with the same destination and sequence number, the cost of the newly arrived RREQ packet is compared to the Min-cost. If the new packet has a lower cost Min-cost is changed to this new value and the new RREQ packet is forwarded. Otherwise, the new RREQ packet is dropped. In NLPR, the destination waits for a threshold number(Tr) of seconds after the first RREQ packet arrives. During that time, the destination examines the cost of the route of every arrived RREQ packet. When the timer (Tr) expires, the destination node selects the route with the minimum cost and replies. Subsequently, it will drop any received RREQs. The reply also contains the cost of the selected path appended to it. Every node that hears this route reply

adds this route along with its cost to its route cache table. Although this scheme can somewhat increase the latency of the data transfer, it results in a significant power saving. NLPR has a route invalidation timer that invalidates old routes. This helps in removing old routes. This also avoids over usage of particular routes in cases of low mobility.

B. Route maintenance

Route maintenance is needed for two reasons: Mobility, Connections between some nodes on the path are lost due to their movement. In this case, a new RREQ is sent out and the entry in the route cache corresponding to the node that has moved out of range is purged. Change in predicted lifetime. In this case, Once the route is established, the weakest node in the path monitors the decrease in its battery lifetime. When this remaining lifetime decrease goes beyond a threshold level, the node sends a route error back to the destination as if the route was rendered invalid. The destination sends this route error message to the source. This route error message forces the source to initiate route discovery again. This decision is only dependent on the remaining battery capacity of the current node and its discharge rate in the short history. Hence it is a local decision. NLPRDSR adopts this local approach because this approach minimizes control traffic. Node i generates a route error at time t when the following condition is met:

$$T_i(t_0) - T_i(t) \geq \delta \quad (2)$$

Where

T : predicted life time of node i

t : current time

t_0 : route discovery time

S : change threshold for life time prediction

IV. AODV BACKGROUND

A. Route Discovery

When a source node needs a route to communicate with some destination node, it would broadcast a route request packet(RREQ) to the destination node just as in case that the route to the destination is not in the routing table. Every downstream node creates or updates a reverse route to source node when it receives the RREQ, and then continuously broadcasts the RREQ. An intermediate node that has a valid route to the specified destination, or the destination node itself, is allowed to send corresponding RREP back to the source node following the same reverse route. Every node along the established route updates the received RREP and its routing table with the broadcast ID and the next nodes address of the reverse route. A route is completely constructed after the source node received the related RREP. Eventually source node creates or updates a forward route to the specified destination into the routing table and begins data transmission. In NLPRADSV, if there are two nodes that have same

energy level, an active node that is used in many data forwarding paths consumes energy more quickly, and thus, it has a shorter lifetime than the remaining inactive node. We evaluate the node lifetime that is based on its current energy and its past activity; however, we present a much simpler solution that does not need to calculate the predicted node lifetime from each data packet. NLPRADSV detects congested nodes and forward the data packets along the less congested nodes.

B. Route Maintenance

When a link breaks during data transmission, An upstream node which could detect a risk of link break should perform a local repair by broadcasting RREQ downstream if its location is rather close to the destination node. During the local repair process, the upstream node has to buffer packets of the affected connections until it receives the corresponding RREP that means the new route is constructed. However, if local repair is unsuccessful the intermediate node sends a Route error packet(RERR) to the source node. As a result the source node would try to re-establish a new route. In such case, the intermediate node which detects the link break is close to the source node, the node would send a RERR that contains address of the unreachable destination to the source node and the source node would do the same process as the case of unsuccessful local repairing.

The term NE_i represents the current energy of node i and E_{red} is the rate of energy reduction. NE_i can simply be obtained online from a battery management instrument, and E_{red} is the statistical value that is obtained from recent history. We use an exponentially weighted moving average methods to estimate the energy drain rate E_{red} . Every T seconds, node i reads the instantaneous energy in every period and the corresponding estimated energy drain rate E_{red} can be obtained, and at time t , we can obtain the estimated node lifetime as TN_i .

$$TN_i = \frac{NE_i}{E_{red}} \quad (3)$$

We set a threshold value for life time. When this remaining lifetime decrease goes beyond a threshold level, the upstream node detects this and node sends a RREQ with a field indicating that the downstream node is about to die and it has to find the alternate route with the good route life time and less congested node. The destination node which receives the RREQ sends back the RREP with a field indicating less congested nodes and good route life time. When the source node receives the RREP from various routes it sends the data packet along a route with less congested nodes and good route life time NLPRADSV adopts this local approach which improves the performance over DSR.

V. EXPERIMENTAL RESULTS

A. Simulation Setup

We used ns-2 Simulator [13] along with the extensions provided by CMU [14]. The simulation consists of 50 nodes confined in a 1000*1000 m² area. random connections were established using CBR traffic such that each node has a chance to connect to every other node. Packet size considered is 512 bytes and each simulation executed for 20000 sec. The initial battery capacity of each node is 100 units. nodes followed random waypoint mobility model with a specific max velocity and no pause time. each packet relayed or transmitted consumes a fixed amount of energy from the battery as given in equation 4

$$E(packet) = b * packet_size + c \quad (4)$$

B. Performance Metrics

Following performance metrics are used to evaluate the performance of our Scheme

- (1) Network life Time: It is defined as the time taken for a fixed percentage of the nodes to die due to energy resource exhaustion.
- (2) Packet delivery ratio: It is defined as the number of delivered data packets to the number of generated data packets in all nodes.
- (3) Control overhead: It is defined as the ratio of number of control packets to the delivered packets

C. Simulation Results

Network lifetime of DSR, NLPRDSR, AODV and NLPRADV are compared for a given scenario. here the speed of each node is 10m/s and radio transmission range is 125 km. Fig 1 shows that the time instances at which certain number of nodes have died when simulating DSR, NLPRDSR, AODV and NLPRADV. As can be seen, the nodes in DSR dies earlier than nodes in NLPDS but much earlier than nodes in AODV. It is also observed that Our schemes shows nodes in AODV also dies earlier than our new scheme NLPRADV. It is also observed that as the velocity of node movement increases, rate of energy consumption in the network goes up. This is expected since higher velocity of movement implies more route discoveries being performed and as a consequences higher energy consumption in the network.

We generate 200,000 data packets during the simulation. They are generated between random source and destination pairs at random times. many of these might not have reached their intended destination because of lack of existence of a route between the source and destination for various reasons. Also the network lifetime cleraly affects this ratio. If the network was alive for longer time, it implies that more data traffic goes through since we establish random connections

throughout the time of simulation. In Fig 2 Packet delivery ratio for NLPDSR is more compared to DSR at lower velocity. As the velocity increases the ratio decreases as shown. It is also observed that at velocity 15 m/s the ratio is almost equal. As can be seen the performance of AODV is better compared to DSR and NLPRDSR. It is also observed that NLPRADV perform better compared to all other protocols. As the velocity increases the link breaks frequently and it leads to more number of route discoveries and as a result the packet delivery ratio decreases in all the cases.

In NLPDSR route discovery process needs more control packets to be propogated in the network since it needs to compare all the possible paths between source and a sink and selects a path with maximum lifetime. As DSR uses source routing, control overhead is more compared to AODV. In AODV, when the link breaks local route repair algorithm is performed which reduces the number of control packets generated compared to source routing DSR. To show the overhead of NLPRDSR and NLPRADV on the network we have measured the ratio of number of control packets to number of number of delivered packets in the network. Fig 3 shows

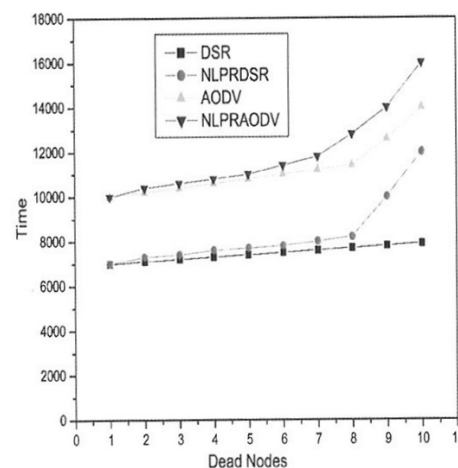


Fig. 1 Network lifetime

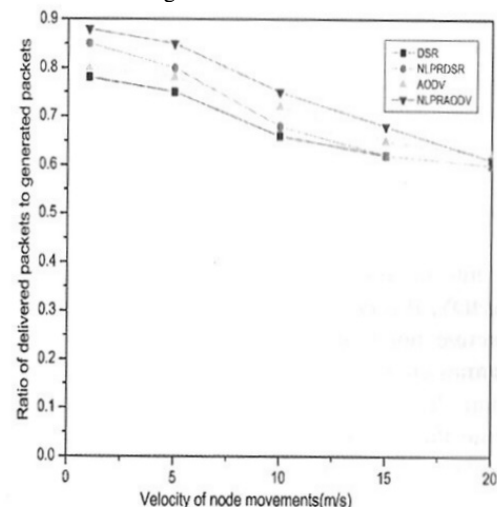


Fig. 2 Packet Delivery ratio

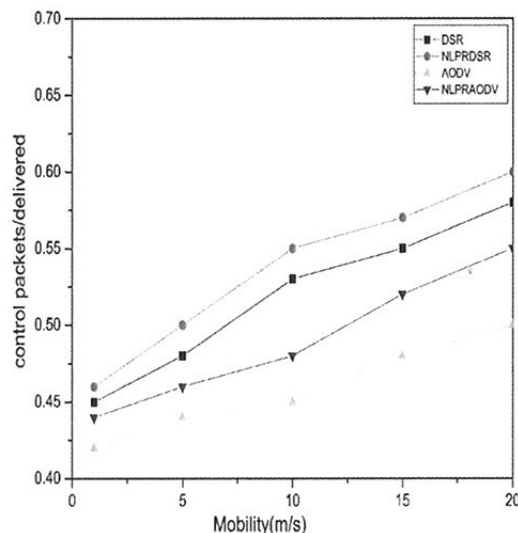


Fig. 3 Control Overhead

this ratio for DSR, NLPRDSR, AODV and NLPRAODV for different velocities of node movement. As the velocity of movement increases, routes are valid for shorter time and more route discoveries are done in the network resulting in more control packets. It is observed that NLPRDSR increases the ratio of control packets to transmit packets compared to DSR as control packets need to be exchanged for predicting the Node life. It is also observed that NLPRAODV generates more number of control packets compared to AODV due to the same reason. The control overhead of NLPRDSR is more compared to all the protocols described in our paper.

CONCLUSION

In this paper, We elaborate on Lifetime prediction Routing, a routing protocol to enhance the lifetime of a given mobile ad hoc network. We compare it with Dynamic source routing (DSR), Ad Hoc On Demand Distance vector routing (AODV). The main Objective of the work is to prolong the network lifetime. We show that NLPRAODV increases the network lifetime and increase the packet delivery ratio when compared with all the routing protocols described in this paper. It is being observed that the new route maintenance schemes NLPRDSR and NLPRAODV increase the control overhead. Finally we also point out some of the overheads of using Node Life Prediction Routing (NLPR) and show clearly when NLPR performs better and what are the factors that affect it.

REFERENCES

- [1] Ad hoc wireless networks, C Siva Ram Murthy and B.S.Manoj, Pearson Education, 2nd edition ,reprint 2005.
- [2] E.M. Rover and C.K.Toth. "A review of current routing protocols for ad hoc mobile wireless networks," IEEE pers commun.mag., vol. 6 no.2.pp 46-55, Apr.1999

- [3] IETF RFC 3626, Optimized Link State Routing protocol (OLSR), oct.2003
- [4] C.perkins, P. Bhagwat, "Highly dynamic destination sequenced distance- vector routing (DSDV) for mobile computers". ACM SIGCOMM 94, London, pp. 234-244, sep, 1994
- [5] Gupta P, Kumar. P, "A system and traffic dependent adaptive routing algorithm for ad hoc networks", The 36th Conference on Decision and Control, san Diego, california, Dec 1997
- [6] IETF RFC 3561, C perkins, e Belding Royer, s das, Ad Hoc On Demand Distance Vector (AODV) Routing, July 2003
- [7] IETF Internet draft, david R Johnson, david A maltz, Yih-Chun Hu, The Dynamic source Routing Protocol for Mobile Ad Hoc networks (DSR), draft-ietf-manet-dsr-09.Lxi
- [8] park V. carson M, " A highly adaptive distributed routing algorithm for mobile wireless networks", IEEE Infocom 97, Kobe Japan, 1997
- [9] J.Yackoski and C.C.Shen, Cross-layer inference-based fast link error recovery for MANETs," Proc. IEEE wireless communications and networking conference(WCNC 06), IEEE Press, April 2006, pp.715-722.
- [10] Y.M Wang, Z.K.Yang, and D.S.Zhao, Cache bypass and local recovery based reconstruction scheme for mobile ad hoc networks, journal of Electronics and Information Technology. Vol.29 july 2007, pp.1 16221627.
- [11] X.Q.Zheng, W.Guo and F.Li, Improvement of Link Break of AODV in Ad-hoc Networks, Journal of University of Electronic Science and Technology, vol32, oct.2003,pp. 569-573.
- [12] H.Lu and W.Guo, Keep the second Hop node in the AODV Routing
- [13] Protocol, Computer Science, vol.32, 2005, pp.110-112 [13] ns-2 manual. <http://www.isi.edu/nsnam/ns/doc/index.html>
- [14] CMU Monarch Extensions to ns. <http://www.monarch.cs.riee.edu/>

DYNAMIC MODELING OF ARM

Dr. T. C. MANJUNATH

Principal, HKBK College of
Engineering, S.No. 22/1, Nagawara,
Arabic College Post,
Bangalore-45, Karnataka.
manjuiitb@yahoo.com

Dr. VAIBHAV A MESRAM

Professor, PG Studies & Research,
HKBK College of Engg.,
Bangalore-45, Karnataka.
vam_25@yahoo.com

Dr. G. V. JAYARAMAIAH

Professor & HOD, Dept. of ECE,
Dr. Ambedkar Inst. of Tech.,
Bangalore, Karnataka.
jramgv@yahoo.com
gvjayaram@gmail.com
gv_jayaram@rediffmail.com

ABSTRACT

This paper features the dynamic modelling of a indigenously developed 3-axis SCARA robot arm which is used for doing successful robotic manipulation task in the laboratory. A 3-axis stationary robotic arm was developed in the laboratory. For this robotic system, a in-depth dynamic model of the system was carried out to study the dynamics of the designed and fabricated system. The dynamic concept was also implemented on the designed robot successfully.

Index Terms— SCARA, Dynamic model, Jacobean, Euler-Lagrange.

I. INTRODUCTION

ROBOTICS is an interdisciplinary field that mixes various engineering disciplines into one. In this, work a unique 3 axes system is designed and fabricated with indigenous components with a brief dynamic analysis of the designed robot. The dynamically modeled and designed robot is used for some PNP operations and was named as a Selective Compliance Assembly Robot Arm (SCARA). The primary motive behind the work was to develop a modular educational robotic system, the **CRUST 2007** (Computerized Robotic Unit with Selective Tractability system) with the help of locally available components, sub-systems as shown in Fig. 1.

The paper is organized as follows. In section 2, a brief introduction about the constructed robot is given. Next, the mathematical modeling of the dynamics is presented along with the algo. The conclusions are finally presented in the last section with the simulation results followed by the references.

II. PHYSICAL STRUCTURE DESIGN

The designed SCARA robot is a 4 DOF stationary robot arm having base, elbow, vertical extension and tool roll and consisting of both rotary and prismatic joints. There is no tool yaw and tool pitch (only tool roll) [1]. There are 3 joints, 3 axis (2 major axes - base, elbow, and one minor axis vertical extension). The 3 DOF's are given by Base, Elbow, Vertical Extension as shown in the Fig. 2, i.e., there are 3 rotary joints and 1 prismatic joint [9].

Since $n = 3$; 12 KP's are to be obtained and 4 RHOCP's are to be attached to the various joints [2] as shown in the LCD in Fig. 3 [10].

The vector of joint variables is a combination of θ and d ,

$$\text{i.e., } q = \{ \theta, d \}^T.$$

Vector of joint variables are

$$q = \{ \theta_1, \theta_2, d_3 \}^T.$$

Vector of joint distances are

$$d = \{ d_1, 0, d_3 \}^T = \{ 500, 0, d_3 \}^T \text{ mm.}$$

Vector of link lengths are

$$a = \{ a_1, a_2, 0 \}^T = \{ 300, 250, 0 \}^T \text{ mm.}$$

Vector of link twist angles are

$$\alpha = \{ \alpha_1, \alpha_2, 0 \}^T = \{ \pm \pi, 0, 0 \}^T.$$

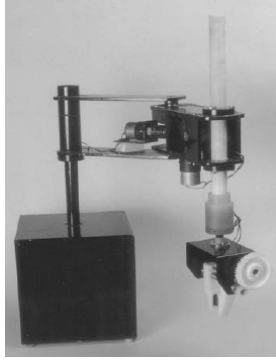


Fig. 1. SCARA robot

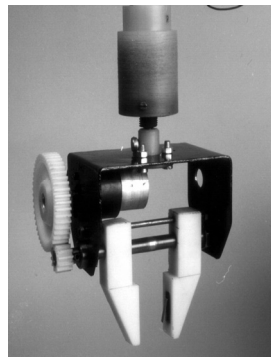


Fig. 2. Robot's gripper

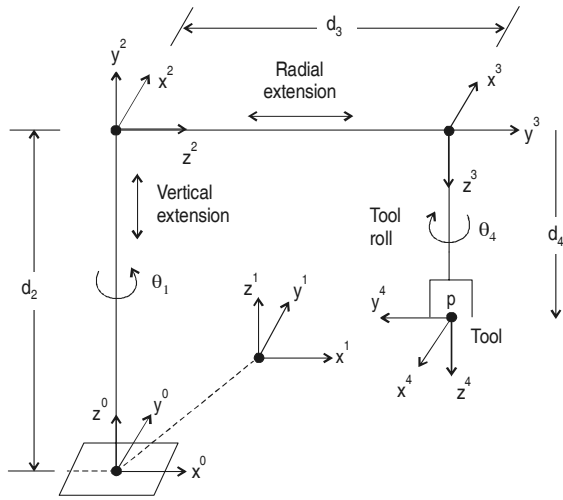


Fig. 3 The link coordinate diagram (LCD) of the SCARA robot

All the 3 joint axes are vertical in nature (all the z - axes can be pointing down or up) as shown in Fig. 3. The 3 (B, E, VE) axes are called as the major axes and are used for positioning the wrist and the last one is used to orient the tool or the gripper in the direction of the object. The first 3 axes determine the shape and size of work envelope [8]. It consists of a L shaped structure, to the end of which the second link is attached. There are 2 links a_1 and a_2 which move parallel to work surface; The vertical extension d_3 is variable and moves in a direction $\perp^r$ to the work surface ; length of the gripper / tool / EE is d_3 and its maximum protrusion is 300 mm [2].

The tool / gripper / EE is permanently pointing down as shown in Fig. 2 and can rotate in a plane $\perp^r$ to the work surface plane $x^0 y^0$. The approach vector r^3 is fixed, i.e., $r^3 \perp^r x^0 y^0$ (work surface) plane ; $r^3 = -z^0$. Because of this reason, our designed and dynamically modeled SCARA robot can do robotic manipulation directly from above the object when exact perpendicularity is required [11]. The SCARA robot is a minimal representation of any robot [3]. Our SCARA robot is a special type of polar / spherical coordinate robot in which the major axes are R R P [7].

III. ALGORITHM

In this section, the Euler-Lagrange Algorithm [2] is presented which is used to obtain the dynamic model of the designed and fabricated 3-axis robot arm. This algorithm is a 2-pass algorithm and consisting of 10-step algo which is shown as follows [2] :

1. Apply the DH algorithm to assign the link coordinates orienting frame L_i with the principal axes of the link i . Set $T_0^0 = I, i = 1, D(q) = 0$.
2. Find Δc^i , the homogeneous coordinates of the center of mass of link i w.r.t the frame L_i .
3. Let L_c be the translation of the frame L_i to the center of the mass of link i .
4. Compute $\bar{D}_i$, the inertia tensor of the link i about its center of mass expressed wr.t. the frame L_c .
5. Compute $z^{i-1}(q) = R_0^{i-1}(q) i^3$,
 $T_0^i(q) = T_0^{i-1}(q) T_{i-1}^i(q)$,
 $\bar{c}^i(q) = H_1 T_0^i(q) \Delta c^i$,
 $D_i(q) = R_0^i(q) \bar{D}_i [R_0^i(q)]^T$
6. Compute the Jacobean matrix $J^i(q)$. Partition it and compute
 $D(q) = D(q) + [A^k(q)]^T m_k A^k(q) + [B^k(q)]^T D_k(q) B^k(q)$
7. Increment the count as $i = i + 1$. Note that if $i \leq n$, go to step no. 3, else, $i = 1$ and continue.
8. Compute the matrices $C^i(q), h_i(q), b_i(\dot{q})$
9. Obtain the i^{th} equation as

$$\sum_{j=1}^n D_{ij}(q) \ddot{q}_j + \sum_{k=1}^n \sum_{j=1}^n C_{kj}^i \dot{q}_k \dot{q}_j + h_i(q) + b_i(\dot{q}) = \tau_i$$
10. Increment the count as $i = i + 1$. Go to step no. 9.

IV. INTRODUCTION TO THE DYNAMIC MODEL

In this section, we obtain the dynamic model of the designed and fabricated 3-axis robot. To keep the dynamic model relatively simple, it is assumed that the mass of the last link and that of the gripper attached to the tip of the arm is very small compared with the masses of the other links and

hence, it can be ignored from the dynamic point of view. This is an important point to be noted because the deflection ratio of the arm has to be maintained, i.e., the links of the arm goes on reducing as it moves away from the base, in other words to say, the link lengths become progressively smaller and less massive as one moves away from the base till the tip. An assumption is made while obtaining the dynamic model of the designed and fabricated robot, i.e., the three links are thin cylinders of rods of masses m_1 , m_2 and m_3 respectively. Note that the coordinate frames are assigned in such a manner that the frame L_k is aligned with the principal axes of the rod m_k for $1 \leq k \leq 3$. The KP table is shown below in Table 1 [2].

Joint	d	a	θ	α	Soft home position
1	d_1	a_1	q_1	π	0
2	0	a_2	q_2	0	0
3	q_3	0	0	0	$l_3/2$

Table 1 : KP table of the designed robot arm

Using the algorithm, we start with $i=1$. Assume that the links are homogeneous. When the joint 1 is activated, the rod of length a_1 and mass m_1 rotates. The center of mass of link 1 in the frame L_1 is $\Delta c^1 = \begin{bmatrix} -\frac{a_1}{2} & 0 & 0 & 1 \end{bmatrix}^T$. If L_c is the CF obtained by translating the frame L_1 along the axis x^1 by an amount equal to $-\frac{a_1}{2}$, then, the inertia tensor

$$\bar{D}_1 = \left(\frac{m_1 a_1}{12} \right) \text{diag}\{0, 1, 1\}.$$

Applying the 5th step of the algorithm, we get the homogeneous coordinate transformation matrix as [2]

$$T_0^1 = \begin{bmatrix} C_1 & S_1 & 0 & a_1 C_1 \\ S_1 & -C_1 & 0 & a_1 S_1 \\ 0 & 0 & -1 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Combining the above two equations, we get the center of mass of link 1 in the base frame as [6]

$$\begin{aligned} \bar{c}_1 &= H_1 T_0^1 \Delta c^1 \\ &= \begin{bmatrix} \frac{a_1 C_1}{2} & \frac{a_1 S_1}{2} & d_1 \end{bmatrix} \end{aligned} \quad (2)$$

Use the rotation matrix of T_0^1 , the inertia tensor matrix is obtained as

$$\begin{aligned} D_1 &= R_0^1 \bar{D}_1 (R_0^1)^T \\ &= \frac{m_1 a_1}{12} \begin{bmatrix} S_1^2 & -S_1 C_1 & 0 \\ -S_1 C_1 & C_1^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3)$$

Finally, the Jacobean matrix of size (6×3) for the link 1 is given by [2]

$$J^1 = \begin{bmatrix} -a_1 S_1/2 & 0 & 0 \\ -a_1 C_1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (4)$$

This Jacobean matrix is partitioned into two parts as A^1 and B^1 . Therefore, the manipulator inertia tensor matrix is given by [2]

$$D(q) = (A^1)^T m_1 A^1 + (B^1)^T D_1 B^1 \quad (5)$$

$$= \left(\frac{m_1 a_1^2}{4} \right) \text{diag}\{1 \ 0 \ 0\} + \left(\frac{m_1 a_1^2}{12} \right) \text{diag}\{1 \ 0 \ 0\} \quad (6)$$

$$= \left(\frac{m_1 a_1^2}{3} \right) \text{diag}\{1 \ 0 \ 0\} \quad (7)$$

Still now, we have obtained the dynamic model of only the link 1. Now, we make $i=2$ and repeat the steps of the algorithm given in section 3 to obtain the dynamic model of the link 2. Now, considering the analysis of link 2 in the frame L_2 coordinates, we get [2]

$$\Delta c^2 = \begin{bmatrix} -\frac{a_2}{2} & 0 & 0 & 1 \end{bmatrix}^T. \quad (8)$$

Similarly, the inertia tensor

$$\bar{D}_2 = \left(\frac{m_2 a_2^2}{12} \right) \text{diag}\{0, 1, 1\}.$$

Applying the 5th step of the algorithm again to the link 2, we get the homogeneous coordinate transformation matrix as [5]

$$T_0^2 = \begin{bmatrix} C_{1-2} & S_{1-2} & 0 & a_1 C_1 + a_2 C_{1-2} \\ S_{1-2} & -C_{1-2} & 0 & a_1 S_1 + a_2 S_{1-2} \\ 0 & 0 & -1 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9)$$

Combining the above two equations, we get the center of mass of link 2 in the base frame as

$$\begin{aligned} \bar{c}_2 &= H_1 T_0^2 \Delta c^2 \\ &= \begin{bmatrix} a_1 C_1 + \frac{a_2 C_{1-2}}{2} & a_1 S_1 + \frac{a_2 S_{1-2}}{2} & d_1 \end{bmatrix}^T \end{aligned} \quad (10)$$

Use the rotation matrix of T_0^2 , the inertia tensor matrix is obtained as [2]

$$D_2 = R_0^2 \bar{D}_2 (R_0^2)^T$$

$$= \frac{m_2 a_2}{12} \begin{bmatrix} S_{1-2}^2 & -S_{1-2}C_{1-2} & 0 \\ -S_{1-2}C_{1-2} & C_{1-2}^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

Finally, the Jacobean matrix of size (6×3) for the link 2 is given by

$$J^2 = \begin{bmatrix} -a_1 S_1 - \frac{a_2 S_1}{2} & 0 & 0 \\ a_1 C_1 + \frac{a_2 C_1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \quad (12)$$

This Jacobean matrix is partitioned into two parts as A^2 and B^2 . Therefore, the manipulator inertia tensor matrix is given by adding the inertia matrix of link 2 with that of link 1 and is finally given by [2]

$$D(q) = (A^2)^T m_2 A^2 + (B^2)^T D_2 B^2 \quad (13)$$

$$= \begin{bmatrix} \left(\frac{m_1}{3} + m_2\right) a_1^2 + m_2 C_2 a_1 a_2 + \frac{m_2 a_2^2}{3} & -\frac{m_2 C_2 a_1 a_2}{2} - \frac{m_2 a_2^2}{3} & 0 \\ -\frac{m_2 C_2 a_1 a_2}{2} - \frac{m_2 a_2^2}{3} & \frac{m_2 a_2^2}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Still now, we have obtained the dynamic model of only the link 2. Now, we make $i=3$ and repeat the steps of the algorithm given in section 3 to obtain the dynamic model of the link 3. Now, considering the analysis of link 2 in the frame L_2 coordinates, we get [2]

$$\Delta c^3 = \begin{bmatrix} 0 & 0 & -\frac{l_3}{2} & 1 \end{bmatrix}^T. \quad (14)$$

Similarly, the inertia tensor

$$\bar{D}_3 = \left(\frac{m_3 l_3^2}{12}\right) \text{diag}\{1, 1, 0\}.$$

Applying the 5th step of the algorithm again to the link 3, we get the homogeneous coordinate transformation matrix as

$$T_0^3 = \begin{bmatrix} C_{1-2} & S_{1-2} & 0 & a_1 C_1 + a_2 C_{1-2} \\ S_{1-2} & -C_{1-2} & 0 & a_1 S_1 + a_2 S_{1-2} \\ 0 & 0 & -1 & d_1 - q_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (15)$$

Combining the above two equations, we get the center of mass of link 3 in the base frame as [2]

$$\bar{c}_3 = H_1 T_0^3 \Delta c^3$$

$$= \begin{bmatrix} a_1 C_1 + a_2 C_{1-2} & a_1 S_1 + a_2 S_{1-2} & d_1 - q_3 + \frac{l_3}{2} \end{bmatrix}^T \quad (16)$$

Use the rotation matrix of T_0^3 , the inertia tensor matrix is obtained as [2]

$$D_3 = R_0^3 \bar{D}_3 (R_0^3)^T \quad (17)$$

Finally, the Jacobean matrix of size (6×3) for the link 3 is given by

$$J^3 = \begin{bmatrix} -a_1 S_1 - a_2 S_{1-2} & a_2 S_{1-2} & 0 \\ a_1 C_1 + a_2 C_{1-2} & -a_2 C_{1-2} & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \quad (18)$$

This Jacobean matrix is partitioned into two parts as A^3 and B^3 . Therefore, the overall manipulator inertia tensor matrix is given by adding the inertia matrix of link 3 with that of link 2 and link 1 and is finally given by [2]

$$D(q) = D(q) + (A_3)^T m_3 A^3 + (B^3)^T D_3 B^3 \quad (19)$$

$$= D(q) + m_3 \begin{bmatrix} a_1^2 + 2a_1 a_2 C_2 + a_2^2 & -(a_1 a_2 C_2 + a_2^2) & 0 \\ -(a_1 a_2 C_2 + a_2^2) & a_2^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (20)$$

Thus, the final dynamic model of the inertia tensor of the 3-axis stationary system is given by [2]

$$D(q) = \begin{bmatrix} d^1(q) & d^2(q) & d^3(q) \end{bmatrix}, \quad (21)$$

where the 3 columns of the dynamic model expression are given in the previous equations. Thus, we can see from the above equation that the motion of the joints depends completely on the mass of the links. Hence, to reduce the overall inertia of the system and improve the dynamic motion, the links are made of very lighter materials such as high grade aircraft aluminum so that the overall weight is less and less torque is required to move the links and hence improves the dynamic performance of the dynamic system. [3]

V. DYNAMIC SIMULATIONS

Once, the dynamic model was obtained, for a given set of joint variables, say, $q = \{30^\circ \quad -60^\circ \quad 45^\circ\}^T$, the dynamic equations are obtained as

$$\text{output} = \begin{bmatrix} 0.866 & 0 & 0.5 \\ 0 & 1 & 0 \\ -0.5 & 0 & 0.866 \end{bmatrix} \begin{bmatrix} 0.866 & -0.5 & 0 \\ 0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (22)$$

The output responses of the 3 joints were obtained using MATLAB simulations and also implemented in real time using interfacing cards, driver, computers, etc.,. The experimental results obtained are as shown below.

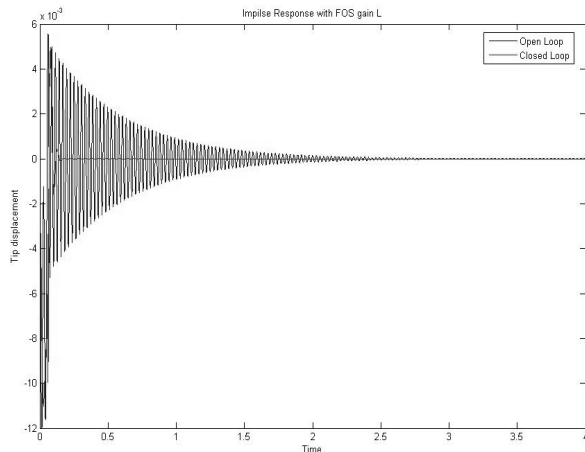


Fig. 4 : Dynamic response of the system

These dynamic equations are used to obtain the dynamic performance of the system, say, when a input of certain parameter is given, how the output settles quickly. The system responses settles so quickly that the damping was taken very low. The settling times of the joints when actuated were very low. The settling time of almost all the joints were found to be within 3 - 4 seconds [4].

CONCLUSIONS

A brief investigation into the dynamic analysis of the designed and fabricated SCARA robot in the lab is carried out. The mathematical analysis of the same is also presented in this paper. A numerical analysis is also done. An experiment was also conducted and the simulated dynamic response and the experimental results were approximately the same. This has proved the effectiveness of the dynamic model. The knowledge of workspace of a robot is very much essential in order to plan the trajectories in 3D space.

REFERENCES

- [1] Craig J, *Introduction to Robotics : Mechanics, Dynamics & Control*, Addison Wessely, USA, 1986.
- [2] Robert, J. S., *Fundamentals of Robotics - Analysis and Control*, PHI, New Delhi.
- [3] Klafter, Thomas and Negin, *Robotic Engineering*, PHI, New Delhi.

- [4] Fu, Gonzalez and Lee, *Robotics: Control, Sensing, Vision and Intelligence*, McGraw Hill.
- [5] Groover, Weiss, Nagel and Odrey, *Industrial Robotics*, McGraw Hill.
- [6] Ranky, P. G., C. Y. Ho, *Robot Modeling, Control & Applications*, IFS Publishers, Springer, UK.
- [7] Crane, Joseph Duffy, *Kinematic Analysis of Robotic Manipulators*, Cambridge Press, UK.
- [8] Manjunath, T.C., (2005), *Fundamentals of Robotics*, Fourth edn., Nandu Publishers, Mumbai.
- [9] Manjunath, T.C., (2005), *Fast Track to Robotics*, Second edn., Nandu Publishers, Mumbai.
- [10] P G Ranky and C Y Ho, *Robot Modeling, control and applications*, IFS publishers, Springer, UK.
- [11] Asada, H., and J.E. Slotine, *Robot Dynamics & Control*, Wiley, NY.

PUBLIC KEY CRYPTOGRAPHIC COMMUNICATION SYSTEM

Dr. T. C. MANJUNATH

Principal, HKBK College of
Engineering, S.No. 22/1, Nagawara,
Arabic College Post,
Bangalore-45, Karnataka.
manjuiitb@yahoo.com

Dr. VAIBHAV A MESRAM

Professor, PG Studies & Research,
HKBK College of Engg.,
Bangalore-45, Karnataka.
vam_25@yahoo.com

Dr. G. V. JAYARAMAIAH

Professor & HOD, Dept. of ECE,
Dr. Ambedkar Inst. of Tech.,
Bangalore, Karnataka.
jramgv@yahoo.com
gvjayaram@gmail.com
gv_jayaram@rediffmail.com

ABSTRACT

Quantum Cryptography is the recent application of the principles of quantum mechanics to cryptography has led to a remarkable new dimension in secret communication since it is not a crypto algorithm, a better name for quantum cryptography would be Quantum Key Distribution. It exploits the quantum mechanical properties of light particles (photons) to generate secret keys (strings of random numbers) that can be shared between two parties (for example, you and your bank) and used to encrypt data to safeguard it from snoopers. The quantum part gives you the promise that when you've topped up your secrets, only you and your service provider own this particular random digit string. Typically, QKD systems transmit a stream of differently oriented photons to represent 1s and 0s through an optical fibre or a free-space link. The snooper-proofing is intrinsic due to the fragility of quantum states: if you try to measure them they collapse, which is a marker for tampering, alerting the legitimate users to the presence of an eavesdropper. In this paper, a brief review of the same is presented.

I. INTRODUCTION

QUANTUM cryptography uses our current knowledge of physics to develop a cryptosystem that is not able to be defeated - that is, one that is completely secure against being compromised without knowledge of the sender or the receiver of the messages. The word quantum itself refers to the most fundamental behavior of the smallest particles of matter and energy: quantum theory explains everything that exists and nothing can be in violation of it. Quantum cryptography is different from traditional cryptographic systems in that it relies more on physics, rather than mathematics, as a key aspect of its security model. Essentially, quantum cryptography is based on the usage of individual particles/waves of light (photon) and their intrinsic quantum properties to develop an unbreakable cryptosystem - essentially because it is impossible to

measure the quantum state of any system without disturbing that system. It is theoretically possible that other particles could be used, but photons offer all the necessary qualities needed, their behavior is comparatively well-understood, and they are the information carriers in optical fiber cables, the most promising medium for extremely high-bandwidth communications.

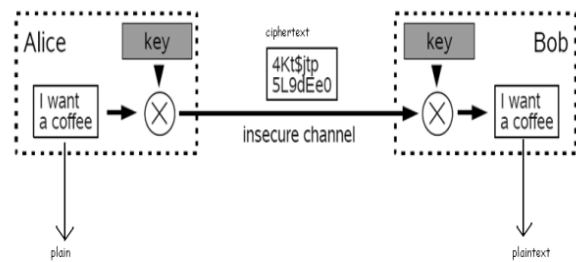


Fig. 1 : CCS Block diag.

A brief description of a classical cryptographic system (CCS) is illustrated in Fig. 1. Following are few important terms that are commonly used in cryptography:

- i. **Plaintext** : Plaintext is the original intelligible message, which is fed as input to the encryption algorithm
- ii. **Encryption algorithm** : Encryption algorithm is the algorithm which performs various substitutions and transformation on plaintext
- iii. **Secret key** : Secret key is the second input used by the encryption algorithm to generate encrypted message, and to decryption algorithm to recover the plaintext from the encrypted message
- iv. **Ciphertext** : Ciphertext is the scrambled message generated by the encryption algorithm using plaintext and secret key.
- v. **Decryption algorithm** : Decryption algorithm is the algorithm that extracts the plaintext message from the Ciphertext using a copy of the same secret key that was used for generating that message

In entire discussion for our convenience we call the sender as Alice and receiver as Bob. For almost all

practical cryptographic systems, the key K is substantially shorter than the length of the plaintext. As a result, the Ciphertext is no longer perfectly secure. However, if the encryption method and key K are wisely chosen, then Alice's communication to Bob will be practically secure. By practically secure," we mean that, although adversary Eve is theoretically able to decrypt Alice and Bob's communication without any knowledge of their key, she can not do so because the required computational time and resources are simply beyond her capability and means. The Data Encryption Standard (DES) is believed to be an example of such a practically secure encryption system.

In any case, one weakness of classical cryptographic communication systems is that secret communication can only take place after a key is communicated in secret over a totally secure communication channel. This is frequently referred to as the catch 22 of cryptography.

Catch 22: Before Alice and Bob can communicate over unsecure channel, they must communicate in secret.

There is even more to this catch 22, namely:

Catch 22a: Even if Alice and Bob somehow succeed in communicating their key over a secure communication channel, there is simply no classical cryptographic mechanism guaranteeing with total certainty that their key was transmitted securely, i.e., that their "secure" communication channel is free of Eve's unauthorized intrusion.

A proposed solution to the catch 22 of classical cryptographic communication systems is the modern public key cryptographic system (PKCS) as illustrated in Fig. 2. For public key cryptographic systems, it is no longer necessary for Alice and Bob to exchange key over a secure channel. Instead, Alice and Bob both create their own individual encryption/decryption key pairs ($EA;DA$) and ($EB;DB$), respectively. Then they both keep their decryption keys DA and DB secret from everyone, including each other, and publish or publicly broadcast their encryption keys EA and EB for the entire world to see. The security of such a public key cryptographic system depends on the selection of an encryption/decryption algorithm which is a trapdoor function. As a result, recovering the decryption key from the encryption key is computationally infeasible. The RSA public key cryptographic system is believed to be an example of such a cryptographic system.

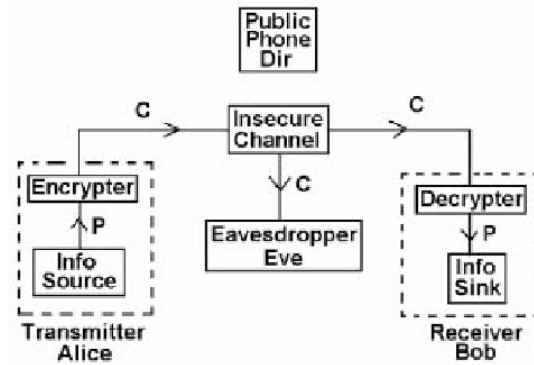


Fig. 2 : A Public Key Cryptographic Communication System

One major drawback to public key cryptographic systems is that no one has yet been able to prove that practical trapdoor functions exist. As a result, no one is really sure how secure such public key cryptographic systems are. Moreover, if researchers succeed in building a feasible quantum computer, Shor's quantum factoring algorithm could break RSA easily, i.e., in polynomial time.

Yet another drawback to public key cryptographic systems is that, in terms of some everyday implementations, such systems frequently do not circumvent the catch 22 of classical cryptography after all. The keys for many practical public key cryptographic systems are frequently managed by a key bank that is independent of Alice and Bob. Thus, secret communications over a secure channel from the key bank to Alice and Bob are required before Alice and Bob can secretly communicate.

II. QUANTUM CRYPTOGRAPHY

The basic principles involved in quantum cryptography can be explained as follows in the following paragraphs. In Quantum cryptography we shall be using quantum particles i.e. photons as the courier for the keys to be transferred. Since in this method we are not encrypting any data, we are just dealing with transfer or distribution of key it would be more probable to call this method as Quantum Key Distribution (QKD) rather than Quantum Cryptography (although generally it is so referred). The basic principle of quantum cryptography revolves around Heisenberg's Uncertainty Principle.

The order to make key distribution effective, in QKD we try to use quantum particles as the courier for the information. So we encode the secret key into quantum particle. Now once this is done it is not possible to copy these quantum particles perfectly or clone perfectly. If tried copying or measuring the quantum particle, would lead to disturbance in the state of the particles.

• polarization for 0 and 1:

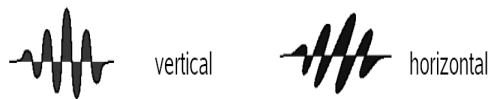


Fig. 3 : Encryption using the polarization of photons.

In case of particle being tampered, Bob tries to read the state of the quantum particle, he finds that the state that he received is different from the state that Alice had sent. So eavesdropping leaves out traces. This can be achieved by encrypting the keys by the polarization state of photons which we shall discuss in the later sections.

The key encryption is done on the polarization of single photons. Photons can be is the smallest indivisible particle of electromagnetic energy and electromagnetic energy can be viewed as the wave. In quantum key distribution we can encode a '1' bit by a horizontally polarized photon and a '0' bit by a vertically polarized photon or vice versa. This is shown in Fig. 3.

During decryption we send the photon through a polarizing beam splitter and this sorts out the photons which are vertically polarized and horizontally polarized. As shown in Fig. 4, by this we can separate the photons which are carrying '0' bit and '1' bit. The key point is that if the eavesdropper tries to detect the photon it is destroyed when tried to measure. So at the receiving end Alice will not receive all the photons and thus can make out the eavesdropping effect. The message then can be terminated and resent.

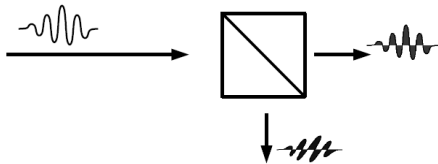


Fig. 4 : Decryption using polarizing beam splitter

While encryption of 0's and 1's it is also possible to have $+45^\circ$ & -45° polarization. This can be imagined as just rotating the vertically and horizontally polarized photons by 45° . It can be viewed as shown below in Fig. 5.



Fig. 5 : Inclined polarization

For the generation of polarized photons diode LASERs are used. The setup of photon generator is as shown in Fig. 6.

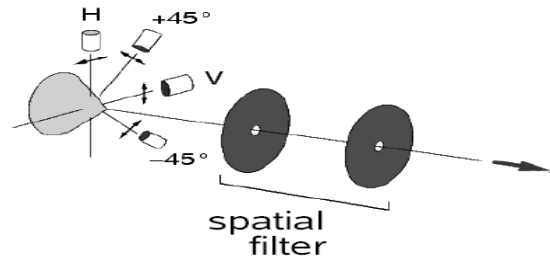


Fig. 6 : Preparation of polarized photons

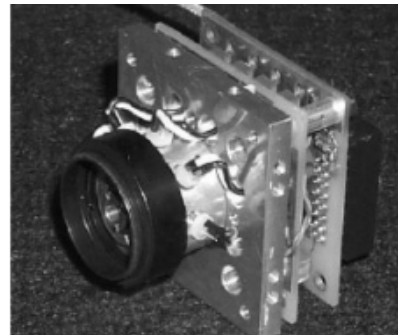


Fig. 7 : Photon Diode Laser

By simply rotating the diode we can get a horizontal, vertical, $+45^\circ$ & -45° polarized photons. Another method is to use four diodes each for a different kind of polarization. This helps in a more accurate preparation of polarized photons. The spatial filters are used to ensure that the light coming out of all the diodes have the same mode. These photons generated are then sent to bob.

Alice has this setup of four diodes and the spatial filters, the photons are generated and then sent to bob. This setup is usually used in the implementation of the standard protocols of quantum cryptography called BB84 protocol. In the next chapter we shall study about the BB84 protocol.

III. BB84 QUANTUM CRYPTOGRAPHIC PROTOCOL (USING PHOTONS AS TRUSTED CARRIERS)

The first quantum cryptographic communication protocol, called BB84, was invented in 1984 by Bennett and Brassard . This protocol has been experimentally demonstrated to work for a transmission over 30 km of fiber optic cable, and also over free space for a distance of over one hundred meters. This kind of QKD uses photons as a trusted carrier.

• trusted courier



Fig. 8 : QKD through trusted carriers

We now describe the BB84 protocol in terms of the polarization states of a single photon. Please note that the BB84 protocol could equally well be described in terms of any other two-state quantum system. Let H be the two dimensional Hilbert space whose elements represent the polarization states of a single photon. In describing BB84, we use two different orthogonal bases of H . They are the circular polarization basis which consists of kets

$$|\rightarrow\rangle \text{ and } |\leftarrow\rangle$$

for right and left circular polarization states, respectively, and the linear polarization basis which consists of the kets

$$|\uparrow\rangle \text{ and } |\downarrow\rangle$$

for vertical and horizontal linear polarization states, respectively. The BB84 protocol utilizes any two incompatible orthogonal quantum alphabets in the Hilbert space H . For our description of BB84, we have selected the circular polarization quantum alphabet A_0 .

Symbol	Bit
$ \rightarrow\rangle$	1
$ \leftarrow\rangle$	0

Circular polarization quantum alphabet A_0

And the linear polarization quantum alphabet.

Symbol	Bit
$ \uparrow\rangle$	1
$ \leftrightarrow\rangle$	0

Circular polarization quantum alphabet A_0

Bennett and Brassard note that, if Alice were to use only one specific orthogonal quantum alphabet for her communication to Bob, then Eve's eavesdropping could go undetected. For Eve could intercept Alice's transmission with 100% accuracy, and then imitate Alice by retransmitting her measurements to Bob. If, for example, Alice used only the orthogonal quantum alphabet, then Eve could measure each bit of Alice's transmission with a device based on some circular polarization measurement operator such as

$$|\rightarrow\rangle\langle\rightarrow| \quad \text{or} \quad |\leftarrow\rangle\langle\leftarrow|$$

Or if, Alice used only the orthogonal quantum alphabet, then Eve could measure each transmitted bit with a device based on some linear polarization measurement operator such as

$$|\uparrow\rangle\langle\uparrow| \quad \text{or} \quad |\leftrightarrow\rangle\langle\leftrightarrow|$$

The above strategy used by Eve is called opaque eavesdropping. (We will consider other and more sophisticated eavesdropping strategies later.)

To assure the detection of Eve's eavesdropping, Bennett and Brassard require Alice and Bob to communicate in two stages, the first stage over a one-way quantum communication channel from Alice to Bob, the second stage over a two-way public communication channel. (Please refer to Fig. 9)

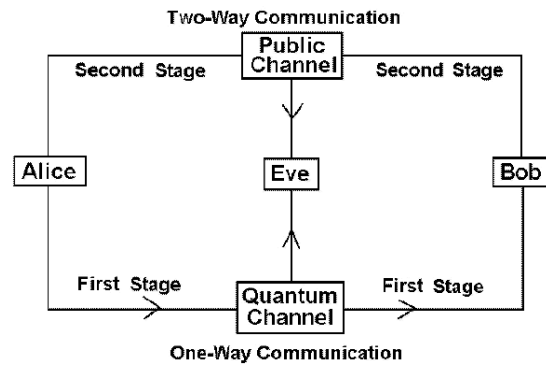


Fig. 9 : A quantum cryptographic system for securely transferring quantum key

Stage - 1. Communication over a quantum channel

In the first stage, Alice is required, each time she transmits a single bit, to use randomly with equal probability one of the two orthogonal alphabets A_0 or $A_\#$. Since no measurement operator of A_0 is compatible with any measurement operator of $A_\#$, it follows from the Heisenberg uncertainty principle that no one, not even Bob or Eve, can receive Alice's transmission with an accuracy greater than 75%.

This can be seen as follows. For each bit transmitted by Alice, one can choose a measurement operator compatible with either A_0 or $A_\#$, but not both. Because of incompatibility, there is no simultaneous measurement operator for both A_0 and $A_\#$. Since one has no knowledge of Alice's secret choice of quantum alphabet, 50% of the time (i.e., with probability 1/2) one will guess correctly, i.e., choose a measurement operator compatible with Alice's choice, and 50% of the time (i.e., with probability 1/2) one will guess incorrectly. If one guesses correctly, then Alice's transmitted bit is received with probability 1. On the other hand, if one guesses incorrectly, then Alice's transmitted bit is received correctly with probability 0.5. Thus in general, the probability of correctly receiving Alice's transmitted bit is

$$P = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$$

For each bit transmitted by Alice, we assume that Eve performs one of two actions, opaque eavesdropping with probability λ , $0 < \lambda < 1$, or no eavesdropping with probability $1-\lambda$. Thus, if $\lambda = 1$, Eve is eavesdropping on each transmitted bit; and if $\lambda = 0$, Eve is not eavesdropping at all. Because Bob's and Eve's choice of measurement operators are stochastically independent of

each other and of Alice's choice of alphabet, Eve's eavesdropping has an immediate and detectable impact on Bob's received bits. Eve's eavesdropping causes Bob's error rate to jump from 0.25 to

$$\frac{1}{4}(1-\lambda) + \frac{3}{8}\lambda + \frac{\lambda}{8}$$

Thus, if Eve eavesdrops on every bit, i.e., if $\lambda = 1$, then Bob's error rate jumps from 1/4 to 3/8 a 50% increase.

Stage 2 - Communication in two phases over a public channel

In stage 2, Alice and Bob communicate in two phases over a public channel to check for Eve's presence by analyzing Bob's error rate.

Phase 1 of Stage 2 : Extraction of raw key

Phase 1 of stage 2 is dedicated to eliminating the bit locations (and hence the bits at these locations) at which error could have occurred without Eves eavesdropping. Bob begins by publicly communicating to Alice which measurement operators he used for each of the received bits. Alice then in turn publicly communicates to Bob which of his measurement operator choices were correct. After this two way communication, Alice and Bob delete the bits corresponding to the incompatible measurement choices to produce shorter sequences of bits which we call respectively Alice's raw key and Bob's raw key.

If there is no intrusion, then Alice's and Bob's raw keys will be in total agreement. However, if Eve has been at work, then corresponding bits of Alice's and Bob's raw keys will not agree with probability

$$0.(1-\lambda) + \frac{1}{4}\lambda = \frac{\lambda}{4}$$

Phase 2 of Stage 2 : Detection of Eve's intrusion via error detection

Alice and Bob now initiate a two way conversation over the public channel to test for Eve's presence. In the absence of noise, any discrepancy between Alice's and Bob's raw keys is proof of Eve's intrusion. So to detect Eve, Alice and Bob select a publicly agreed upon random subset of m bit locations in the raw key, and publicly compare corresponding bits, making sure to discard from raw key each bit as it is revealed. Should at least one comparison reveal an inconsistency, then Eve's eavesdropping has been detected, in which case Alice and Bob return to stage 1 and start over. On the other hand, if no inconsistencies are uncovered, then the probability that Eve escapes detection is:

$$P_{false} = \left(1 - \frac{\lambda}{4}\right)^m$$

For example, if $\lambda = 1$ and $m = 200$, then

$$P_{false} = \left(\frac{3}{4}\right)^{200} \approx 10^{-25}$$

Thus, if P_{false} is sufficiently small, Alice and Bob agree that Eve has not eavesdropped, and accordingly adopt the remnant raw key as their Final secret key.

IV. QKD USING ENTANGLEMENT CONCEPT

In this Alice and Bob have an individual key and they derive correlation using these temper safe devices. The physical concept that is used in this type of qkd is called entanglement.

Quantum entanglement is a possible property of a quantum mechanical state of a system of two or more objects in which the quantum states of the constituting objects are linked together so that one object can no longer be adequately described without full mention of its counterpart — even though the individual objects may be spatially separated. This interconnection leads to non-classical correlations between observable physical properties of remote systems, often referred to as non-local correlations.

Now we shall see how an entangled photon pair is generated and used in quantum key distribution using a third party tamper safe device. We use a device based on non linear optics, this is called as Q895 method of generating photons in entangled states. The setup is as shown below in Fig. 10.

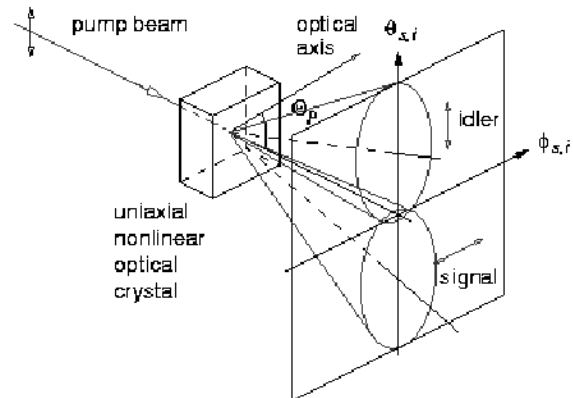


Fig. 10 : Entangled photon generation

The good thing about using this concept is that it is not only entangled in the photon but also entangled in the polarization state. In BB84 Alice had to decide the polarization state for transmission of data encoded in the photons and the selection of the polarization state had to be random so a random key generator was required. But in this case the randomness is provided by the physical nature of the entangled photons. What we do in this is that we take a non linear crystal like BBO (Beta Barium borate) and pass a pump beam, which is very well polarized, through it.

A spontaneous parametric down conversion is done when the pump beam passes through the BBO. So there is a possibility that photon may decay into 2 photons. These 2 photons are highly correlated not only in time but also in energy to maintain the law of conservation of energy. One of the photons is always vertically polarized and the other is always horizontally polarized. Now, as shown in the fig, at the place where the paths of the two photons coincide we get a randomly polarized photon. Thus the need of the random key generator is discarded.

Now that we have a photon pair one of the pairs is sent to Alice and the other to Bob. Then the BB84 protocol is used at Alice's side and the coded key is sent to Bob. At the bob's end the photons are compared and the correlation is used to get back the information. Protocol using this concept is called as Ekert-91.

The main disadvantage of this method that there is no timing information available to Alice or bob. Hence synchronization is a big problem i.e. at bob's end it is difficult to compare the respective photon pairs leading to wrong correlation calculation. We can overcome this problem by 2 methods:

- 1) Generating the photons randomly but at the same time intervals.
- 2) By using a central clock and synchronizing using synchronization bit.

V. ADVANTAGES AND APPLICATIONS

Security on the Internet is a big issue of today, in the days of prospective wireless communications, business transactions, e-learning, e-government, e-health etc. The application of QKD methods and protocols is a new arrival to the networks of today (and references therein). The authors have done work on Key Creation and Distribution both by classical methods (Chaotic Dynamic Systems Behavior for Random Number Generation) as well as in QKD applications.

The advantages of this system are as follows:

- Technical complexity very low as only passive components used and no switches required. Technically it is as complex as a CD writer.
- No random key generators needed this further reduces the complexity and size
- Software only synchronization
- Data impossible to steal
- Detects if someone is listening

The applications of quantum cryptography could be listed as follows.

Video Conferencing: Scientists from Toshiba's Cambridge Research Laboratory unveiled their invention to business leaders and government officials at Britain's Department of Trade and Industry in London on 27 April '08. Their system is capable of generating 100 quantum 'keys' every second. This is fast enough for every individual frame of video to be protected by its own encryption. This makes the system highly secure.

Secure financial transaction: Quantum cryptosystems based on the principle of quantum key distribution provides devices with fail-proof codes and ensures that hackers cannot intercept or tamper with the transmitted message. This is being implemented in some banks Swiss banks.

Grégoire Ribordy by id Quantique SA is the first quantum cryptographic application that is commercially available. The device is called a Vectis link encryptor. This uses a dedicated pair of optical fibres for key exchange, while data are transmitted over a conventional Ethernet connection using AES encryption.

CONCLUSION

It is not easy to emphasize how dramatic an impact the application of quantum mechanics has had and will have on cryptographic communication systems. From the perspective of defensive cryptography, it is now within the realm of possibility to build practical cryptographic systems which check for, detect, and prevent unauthorized intrusion. Quantum mechanics provides an intrusion detection mechanism never thought possible within the world of classical cryptography. Most importantly, the feasibility of these methods has been experimentally verified in a laboratory setting.

Moreover, from the perspective of offensive cryptography, the application of quantum mechanics to computation also holds forth the promise of a dramatic increase of computational parallelism for cryptanalytic attacks. Shor's quantum factoring algorithm is just one example of such potential. However, unlike quantum protocols, quantum computational parallelism has yet to be fully verified in a laboratory setting.

REFERENCES

- [1] IEEE paper for reference: Unconditionally secure key distillation from multi-photons in a single-photon polarization based Quantum Key Distribution Kiyoshi Tamaki (Center for Quantum Information and Quantum Control, Dept. of Electrical & Computer Engineering.

-
- [2] Barenco, Adriano, Charles Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter, Elementary gates for quantum computation, Phys. Rev. A, Vol. 52, No. 5, November, 1995, pp 3457 - 3467.
- [3] Barnett, Stephen M., Rodney Loudon, David T. Pegg, and Simon J.D. Phoenix, Communication using quantum states, Journal of Modern Optics, Vol. 41, No. 12, 1994, pp 2351 - 2373.
- [4] Barnett, Stephen M., and Simon J.D. Phoenix, Informationtheoretic limits to quantum cryptography, Physical Review A, vol. 48, no. 1, July 1993, 1050-2947.
- [5] Bell, J.S., Physics, 1, (1964), pp 195 - 200.
- [6] Benio_, Paul, Quantum mechanical models of Turing machines that dissipate no energy, Physical Review Letters, Vol. 48, No. 23, 7 June 1982, pp 1581 - 1585.
- [7] Bennett, Charles H., Francois Bessette, Gilles Brassard, Louis Salvail, and John Smolin, Experimental quantum cryptography, J. Cryptology (1992) 5: 3 - 28.
- [8] Bennett, Charles H., Quantum cryptography using any two nonorthogonal states, Physical Review Letters, Vol. 68, No. 21, 25 May 1992, pp 3121 - 3124.
- [9] Bennett, Charles H., Quantum cryptography: Uncertainty in the service of privacy, Science, Vol. 257, 7 August 1992, pp 752 - 752.
- [10] Bennett, Charles H., and Stephen J. Weisner, Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states, Physical Review Letters, Vol. 69, No. 20, 16 November 1992, pp 2881 - 2884.

TRANSFER FUNCTION MODELLING

Dr. T.C. MANJUNATH

Principal, HKBK College of
Engineering, S.No. 22/1, Nagawara,
Arabic College Post,
Bangalore-45, Karnataka.
manjuiitb@yahoo.com

Dr. VAIBHAV A MESRAM

Professor, PG Studies & Research,
HKBK College of Engg.,
Bangalore-45, Karnataka.
vam_25@yahoo.com

Dr. G.V. JAYARAMAIAH

Professor & HOD, Dept. of ECE,
Dr. Ambedkar Inst. of Tech.,
Bangalore, Karnataka.
jramgv@yahoo.com
gvjayaram@gmail.com
gv_jayaram@rediffmail.com

ABSTRACT

Moment of inertia is an important property of the mechanical aspect of the robot and affects its dynamic performance. This paper presents the development of a control model in the transfer function form to minimize the moment of inertia of a stationary robotic system.

1. INTRODUCTION

The property by virtue of which it resists any change in its state of rest or of uniform motion is called as inertia. Two types of inertia exist, viz., Translatory inertia and the rotary inertia. Translatory inertia is defined as the inertia that is offered by the translation of a body due to the application of a force and depends on the mass and acceleration ; i.e., $F = ma$. Translatory inertia is due to the mass. Translatory inertia is called as Mass. Rotational inertia is defined as the inertia that is offered by the rotation of a body due to the application of torque and depends on moment of inertia and angular acceleration ; i.e., $T = I\alpha$. Rotational inertia is due to the MI. Rotational inertia is called as Moment of Inertia.

Three important concepts of any body or matter are the concepts of relative distribution of area, mass and the weight. Quantitative estimates of the relative distribution of the area and the mass over the regions of interest are made by the concept of Moments of Inertia [MI] of the object. The concept of inertia is provided by the Newton's laws of motion. Moment of inertia plays a very important role in the dynamics and kinematics of robot mechanisms. A robot arm is made of links + joints and these joints are driven by actuators (motors / pistons).

The arm dynamics of the robot is the spatial displacement or the movement of the robot in the

3D space with a particular velocity and acceleration. It mainly depends on the mass and MI. Moment of inertia depends on the positioning, velocity, acceleration, moments, torques, stresses, strains in links deflection effect, bending effect, buckling effect, bending stresses, vibrations of the links, shear stresses, torsional stresses & torsional strains related to the links of the robot arm.

MI of a part of a robot, say a link gives the information about the mass distribution of the part at a particular distance w.r.t. a particular axis. When the mass of the link increases, its weight increases, thereby MI increases, thereby increasing the torque rating and size of the motors. More power is required to drive the joints. The drive power amplifier required to drive the actuators would be of higher rating. To overcome this, joint - link movement motors are to be light in weight and the link weight has to be reduced (links should be light in weight, but robust, sturdy, should not bend). Hence, robots are designed in such a way that MI of the links is reduced, so that less torque is sufficient to move the links.

Note that MI is responsible for the creation of torque. The MI also depends on the axis. Say, for example, if the area of the link is more, then MI is more (link horizontal) and it will bend due to application of a force at the other end (F_{bending} more), if the area of link is less (vertically placed), then, MI is less and it will not bend due to application of weight at the other end (F_{bending} less). $MI \uparrow$, the T requirement to activate a joint increases in order to manipulate physical objects. The robot manipulation depends on the MI and the mass, weight of its components such as links, motors, grippers, tools, payload weight etc., in addition, the torque developed by the motor depends on the load which in turn depends on the MI which in turn also depends on the column effect / buckling effect in the links, bending stresses and shear stresses that are developed in the links when they are loaded at one end. All these factors have

to be considered for robot stability, i.e., the link design totally depends on the MI.

The spatial displacement of the links should be according to the speed profile curve. So, links should be as light as possible and made up of high grade aluminum, the joint motors should be light or instead of mounting the heavy parts at the joints, mount it at the base and transmit the power from the base to the joint by making use of chains, ropes and pulleys. The joint-link pair mechanism should be designed with less weight so that the robot control will become easy and effect of MI on the links is reduced and less torque is required by the motors to displace the load. The heavy objects such as the motors and the gear mechanisms have to be mounted at the base and power should be transmitted from the base to the respective joints via the transmission devices (the chains, gears, ropes, pulleys, etc.) like in Rhino XR-3 robot arm. This will increase the positional accuracy of the robot and reduce the overall MI of the robot.

The length of the arm should go on reducing as one moves radially away from the robot base, i.e., the deflection ratio should be maintained (Ex. SCARA robot, $a_1 = 425$ mm, $a_2 = 375$ mm). If a third link would have been, then the length of that link would be $< a_2$. The main factors which are affecting the control problems and the MI are the Positioning, Velocity, Acceleration, Moments, Torques, Stresses, Strains in links, Deflection effect, Bending effect, Buckling effect, Bending stresses, Vibrations of the links, Shear stresses, Torsional stresses and strains related to the links of the robot arm and the controller rating. All these factors have to be considered while designing the robot and its control strategy for its successful robotic manipulation.

The paper is organized as follows. A brief introduction about the MI concepts etc., and its effects was discussed in the previous paragraphs. Section 2 gives a small review of the physical construction of the robotic system that was designed and fabricated. Section 3 gives an overview of the concepts in the design of the controller to overcome the MI. In section 4, for the physical model developed in the section 2, a control model in the mathematical form along with the experimental results are presented. Finally, the conclusions are presented in the last followed by the references.

2. INTRODUCTION TO THE ROBOT DESIGN

The designed robot for which we developed the control model is a 3 DOF stationary robot arm having base, elbow, vertical extension and consisting of both rotary and prismatic joints. There is no tool yaw and

tool pitch (only tool roll) [1]. There are 3 joints, 3 axis (2 major axes - base, elbow, and one minor axis vertical extension). The 3 DOF's are given by Base, Elbow, Vertical Extension as shown in the Fig. 2, i.e., there are 3 rotary joints and 1 prismatic joint [9]. Since $n = 3$; 12 KP's are to be obtained and 4 RHOCF's are to be attached to the various joints [2] as shown in the LCD in Fig. 3 [10].

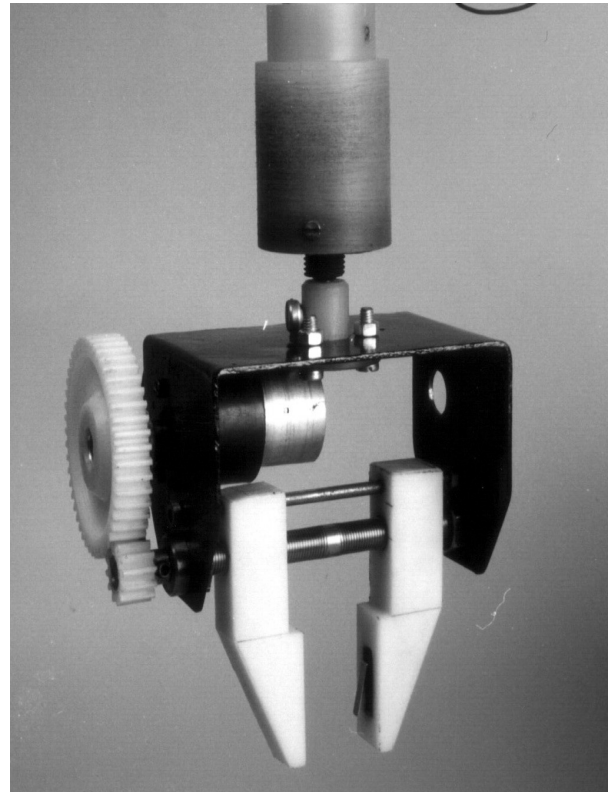


Fig. 2. Robot's gripper

The vector of joint variables is a combination of θ and d ,

$$\text{i.e., } q = \{ \theta, d \}^T.$$

Vector of joint variables are

$$q = \{ \theta_1, \theta_2, d_3 \}^T.$$

Vector of joint distances are

$$d = \{ d_1, 0, d_3 \}^T = \{ 500, 0, d_3 \}^T \text{ mm}.$$

Vector of link lengths are

$$a = \{ a_1, a_2, 0 \}^T = \{ 300, 250, 0 \}^T \text{ mm}.$$

Vector of link twist angles are

$$\alpha = \{ \alpha_1, \alpha_2, 0 \}^T = \{ \pm \pi, 0, 0 \}^T.$$

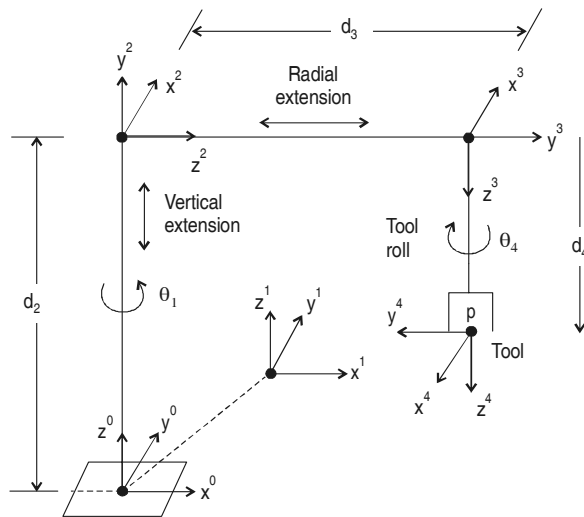


Fig. 3 The link coordinate diagram of the robot

All the 3 joint axes are vertical in nature (all the z - axes can be pointing down or up) as shown in Fig. 3. The 3 (B, E, VE) axes are called as the major axes and are used for positioning the wrist and the last one is used to orient the tool or the gripper in the direction of the object. The first 3 axes determine the shape and size of work envelope [8]. It consists of a L shaped structure, to the end of which the second link is attached. There are 2 links a_1 and a_2 which move parallel to work surface. The vertical extension d_3 is variable and moves in a direction $\perp^r$ to the work surface, length of the gripper / tool / EE is d_3 and its maximum protrusion is 300 mm [2].

The tool / gripper / EE is permanently pointing down as shown in Fig. 2 and can rotate in a plane $\perp^r$ to the work surface plane $x^0 y^0$. The approach vector r^3 is fixed, i.e., $r^3 \perp^r x^0 y^0$ (work surface) plane ; $r^3 = -z^0$. Because of this reason, our designed and dynamically modeled SCARA robot can do robotic manipulation directly from above the object when exact perpendicularity is required [11]. The SCARA robot is a minimal representation of any robot [3]. Our SCARA robot is a special type of polar / spherical coordinate robot in which the major axes are R R P [7].

3. CONTROLLER CONCEPTS

A controller is a device which controls each and every operation that is taken place inside the robotic system making decisions. The functions of a robot controller are : The robot is a mechanical system that must be controlled in order to accomplish a useful task (say , a PNP operation). The task involves the movement of the manipulator arm from the source to the destination, so that the primary function of the robot control system is to

position and orient the wrist with a specific speed, precision, accuracy. The controller serves as a interface between the robot and the computer. It controls each and every operation of the robot making decisions.

The controller consists of a driver unit, sometimes a power supply unit, electronic logic circuits, feedback circuitry, sensing circuitry, few amplifiers, IC chips and the actuators. The task of the control system is to execute the movements and actions of the manipulator according to the robot program and to coordinate the interaction with the environment. Controllers can operate either in open loop or in closed loop. Each axis of motion of the robot arm is separately actuated by a control circuitry which contains a driver unit which converts the electrical command signals of the computer to mechanical motions. Drives for computerized robotic systems designed are of electric type.

The purpose of the controller is to compare the actual output of the plant with the input command and to provide a control signal which will reduce the error to zero or as close to zero as possible. The controller generally consists of a summing junction where the input and the output signals are compared, a control device which determines the control action, the necessary power amplifiers and the associated hardware devices to accomplish the control action in the plant.

Robot control problem means how to control it to do a particular task effectively in its work space. Robot control problem is taken care of by a controller which is used to interface between the robot and the computing system. Various types of controllers such as P, PI, PD, PID, FOS, PO, H_2 , H_∞ , Sliding mode controller, programmable logic controllers, microcontrollers can be designed to overcome the MI & to do a correct task.

The designed controller is shown in the form of a block diagram in the Fig. 4.

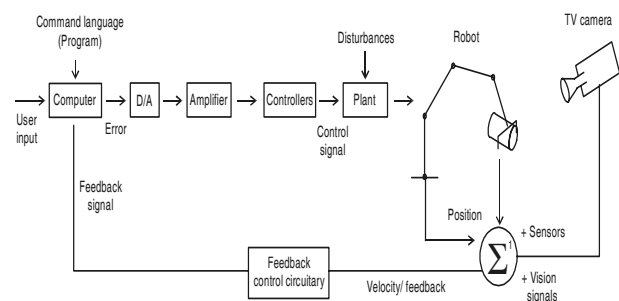


Fig. 4 : Block diagram of the controller

4. CONTROLLER DESIGN

Consider an electromechanical servo system in the designed robot, where the mechanical drive unit consists of an arm controlled DC servomotor and a gear train of high gear reduction ratio 'n' as shown in Fig. 5. The block diagram is in the Fig. 6. It gives the position and velocity feedback of the servo system. The position input signal θ_p radians acting on the system (electrical input to actuator - voltage, v) results in an output signal θ_0 radians (mechanical output : - angular displacement, θ). The servos used for robotic control application is a Type-1 system, i.e., it consists of 1 integrator in the feedback loop.

⇒ Forward path

→ Feedback path

K_e Motor back e.m.f. constant

K_m Motor torque constant

k_p Position feedback gain constant

k_v Velocity feedback gain constant

A Amplifier gain

J_m, J_L Rotor moment inertia and load MI

N Gear reduction ratio

D Viscous friction coefficient

θ_0 Controlled output

θ_p Reference input

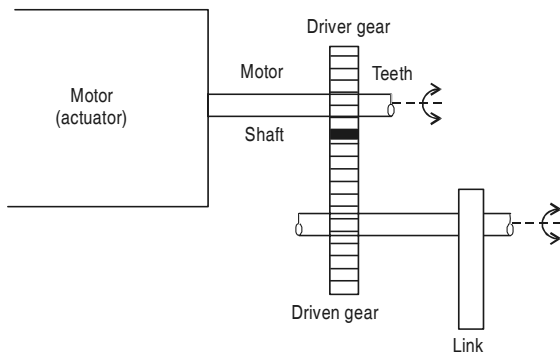


Fig. 5 Gear reduction unit

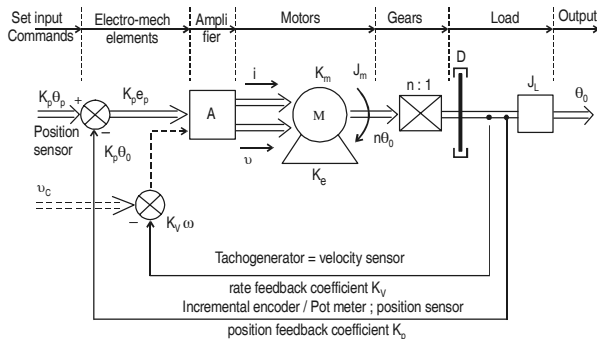


Fig. 6 Schematic block-diagram of a joint-link robot servo system actuated by a DC servo motor via a gear train

Two feedback loops are designed for the controller for the robotic unit. They are as follows :

- Output θ_0 is feedback using position sensor and the difference between the input and the output which is proportional to the position error signal $e_p(v)$ is found out using equation

$$e_p = K_p (\theta_p - \theta_0), \quad (1)$$

where K_p is position feedback gain coefficient (V / rads)

- Output θ_0 is feedback using a tachometer device which gives a signal $e_v(V)$ which is proportional to the velocity of the output shaft rotation and is given by

$$e_v = -K_v \frac{d\theta_0}{dt} \quad (2)$$

where K_v is velocity gain coefficient (V / rad . sec)

The two feedback signals given by Eqns. (1) and (2) are superimposed using a amplifier to give an effective error signal resulting in current control strategy for driving the robot actuators and is given by Eq. (3) as

$$i \propto^{al} (e_p - e_v) \\ i = A \left[K_p (\theta_p - \theta_0) - K_v \frac{d\theta_0}{dt} \right] \quad (3)$$

where A is the gain of the amplifier and is the proportionality constant. The gross torque (total torque) developed by servomotor on its shaft to drive the joint is directly proportional to the current and is given by Eq. (4).

$$T_g \propto^{al} i \\ T_g = k_m i \\ T_g = A K_m \left[K_p (\theta_p - \theta_0) - K_v \frac{d\theta_0}{dt} \right] \quad (4)$$

where k_m is the motor gain constant.

Thus, we have explained how the control torque is generated. Now, we shall use this generated torque to control the individual joint-link arm segments. A substantial part of the generated torque is absorbed in accelerating the rotor MI, $J_m (K_g m^2)$ which also includes the inertia of the gear train on the rotor shaft ; i.e., some part of the total torque produced is wasted in over coming the losses in the servomotor (inertia losses, damping losses, frictional losses, windage losses, copper losses, hysteresis and eddy current losses, mechanical losses).

Consider at this stage the only load to be the MI, J_m . Then, the torque T_m in N-m used to accelerate this MI is given by Eq. (5). The motor torque is given by

$$T_m = J_m \frac{d^2(n\theta_0)}{dt^2} = n J_m \frac{d^2\theta_0}{dt^2} \quad (5)$$

Subtract T_m from the gross torque T_g to give the rotor torque T_r which is used to control all the loads on motor shaft and is given by the Eq. (6).

∴, the useful torque available at the shaft of the rotor for driving the joint is = the gross torque – the wastage torque.

$$\begin{aligned} T_r &= [T_g - T_m] \\ &= T_g - n J_m \frac{d^2\theta_0}{dt^2} \end{aligned} \quad (6)$$

This torque T_r acts on the shaft of the servomotor. Therefore, its action must be reflected in slow running shaft of the gear train. The gear reduction unit magnifies the torque by gear ratio 'n'. The link is connected to the motor shaft via a series of gear trains. It is done so in order to convert the high speed rotation of the motor shaft into the slow speed running of the joint-link axis. Since, the link cannot rotate at such a high speed, the speed of the motor is reduced and torque is magnified using gears (gears are nothing but speed reduction devices and torque magnification devices). Therefore, the useful torque on slow rotation shaft after gear train is given by the Eq. (7) as

$$n T_r = n \left[T_g - n J_m \frac{d^2\theta_0}{dt^2} \right] \quad (7)$$

The load on the slow running shaft consists of

- Inertia's of secondary part of gear unit.
- Effective inertia of arm segments.
- Gripper and tool.
- Pay load / part / object.

Combine all of them, the overall effective moment of inertia J' is obtained and is reflected to the end of the arm segment. Useful torque, $T = nT_r$ is used in accelerating the effective overall inertia J' on the output shaft of load and is given by Eq. (8)

$$T' = J' \frac{d^2\theta_0}{dt^2} B \quad (8)$$

Also, part of the useful torque T' is used to overcome the viscous friction (proportional to velocity), coulomb friction (constant), stiction (instantaneous).

∴, mechanical damping torque with viscous damping factor D is given by Eq. (9).

$$T_D = D \frac{d\theta_0}{dt} \quad (9)$$

where D is the viscous damping factor.

Required torque for driving the effective load and the useful torque generated by the motor are equal. Equate them, we get a differential equation which gives the motion of the robot joint - link arm segments and is given by Eq. (10)

$$J' \frac{d^2\theta_0}{dt^2} + D \frac{d\theta_0}{dt} = n \left(T_g - n J_m \frac{d^2\theta_0}{dt^2} \right) \quad (10)$$

Substituting Eq. (4) in Eq. (10), we get,

$$J' \frac{d^2\theta_0}{dt^2} + D \frac{d\theta_0}{dt} = n \left\{ A K_m \left[K_p (\theta_p - \theta_0) - K_v \frac{d\theta_0}{dt} \right] - n J_m \frac{d^2\theta_0}{dt^2} \right\} \quad (11)$$

Rearrange this Eq. (11) to get a second order linear differential equation. The equation of motion of the controller is finally given by Eq. (12) as

$$\begin{aligned} J' \frac{d^2\theta_0}{dt^2} + \left(\frac{D + A K_m n K_v}{n^2 J''} \right) \frac{d\theta_0}{dt} + \left(\frac{A K_m n K_p}{n^2 J''} \right) \theta_0 \\ = \left(\frac{A K_m n K_p}{n^2 J''} \right) \theta_p \end{aligned} \quad (12)$$

Take the Laplace Transform (LT) of the above second order linear differential equation and get the transfer function of the controller as

$$\left(\frac{D + A K_m n K_v}{n^2 J''} \right) s^2 \theta_0(s) + \left(\frac{A K_m n K_p}{n^2 J''} \right) \theta_0(s) = \left(\frac{A K_m n K_p}{n^2 J''} \right) \theta_p(s) \quad (13)$$

The Transfer Function [TF] of the robot control servo system is given by Eq. (14) and is shown in the form of a block-diagram in the Fig. 7.

$$G(s) = \frac{\theta_0(s)}{\theta_p(s)} = \frac{A K_m n K_p}{s^2 n^2 J'' + (D + A K_m n K_v) s + A n K_m K_p} \quad (14)$$

This TF of the robot control system developed is of the second order form given by Eq. (15) as

$$G(s) = \frac{\omega_n^2}{s^2 + 2 \delta \omega_n s + \omega_n^2} \text{ where} \quad (15)$$

ω_n = Undamped natural frequency of the system or natural frequency of the free or sustained oscillations,

δ = Damping factor (friction),

s = Complex variable consisting of a real and imaginary part = $(\sigma + j \omega)$.

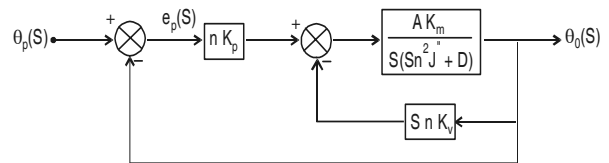


Fig. 7 Block diagram and transfer function model of the servo-controlled joint-link arm segment of the robot arm

Comparing Eqns. (14) and (15), we get

$$\omega_n = \sqrt{\frac{A k_m n k_p}{n^2 J''}} \quad (16)$$

$$\delta = \frac{D + A k_m n k_v}{2 \sqrt{A k_m n k_p n^2 J''}} \quad (17)$$

The robot control system is designed for damping factor in the range $0.4 < \delta < 0.9$, as a result of which the output always exhibits damped oscillations (under damped response) and hence, system will be stable. The problem regarding oscillation is that the structural natural frequencies of the joint-link arm segments vary inversely with the square root of the effective inertias between their minimum and maximum values.

The minimum and maximum values of the inertias are as follows : i.e., $J_{\min}$ in no-load-no-stretch and $J_{\max}$ in full-load-full-stretch positions. Both of these limiting values of inertia are obtainable from measurement. In practice, the variation of the ratio of the joint-link inertias $\frac{J_{\max}}{J_{\min}}$ within the robot

workspace may vary by the ratio 10 : 1. The experimental result of one of the joint of the robot arm when a particular value of torque given is shown below in Fig. 8.

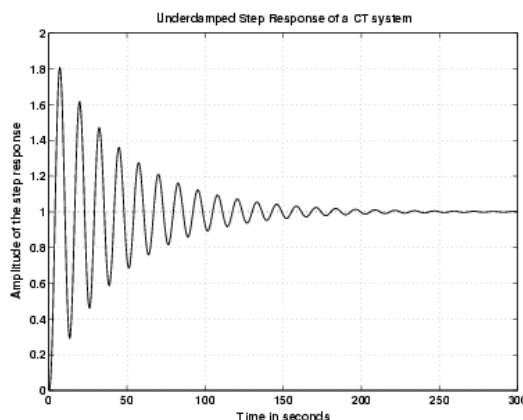


Fig. 8 Experimental result of a joint of the robot system

CONCLUSIONS

A novel design of a mathematical control model was developed for a stationary robotic system. The control model was designed taking into account all the necessary damping, feedback, backlash, hysteresis, etc.,. The successful testing was also done. The controller developed was of the second order form.

REFERENCES

- [1] Craig J, "Introduction to Robotics : Mechanics", *Dynamics & Control*, Addison Wessely, USA, 1986.
- [2] Robert J.S., "Fundamentals of Robotics - Analysis and Control", *PHI*, New Delhi.
- [3] Klafter, Thomas and Negin, "Robotic Engineering", *PHI*, New Delhi.
- [4] Fu, Gonzalez and Lee, "Robotics: Control, Sensing, Vision and Intelligence", *McGraw Hill*, Singapore.
- [5] Groover, Weiss, Nagel and Odrey, "Industrial Robotics", *McGraw Hill*.
- [6] Ranky P.G., C.Y. Ho, "Robot Modeling, Control & Applications", *IFS Publishers*, Springer, UK.
- [7] Crane, Joseph Duffy, "Kinematic Analysis of Robotic Manipulators", *Cambridge Press*, UK.
- [8] Manjunath, T.C., "Fundamentals of Robotics", Fifth edn., *Nandu Publishers*, Mumbai, 2007.
- [9] Ranky P G and C Y Ho, Robot Modeling, control and applications, IFS publishers, Springer, UK.
- [10] Asada, H., and J.E. Slotine, Robot Dynamics & Control, Wiley, NY.
- [11] Phillip Coiffette, Robotics Series (Vol. I to VIII), Kogan Page, London.
- [12] Mohsen Shahinpoor, "Robotic Engineering Text Book", *Harper & Row Publishers*.
- [13] Janakiraman, "Robotics and Image Processing", *Tata McGraw Hill*.
- [14] Richard A Paul, "Robotic Manipulators", *MIT press*, Cambridge.
- [15] Fairhunt, "Computer Vision for Robotic Systems", *CFS Pubs.*, New Delhi.
- [16] Yoram Koren, "Robotics for Engineer", *McGraw Hill*.
- [17] Bernard Hodges, "Industrial Robotics", *Jaico Publishing House*, India.
- [18] Tsuneo Yoshikawa, "Foundations of Robotics : Analysis and Control", *PHI*, India.
- [19] Dr. Jain and Dr. Aggarwal, "Robotics : Principles & Practice", *Khanna Publishers*, Delhi.
- [20] Lorenzo and Siciliano, "Modeling and Control of Robotic Manipulators", *McGraw Hill*.
- [21] Dr. Amitabha Bhattacharya, "Mechanotronics of Robotics Systems", *Kaizen Publishing*, Calculatta, India.
- [22] S.R. Deb, "Industrial Robotics", *Tata MacGraw Hill*.
- [23] Edward Kaffrisen and Mark Stephans, "Industrial Robots and Robotics", *Prentice Hall Inc.*, Virginia, USA.
- [24] Rex Miller, "Fundamentals of Industrial Robots and Robotics", *PWS Kent Pub Co.*, Boston, USA.
- [25] Doughlas R Malcom Jr., "Robotics ... An introduction", *Breton Publishing Co.*, Boston.
- [26] Wesseley E Synder, "Industrial Robots : Computer Interfacing and Control", *Prentice Hall*.
- [27] Carl D Crane and Joseph Duffy, "Kinematic Analysis of Robot Manipulators", *Cambridge Press*, UK.
- [28] CY Ho and Jen Sriwattamathamma, "Robotic Kinematics ... Symbolic Automatic and Numeric Synthesis", *Alex Publishing Corp*, New Jersey.
- [29] Francis N Nagy, Andreas Siegler, "Engineering Foundations of Robotics", *Prentice Hall*.
- [30] William Burns and Janet Evans, "Practical Robotics – Systems Interfacing, Applications", *Reston Publishing Co*, USA.
- [31] Robert H Hoekstra, "Robotics and Automated Systems".
- [32] Lee C S G, "Robotics, Kinematics and Dynamics".
- [33] Gonzalez and Woods, "Digital Image Processing", *Addison Wesseley*.
- [34] Anil K Jain, "Digital Image Processing", *PHI*.
- [35] Joseph Engelberger, "Robotics for Practice and for Engineers".

3D FINGERPRINT RECOGNITION CODE

Aneesh Rangnekar

Dept. of Electronics and
Telecommunication Engineering
Vidyalankar Institute of Technology
Wadala, Mumbai 400037
aneesh.rangnekar@vit.edu.in

Tanmay Ingle

Dept. of Electronics and
Telecommunication Engineering
Vidyalankar Institute of Technology
Wadala, Mumbai 400037
tanmay.ingle@vit.edu.in

Vikas Gupta

Dept. of Electronics and
Telecommunication Engineering
Vidyalankar Institute of Technology
Wadala, Mumbai 400037
vikas.gupta@vit.edu.in

ABSTRACT

In this paper, we've explained 3D Fingerprint Recognition code (3D-FRC) developed by the authors using efficient computational method developed for fingerprint matching - 'Minutia Cylinder Code' that uses 3D data sets (referred to as Minutiae cylinders). In biometrics and forensic science, minutiae are major features of a fingerprint, using which comparisons of one print with another can be made. (Example: ridge endings, ridge bifurcations, etc.) The cylinders are developed by taking into consideration the valid minutiae from a fingerprint and performing mathematical operations on them, keeping into consideration the relative distances, angles and radial angles. Due to the fingerprint elasticity and lack of detailed descriptions, logistic functions have to be defined to compute local similarities and to amalgamate them into a global fingerprint match score. Thorough experiments over inked and chance fingerprint databases prove the superiority of our code over current recognition techniques and demonstrated the feasibility of obtaining a very effective, computationally light and tamper-proof fingerprint recognition algorithm that uses fixed length and bit-oriented coding.

Keywords- *cylinder-code, logistic functions, local similarities, chance fingerprints, fixed length, bit-oriented*

I. INTRODUCTION

Although the modern world progressed a lot in the field of fingerprint matching techniques, ideal accuracy is yet to be achieved and there's scope of developing a robust, cross-platform & efficient fingerprint recognition system. Most of the fingerprint matching algorithms are based on counting all types of minutiae leading to unnecessary increase in the data storage and are computationally demanding and lack robustness with respect to fingerprint distortion [1]. A critical step in matching fingerprints is to reliably extract minutiae from the images. However, fingerprint images are rarely of perfect quality. They may be degraded and corrupted due to elasticity of skin and conditions under which

fingerprint is obtained, like in case of criminal databases. Thus, image enhancement techniques like Gabor filtering, segmentation, Histogram equalization, binarization and thinning [3], [4] are employed prior to minutiae extraction to obtain a more reliable estimation of minutiae locations.

3D-FRC works around a 3-dimensional cylinder construction taking into consideration the relative values of minutiae coordinates. The code uses only position and direction of minutiae, specifically ridge endings and bifurcations, leading to compact size of one fingerprint template. Based upon the quality of the fingerprint, the code's parameters can be varied to make it computationally effective. The algorithm yields precision output for any a-priori alignment except that the two templates to be matched should not have a difference greater than $\pi/2$ radians.

The rest of this paper is organized as follows. Section II deals with construction of 3-dimensional cylinders and storing information in them. Section III deals with matching minutiae locally and Section IV offers how local similarities hold globally and gives a final matching score between the two fingerprints. In Section V, large number of authors' implementation results are reported with observations and finally Section VI draws the concluding remarks.

II. GENERATION OF CYLINDERS

3D-FRC associates two 3-dimensional Boolean cylinders to each minutia, referred to as data cylinder and validity cylinder. The information in the data cylinder is computed from the spatial and directional differences between the minutia under consideration and its neighborhood minutiae, while the validity cylinder stores the validity of the corresponding cell in the data cylinder.

There are around 40 to 100 minutiae present in every fingerprint template and they are numbered $m_1, m_2, \dots, m_n$. Each minutia is a triplet $m = \{x_m, y_m, \theta_m\}$, where x_m and y_m are the minutia co-ordinates and θ_m is the minutia direction in the range $(0, 2\pi)$ with respect to

vertical. The radius of the cylinders associated to every minutiae $m = \{x_m, y_m, \theta_m\}$ is R and height is 2π , with centre of the base as (x_m, y_m) (Fig.1). The cylinders are enclosed inside cuboids and their bases are aligned according to the minutiae direction θ_m . The cuboid consists of $N_C = N_S * N_S * N_D$ cells, where N_S = number of cells along the cylinder diameter and N_D = number of cylinder sections, wherein each cell is a small cuboid by itself with $\Delta_S \times \Delta_S$ base and Δ_D height, where $\Delta_S = 2 * R/N_S$ and $\Delta_D = 2 * R/N_D$ (Fig.2). Each cell is identified with its unique (i, j, k) coordinates, where $i, j \in I_S = \{n \in \mathbb{N}, 1 < n < N_S\}$ and $k \in I_D = \{n \in \mathbb{N}, 1 \leq n \leq N_D\}$ and $p_{i,j}^m$, which represents the two-dimensional point corresponding to the center of the cells of both cylinders with indices i, j projected to the base.

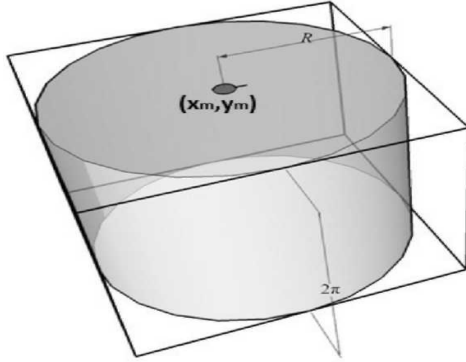


Fig. 1. The figure demonstrates the graphical representation of the cylinder enclosed inside the cuboid

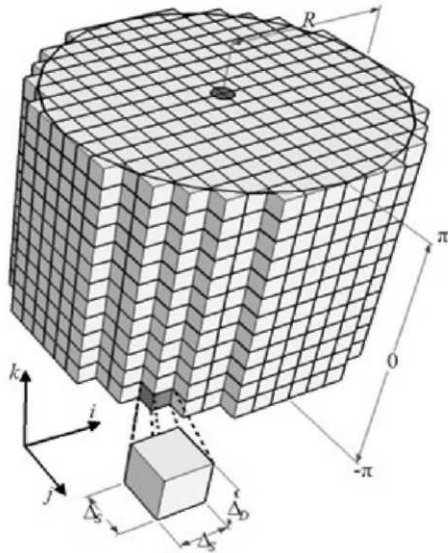


Fig. 2. The figure is an extended version of Fig.1. and shows the discretization of cuboid into numerous cells, with detailed representation of a single cell with dimensions $\Delta_S \times \Delta_S \times \Delta_D$. Note that the cuboid is rotated so that the axis 'i' is aligned to the direction of corresponding minutia.

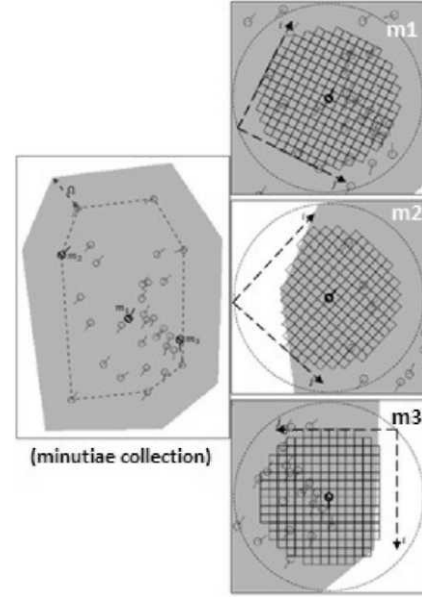


Fig. 3. A fingerprint template with the corresponding convex hull. In all cases, minutiae that lie both, within the gray region as well as dash circle, are those who contribute to cylinder values.

A. Validity Cylinder

The cells in validity cylinder store the validity of the corresponding cells in the data cylinder of each minutiae. Each cell stores a value 1 or 0 (valid or invalid) in the form of $C_{m-v} k$. A cell is considered valid when the centre p^m of the cell, projected at the base is:

- Within the distance of R from the minutiae under consideration.
- Lies inside the convex hull [5] formed by all minutiae and enlarged by an offset of Ω pixels.

The convex hull of the minutiae in the template is the smallest convex set that contains the minutiae. For instance, when $m_1, m_2, \dots, m_n$ is a bounded subset of the template, the convex hull may be visualized as the shape formed by a rubber band stretched around them. This condition is important to avoid considering portions of the cylinder that probably lie outside the fingerprint area and hence cannot contain any relevant information. Thus,

$$C_{m-v} = \begin{cases} 1 & \text{if } ds(m, p_{i,j}^m) < R \text{ and } m \in \text{convhull}(m_1, m_2, \dots, m_n, \Omega) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\text{convhull}(m_1, m_2, \dots, m_n, \Omega)$ is the convex hull of the minutiae in the template, enlarged by adding an offset of Ω pixels. 3D-FRC uses ray-casting algorithm [6] to detect whether cell corresponding to $p_{i,j}^m$ lies within the convex hull or otherwise (Fig.3).

B. Data Cylinder

The data cylinder cells store the value 0 or 1, which is calculated by summing the contributions from each valid

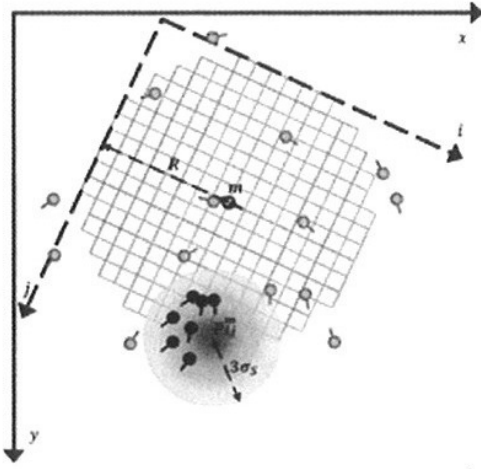


Fig. 4. Sectino of a cylinder associated to minutia m . All the minutiae involved in construction of cylinder are darkened. Note. An offset of $3\sigma_s$ is allowed.

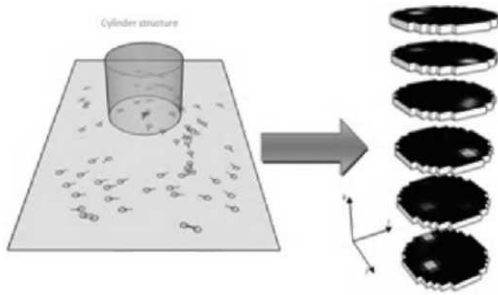


Fig. 5. A simplified case where only one minutia is taken under consideration. cylinder is graphically constructed and information is stored.

minutia m_n belonging to the neighbourhood of p^{TM} , henceforth denoted as $N_{pi,j}$. Let each cell, at height k , have an angle dp_k associated with it given by:

$$d\varphi_k = -\pi + (k - \frac{1}{2}) \cdot \Delta_D \quad (2)$$

Similar to validity cylinder, the data cylinder cells store value in the form of $C_m(i, j, k)$ (Fig.5) and is calculated as:

$$C_m(i, j, k) = \psi \left(\sum_{mt \in N_{pi,j}^{p_m}} (C_m^S(m_t, p_m^{i,j}) \cdot C_m^D(m_t, d\varphi_k)) \right) \quad (3)$$

where,

- C_m [2] represents the spatial contribution between the minutia m_t and cell (i, j, k) and is defined as a function of the Euclidean distance between m_t and p^m
- C_m [2] represents the directional contribution of m_t and is a function of dp_k (Eqn.3) and the difference between g_m and g_{mt} .

- $ip(v)$ is an unit step function given by:

$$\psi(v) = \begin{cases} 1 & ; \text{if } v \geq \mu_\psi \\ 0 & ; \text{otherwise} \end{cases} \quad (4)$$

There are around 40 to 100 minutiae that are extracted in fingerprint template. The validity of the minutiae is judged by the following factors:

- There are more than min_M minutiae in its neighborhood (Fig.5, Fig.4) given by:

$$d_S(m_t, m) \leq \{R + 3\sigma_s\}, \text{ with } m_t \neq m \quad (5)$$

They should've more than min_{VC} cells in the cylinder. Both, data and validity cylinders are valid if and only if the corresponding minutia under consideration is valid.

III. LOCAL SIMILARITY COMPUTATION

Generation of minutiae cylinders leads to creation of unique fingerprint databases that can be now referenced for matching. All databases are immune to translational, rotational and skin distortion changes since the mathematical functions working on them encrypt distances and directional differences between minutiae and are limited by the unit step function $ip(v)$.

A. Conditions for minutiae matching

When two fingerprint templates are compared, there's a need to compare corresponding pairs of minutiae. However, prior knowledge about corresponding pairs of minutiae is unknown and hence there's a necessity to compare each and every minutiae from first template to all minutiae in the database template to be matched. Thus, to save time, a few conditions are enforced on the pairing of minutiae so that unnecessary comparisons can be avoided. The conditions are as follows:

- At least min_{ME} corresponding cells between C_{m-v} of the two minutiae are valid.
- For rotational immunity, maximum allowed directional difference between the 2 minutiae is 8φ .

The first condition avoids to compare minutiae cylinders with too small valid intersection. The second condition helps to reduce the number of matchable cylinders by assuming the maximum possible rotation between the two fingerprints.

B. Similarity between minutiae

We scan each cell of validity cylinders (C_{m-v}) of both the minutiae and a temporary variable $V_{i,2}$ is created which stores the value 1 if both the corresponding cells are valid. Two temporary Boolean cylinders, each for both the minutiae under consideration, are constructed and data is stored in them as follows:

$$V_{1,2} = C_{m_{v_1}}(i, j, k) \text{ AND } C_{m_{v_2}}(i, j, k) \quad (6)$$

$$C_{m_{temp1}}(i, j, k) = C_{m_1}(i, j, k) \text{ AND } V_{1,2} \quad (7)$$

$$C_{m_{temp2}}(i, j, k) = C_{m_2}(i, j, k) \text{ AND } V_{1,2} \quad (8)$$

Norm ($\|a\|$) of a vector is defined as square root of sum of squares of all the elements of the vector. Since we're using bit-based implementation, norm of a bit-vector can be computed

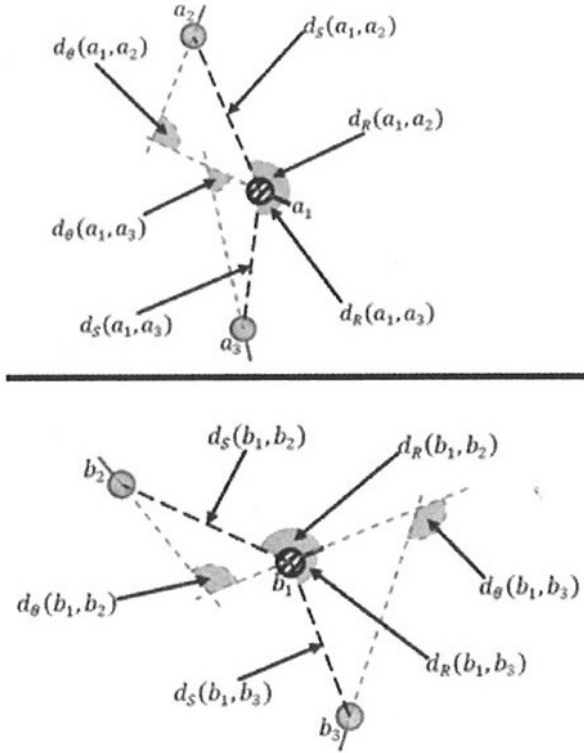


Fig. 6. An example of the global relationships considered in the relaxation procedure. The similarity between minutiae templates *a* and *b* is modified according to the compatibility between the global relationships given by the following features: i) the spatial distances (dashed black lines), ii) the directional differences (gray angles with dashed border), and iii) the radial angles (gray angles with dotted border).

by calculating the square root of number of bits with value 1. Let,

$$diff(m1, m2) = \frac{\|C_{m_{temp1}} \text{ XOR } C_{m_{temp2}}\|}{\|C_{m_{temp1}}\| + \|C_{m_{temp2}}\|} \quad (9)$$

where, $diff(m1, m2) \in [0, 1]$ quantifies the difference between the two minutiae. The numerator is the measure of difference between the two cylinders and the denominator is approximate total of the information in both the cylinders. From the equation, the minutiae are matchable if:

$$\|C_{m_{temp1}}\| + \|C_{m_{temp2}}\| \neq 0 \quad (10)$$

This condition excludes the case where a sufficiently large valid intersection of two valid cylinders does not contain any information. Thus, from this equation, the

similarity between two minutiae is computed by the following formula:

$$SIM(m1, m2) = \begin{cases} 1 - diff(m1, m2) & ; \text{ if all conditions are satisfied} \\ 0 & ; \text{ otherwise} \end{cases} \quad (11)$$

where, 0 denotes no similarity and 1 denotes maximum similarity between the two minutiae under consideration.

IV. GLOBAL CONSOLIDATION

In the previous section, we observed local similarity between the minutiae cylinders. Now, we must progress to match the two templates i.e. the overall similarity of the fingerprint.

This part is implemented using recent relaxation techniques discussed in [7]. The technique used basically modifies the local similarities based on compatibility among the minutia relationships. In simpler words, in a given pair of minutiae (*x*, *y*), if the global relationships among *x* and some other minutia in template *X* are compatible with the global relationships among *y* and the corresponding minutia in template *Y*, then the local similarity between *x* and *y* is strengthened, else weakened. This is a very good technique to reduce the score for imposters and raise that for genuine hence enabling us find the perfect match.

A. Construction of Similarity Matrix

To understand the technique, let us take two minutiae templates $X = x_1, x_2, x_3, \dots, x_n$ and $Y = y_1, y_2, y_3, \dots, y_n$ and let $SIM[r, c]$ be a two dimensional matrix; with the template having less number of minutiae as rows and the other as columns, that stores the local similarity between minutia $x_n \in X$ and $y_n \in Y$.

B. Global Strengthening via iterations

The template with the least number of minutiae is selected, in our case, template *X* having n_x minutiae. Let n_R be number of valid minutiae out of n_x minutiae. From SIM , each row is scanned to find the maximum similarity match in that row and corresponding indices are stored, resulting in generation of n_R pairs. Let $A^0 = SIM[r_t, c_t]$ be the initial similarity of pair *t*. The similarity at iteration *i* is given by:

$$\lambda_t^i = w_R \cdot \lambda_t^{i-1} + (1 - w_R) \cdot \left(\sum_{k=1, k \neq t}^{n_R} \frac{\rho(t, k) \cdot \lambda_k^{i-1}}{n_R - 1} \right) \quad (12)$$

where $w_R \in [0, 1]$ is the weighting factor. To compute iterations, there's a need to first calculate the compatibility $P(t, k)$ [2] between the two pairs of minutiae which is based on three features (Fig.6) as follows:

- d_1 , that denotes the similarity between the spatial distances of minutiae
- d_2 , that compares the directional differences between the two minutiae

- d3, that determines the difference between the radial angles between minutiae

Taking all 3 features into consideration, their product is taken and normalized by the means of sigmoid function that gives us the value of $p(t, k)$ for eqn. 12.

C. Efficiency and final score

After computing n_{rel} iterations for n_R pairs, the efficiency of the pair is calculated as:

$$\epsilon_t = \frac{\lambda_t^{n_{rel}}}{\lambda_t^0} \quad (13)$$

Let

$$n_P = \min_{n_P} + \left[(Z(\min(nx, ny), \mu_p, \tau_p)) \cdot (\max_{n_P} - \min_{n_P}) \right] \quad (14)$$

where Z is the sigmoid function with parameter $r \cdot (v - /i)$ and $[\cdot]$ is the rounding operator. The value of n_P is not an overall constant since it partially depends on the number of minutiae in the two templates. Thereafter, we've a consolidated database of n_R matching minutiae pairs from both templates, their efficiency and post-

iteration similarity value $\left(\lambda_t^{n_{rel}} \right)$. Now, top n_P pairs with highest efficiency are selected and the global score is calculated as:

$$Score(X, Y) = \sum_{n_P}^{f=1} \frac{\text{similarity value}_f}{n_P} * 100\% \quad (15)$$

which gives us the final match percentage.

TABLE 1		
Parameter Values		
Parameter	Description	Value
R	Radius of the cylinder in pixels	70
NS	Number of cells along the cylinder diameter	16
ND	Number of cylinder sections	6
	Unit step function in 4	1 100
	Offset applied to enlarge convex hull (in pixels)	50
min _{vc}	Minimum number of valid cells for cylinder to be valid	75% of the maximum number of cells in cylinder
min _M	Minimum number of surrounding minutiae for reference minutia to be valid	2
min _{ME}	Minimum number of matching elements in two matchable cylinders	60% of the maximum number of matching elements
Se	Maximum global rotation allowed between two templates	n 2
wR	Weight parameter in 12	1 2
Mp ^r p	Sigmoid parameters in 14	20 ²
maXnp	Maximum number of minutiae in 14	12
minnp	Minimum number of minutiae in 14	4
nrel	Number of relaxation iterations	G [1, 5]

V. RESULTS AND OBSERVATIONS

In this section, excerpts from experiments conducted using well-known local minutiae matching methods [8], [9], [11] are

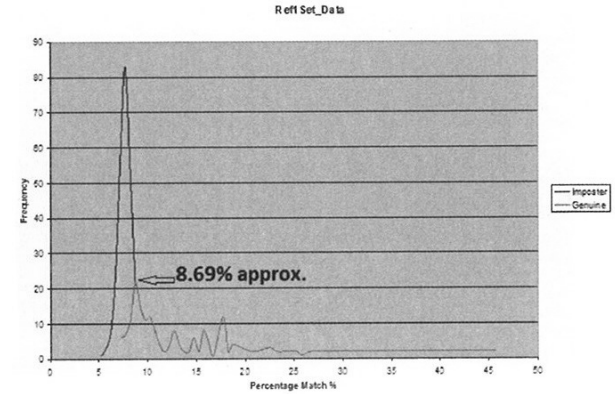


Fig. 7. The figure is a graph of cumulative score obtained by processing 200 fingerprint templates & thus a threshold of 8.69% is obtained for $n_{rel} = 3$ for our code.

reported to analyze accuracy and efficiency of 3D-FRC.

A. Results

For all databases tested, the following protocol has been adopted:

- Each template is compared against the remaining ones of the same finger to obtain the False Non Match Rate (FNMR).
- If template T_{10} is compared against T_{2a} , then to avoid correlation in the matching scores, T_{20} is not compared with T_{10} . The first template of each finger is compared against the first template of the remaining fingers in the dataset, to determine the False Match Rate (FMR).

By individually calculating the similarity for various imposters and genuine matches, the percentage threshold for genuine v/s imposter can be set at **8.69%** (Fig.7), where 8.69% to 99.95% indicates genuine fingerprint match considering $n_{rel} = 3$. The database contains inked as well as chance prints and results prove the superiority and efficiency of the algorithm.

B. Observations

From the obtained data and our experimentation, we observe:

- While finding the directional contribution of minutiae, which is a function of i) $d_k^{\wedge}$ and ii) directional difference between g_m and g_{mt} , it is observed that the contribution is high when i) and ii) are close to each other.
- While calculating global score, high efficiency is achieved for the pairs of minutiae whose similarity is substantially strengthened because of high

compatibility with other pairs, whereas pairs of local structures that initially obtained a high similarity by chance, will be penalized by the relaxation process and their final efficiency will be quite low.

```

C:\Users\Aneesh\Desktop\ATV_app\ATVTestApp\Debug\MultipleMatcher.exe
dbcurrent directory is C:\Users\Aneesh\Desktop\ATV_app\ATVTestApp
query dir C:\Users\Aneesh\Desktop\ATV_app\ATVTestApp
query filename 2_3.mnt

Aneesh Tanmay Vikas proudly present : 3D-Fingerprint Recognition Code

minutiae in template 1 are 36
minutiae in template 2 are 41
now onto similarity of 0 with 0
percentage match is : 3.748699
It took me 50.308000 seconds for n=4 computation
Press any key to continue

```

Fig. 8. A sample output taken with imposter fingerprints. As observed, database file has 36 minutiae while query has 41 minutiae & it took 50.38 seconds for computing their match at $nrel = 4$.

- By varying the values of $nrel$, we can reduce the execution time by compromising on the efficiency and vice versa (Table 2, Fig. 8)

Table – 2			
Variation in computational time with reference to $nrel$			
$nrel$	quality of fingerprints	preferable applications	time(in seconds) *
5	very low quality	criminal chance prints	300
4	low quality	criminal chance prints	35.64
3	average quality	criminal chance prints/inked & optical scanner prints	2.162
2	good quality	inked & optical scanner prints	1.324
1	very good quality	inked & optical scanner prints	1.279
* average time is specified after processing on 200 fingerprint templates, with average minutiae size of 36			

CONCLUSION

In this paper, we acquaint readers with 3D Fingerprint Recognition Code (3D-FRC): a novel minutiae-only representation and matching technique for fingerprint recognition. 3D-FRC relies on a robust discretization of the neighborhood of each minutia into a 3D cell-based structure. While in this paper we focused on the problem of robustly and efficiently matching two fingerprints, there exists a strong chance that the peculiarities of 3D-FRC can develop new effective techniques for

fingerprint indexing and template protection. Varying the sigmoid parameters to understand their contribution, using Hungarian algorithm to adaptively obtain the value of n_p and developing a robust fingerprint technique are the main targets of our future research.

REFERENCES

- [1] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Handbook of Fingerprint Recognition (2nd edition)"
- [2] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni, "Minutia Cylinder-Code: a new representation and matching technique for fingerprint recognition"
- [3] Raman Maini and Himanshu Aggarwal, "A Comprehensive Review of Image Enhancement Techniques"
- [4] Aroop Mukherjee and Soumen Kanrar, "Enhancement of Image Resolution by Binarization"
- [5] Andrew, A. M. (1979), "Another efficient algorithm for convex hulls in two dimensions"
- [6] Roth, Scott D. (February 1982), "Ray Casting for Modeling Solids", Computer Graphics and Image Processing.
- [7] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene Labeling by Relaxation Operations," Systems, Man and Cybernetics
- [8] Y. He, J. Tian, L. Li, H. Chen, and X. Yang, "Fingerprint Matching Based on Global Comprehensive Similarity"
- [9] X. Jiang and W. Y. Yau, "Fingerprint Minutiae Matching Based on the Local and Global Structures," Pattern Rec.
- [10] ognition
- [11] N. K. Ratha, V. D. Pandit, R. M. Bolle, and V. Vaish, "Robust Fingerprint Authentication Using Local Structural Similarity"
- [12] H. Chen, J. Tian, and X. Yang, "Fingerprint Matching with Registration Pattern Inspection"
- [13] L. Sha and X. Tang, "Orientation-improved minutiae for fingerprint matching"

AUTOMATION OF RAW MILL DRIVE IN CEMENT INDUSTRY USING PLC AND SCADA

Ruchi Harchandani

Department of Electrical Engineering,
Fr.CRIT, Vashi, NaviMumbai.
ruchi.harchandani@rediffmail.com

Bindu R.

Department of Electrical Engineering,
Fr.CRIT, Vashi, NaviMumbai.
rbinducrit@gmail.com

ABSTRACT

Automation of manufacturing operations helps in increasing the production, improving the product quality, enhancing the safety of the operators as their role get transformed to supervisors and decreasing the cost of production. A Programmable Logic Controller (PLC) is nowadays key component for Industrial Automation. PLC helps in making the automation more flexible. SCADA systems became popular to arise the efficient monitoring and control of distributed remote equipment. Today SCADA systems include operator-level software applications for viewing, supervising and troubleshooting local machines and process activities. Powerful software technologies are used for controlling and monitoring equipment. The PLC correlates the operational parameters to that requested by the user and monitors the system during normal operation and under trip conditions. SCADA encompasses the collection of information transferring to the central site, carrying out any necessary analysis and control on a number of operator screens or displays. Thus, Control system of drives with PLC and SCADA system has the advantages of local as well as centralized control.

Based on the above concept a control system is developed for the raw mill drive used in Cement Industry using PLC, and which can be interfaced using SCADA.

Key Words- PLC, SCADA, Cement, Raw mill

I. INTRODUCTION

The PLC is a programmable device, which executes functions such as logic, timing, counting, arithmetic operations and data manipulations. It continuously monitors the status of input devices, processes the input data as programmed, makes decisions and modifies the outputs accordingly. The program stored in the memory can be readily modified. The novelty of PLC is that it can operate in an industrial environment with temperature ranging from 0 to 60°C and humidity ranging from 0 to 90%.

The use of programmable logic controllers (PLCs) with power electronics in electric machines applications has been introduced in the manufacturing

automation. Many factories use PLCs in automation processes to diminish production cost and to increase quality and reliability. To obtain accurate industrial electric drive systems, it is necessary to use PLCs interfaced with power converters, personal computers, and other electric equipment. Input devices and output devices of the process are connected to the PLC and the control program is entered into the PLC memory. PLC controls analog and digital inputs and outputs in varying load operation of motors. PLC continuously monitors the inputs and activates the outputs according to the control program. The PLC correlates the operational parameters to the speed requested by the user and monitors the system during normal operation and under trip conditions [1-3].

SCADA systems became popular to arise the efficient monitoring and control of distributed remote equipment. Today SCADA systems include operator-level software applications for viewing, supervising and troubleshooting local machines and process activities. Powerful software technologies are used for controlling and monitoring equipment [5].

The cement manufacturing process consists broadly of mining, crushing and grinding, and grinding with gypsum. Two basic processes, the wet process and the dry process, are used for cement manufacturing. In the wet process, each raw material is proportioned to meet a desired chemical composition and fed to a rotating ball mill with water. The raw materials are ground to a size where the majority of the materials are less than 75 microns [17-19].

In the dry process, each raw material is proportioned to meet a desired chemical composition and fed to either a rotating ball mill or vertical roller mill.

II. PLC AS A SYSTEM CONTROLLER

For raw material grinding slip ring induction motor is commonly used. An overload capacity 50% for one minute may be required. Liquid resistance starters are usually used to bring it up to full speed. Three starts from cold conditions and two consecutive starts from hot condition per hour against full load are required. The starting torque is limited to 125% of the rated torque and pull out torque is restricted to 240% of rated torque [1, 2 and 11].

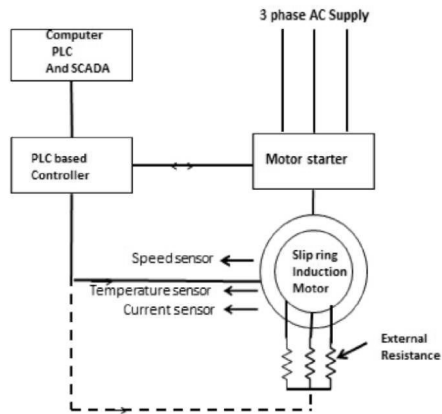


Figure 1 System block diagram for slip ring motor control

In cement industry liquid resistance starter is used. A DC motor can be used to control liquid resistance starter, slip ring induction motor and raw material grinding is controlled using PLC and SCADA Figure 1. To achieve three starts from cold condition and two starts from hot condition per hour, before switching ON, temperature of the motor is sensed using scale with parameter (SCP) block and RTD output. If the sensed temperature is beyond certain value, DC motor will be switched on, which will include the certain value of resistance in the rotor circuit to limit the starting current. A counter is used to restrict the two starts in the hot condition whose accumulator value is set to two [1, 2, and 7]. If the sensed temperature is within limit the slip ring motor is switched on directly. A counter is again used to restrict three starts in the cold condition.

Motor speed is sensed using high speed counter (HSC) or proximity sensor which is connected to one input of the PLC. A counter, timer, multiplier and move blocks of PLC is required to calculate the motor speed. Counter is connected to the proximity input to count the no. of pulses.

Speed of the motor is sensed continuously and simultaneously, rotor external resistance is adjusted using the dc motor. Interlocking is done in such a way that once required speed of slip ring induction motor is reached, status of rung containing DC motor for varying liquid resistance becomes low and the dc motor stops. Now, slip ring IM runs at required constant speed.

Motor current is measured using current transducer and analog input of the PLC. AC current is first stepped down using current transformer and then using rectifier ac current is converted to DC current. By using one resistance this DC current is converted to DC voltage which can be fed to analog input of PLC. This DC voltage is calibrated to corresponding AC motor current. Once the motor is overloaded a trip signal is given to motor through PLC.

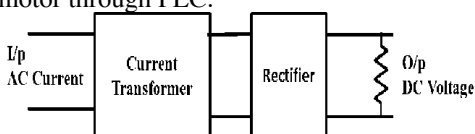


Figure 2 Block diagram of Current Transducer

Various Protections are provided in PLC ladder logic. All the faults are represented as NC contacts in PLC programme. Due to occurrence of any of the faults in the motor NC contact of corresponding fault becomes open and bit in protection rung becomes low, and in turn supply to motor gets interrupted and motor stops.

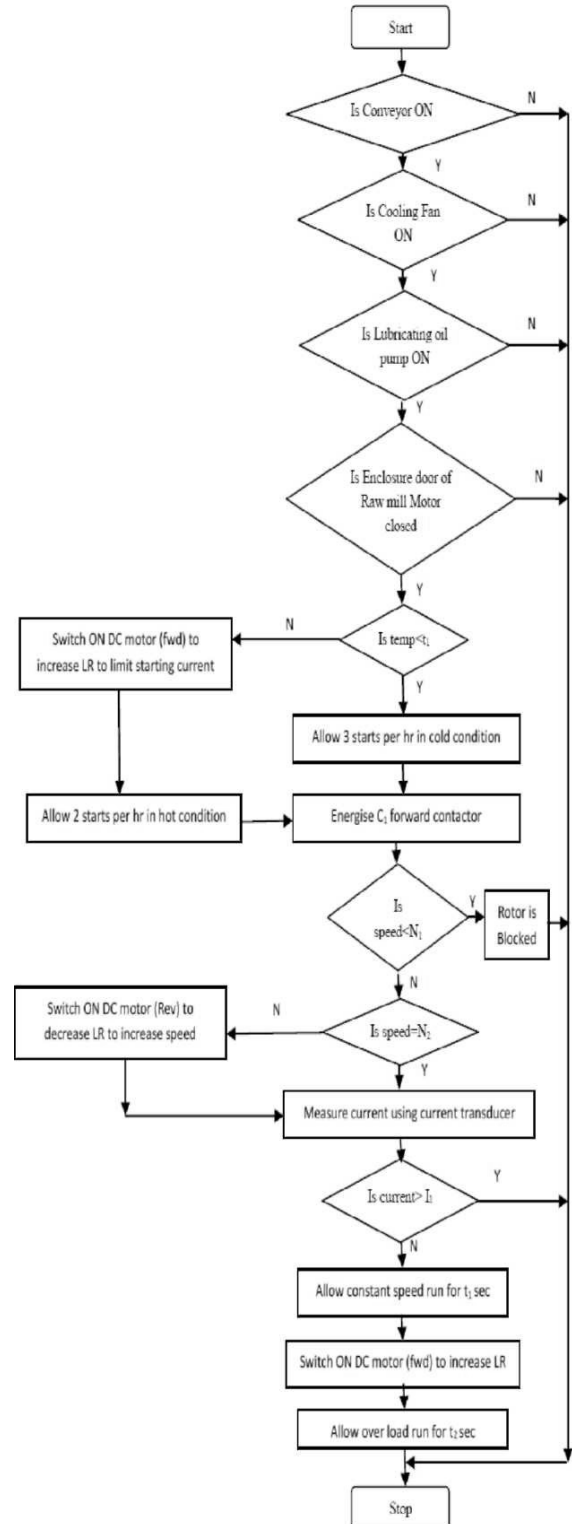


Figure 3 System flow diagram

Temperature of the motor is sensed using scale with parameter (SCP) block and RTD output. If the sensed temperature is beyond limit, a trip signal is given to stop the motor. Also various protections like single phasing, material availability, door guard, conveyor on, lubricating oil pump on, cooling fan on etc. are provided. In case any condition is not fulfilled the motor cannot be switched on even by using physical inputs.

Single phasing protection is provided in hardware. Relays are energised by 3 phase currents, and contactors of these relays are connected in series. One digital input (NC) of PLC is connected in series with these contactors. In case one of the supply lines gets disconnected either due to rupturing of fuse or open circuit in one of the three supply connections, NC contact of PLC becomes NO and motor will stop due interruption in supply.

Depending upon the availability of inputs and outputs of PLC many more protections, interlocking and indicators can be provided in ladder logic.

III. SCADA IMPLEMENTATION

SCADA graphics screen is developed as shown in Figure 4 for raw material grinding. Main Screen contains status of slip ring induction motor (ON/OFF), status of supply, status of DC motor (ON/OFF), duration of motor running, system status (healthy/unhealthy), actual motor speed, motor winding temperature, number of motor starts etc.

One rectangle is animated for ON/OFF switch of Raw Mill motor. Motor running status is animated using motor picture; it turns green in running condition of motor. System healthy condition and supply ON/OFF condition are animated using circles. Using Numerical display actual running duration of the Raw Mill motor is displayed.

Using numerical input enable, in run time, maximum temperature of the motor is changed and also the actual temperature of the motor is displayed using numerical display.

Speed limits of the motor are tagged on the SCADA screen using numerical input enable. Before switching ON the motor, speed limit can be changed in run time.

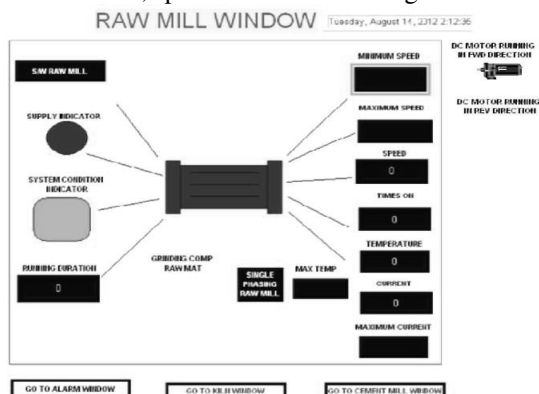


Figure 4 SCADA graphics window for Raw Mill

The actual speed of the Raw Mill motor is displayed using numerical display. Numbers of runs in hot and cold condition are displayed using numerical display blocks on the SCADA graphics windows.

A motor model is used to animate DC motor running in forward direction. Similarly another motor model is used to show DC motor running in reverse direction. Visibility animation is given to these motor models; model will appear only when motor is running.

To display Raw Mill motor current a numerical display is used on the SCADA graphics window which is tagged to floating point address of PLC where the actual motor current value is stored. Using numerical input enable, before switching ON the maximum value of motor current can be entered in run time.

Single phasing protection is provided in hardware and using digital input of PLC. In case single phasing occurs a message will display on the SCADA graphics window which is tagged to digital input.

Alarms for analog signals like temperature, motor current and motor speed is set. If the temperature of motor winding exceeds beyond limit one banner displaying temperature high will appear on the main graphics window.

Once the alarm is acknowledged, acknowledge alarm (ack alarm) button should be clicked. And if the alarm is attended it can be cleared by clicking at clear alarm button.

Similarly alarm trigger values for motor speed and motor current are set. Faulty conditions are displayed in the form of alarm banners. On these banners time of alarm, acknowledged or not, time of acknowledgement are mentioned. Alarms can be cleared separately by clicking at clear alarm or all together at once by clicking at clear all alarms.

On the main graphics window of Raw Mill, navigation buttons are provided for different windows. These are GO TO alarm window, GO TO cement kiln window, GO TO cement mill window. Using these buttons we can monitor other process in run time on the graphics window [5].

CONCLUSION

PLC can be programmed to sense, activate, and control industrial equipment and, therefore, incorporates a number of I/O points, which allow electrical signals to be interfaced.

Input devices and output devices of the process are connected to the PLC and the control program is entered into the PLC memory. PLC controls analog and digital inputs and outputs in varying load operation of motors.

PLC continuously monitors the inputs and activates the outputs according to the control program.

Control system of drives with PLC and SCADA system has the advantages of local as well as centralized control. The monitoring control system of the motors driven by starters and controlled by PLC and SCADA proves its high accuracy in speed regulation control and protection at variable-load operation.

REFERENCES

- [1] Vedam Subrahmanyam *"Electric Drives Concepts and Applications"* Tata McGraw-Hill, sixteenth reprint
- [2] S. K. Pillai *"A First course on Electrical Drives"* New age international Publishers Second edition
- [3] Nathan Schachter Cmentec Engineering Ltd, *"Energy Efficient Speed Control Using Modern Variable Frequency Drives"*
- [4] The ABB website www.abb.com/drives *"Case note drives reduce energy consumption and improve productivity of cement plant"*
- [5] Cristina Anita Bejan, Mihailacobi and Gheorghe-Daniel Andreescu *"SCADA Automation System Laboratory, Elements and Applications"*
- [6] Maria G. Ioannides, Senior Member, IEEE, *"Design and Implementation of PLC-Based Monitoring Control System for Induction Motor"*
- [7] Cristina Anita Bejan, Mihailacobi and Gheorghe-Daniel Andreescu, Member, IEEE *"Politehnica"* University of Timisoara, Dept. of Automation and Applied Informatics, Timisoara, Romania *"SCADA Automation System Laboratory, Elements and Applications"*
- [8] Mini Sreejeth, Pannod Kumar and Madhusudan Singh, 2010 IEEE, *"Development of Supervisory Control for Distributed Drives System"*
- [9] S P Deolekar Handbook for *"Designing Cement Plants"*
- [10] User manual Allen Bradley *"Micrologics 1400"*
- [11] The carrier website www.carrier.com *"Operation and Application of Variable Frequency Drive (VFD) Technology"*, Carrier Variable Speed Screw White paper
- [12] Gary Dunning *"Introduction to Programmable Logic Controller"* Cengage Learning Third edition
- [13] I. CQolac, MemIEEE, R. Bayindir, MemIEEE, A. Bekta, I. Sefa, BSc, MSc, PhD *"Protection of Induction Motor Using PLC"*
- [14] The Siemens website www.siemens.com *"Automation and Drives"*
- [15] G. K. Dubey *"Fundamentals of Electrical Drives"* Narosa Publication, second edition
- [16] Austine Hughes *"Electric Drives and Motors, Fundamental Types and Application"*
- [17] MahendraRane, DhruvChhabra and Rangan Banerjee, *"Industrial DSM for Indian Power"*
- [18] Cement Manufacturer's Association 2003; Worrell 2004 [19] Confederation of Indian Industry - Energy Management Cell 2002

ROBOTIC ARM CONTROLLED BY POTENTIOMETRIC SYSTEM

Rashmi Gaba

Department of Biomedical Engineering
Thadomal Shahani Engineering College
Bandra (West), Mumbai-400050
rashmi.gaba@gmail.com

Feba Susan Sam

Department of Biomedical Engineering
Thadomal Shahani Engineering College
Bandra (West),
Mumbai-400050

Annette Cheenath

Department of Biomedical Engineering
Thadomal Shahani Engineering College
Bandra (West),
Mumbai-400050

Mita Bhowmick

Department of Biomedical Engineering
Thadomal Shahani Engineering College
Bandra (West), Mumbai-400050
mitabhowmick@gmail.com

ABSTRACT

A Robot is a mechanical or virtual intelligent agent that can perform tasks automatically or with guidance. A robotic arm consists of several joints connected together by linkages. Robotic surgery is the use of robots in performing surgery. Robotics in surgery have not only reduced the workload on surgeons in terms of stress free operative procedure and decreased fatigue, but have also helped in reassuring the patient in terms of a faster and more precise surgical procedure. The accuracy and the efficiency of surgeries have improved greatly because of the application of robotics in the medical field. The paper deals with the control of the robotic arm using a potentiometric control system. The potentiometric system gives control signals to the arm which are processed by a microcontroller. The robotic arm has 5 degrees of freedom which are base rotation, shoulder rotation, elbow rotation, wrist pitch and wrist roll.

Keywords: *Robot, Robotic Arm, Robotic Surgery, Haptic Technology, Potentiometric System.*

I. INTRODUCTION

A robot can be defined as a reprogrammable, multifunctional manipulator which is designed to move material, parts, tools, or specialized devices through various programmed motions to perform a variety of tasks. A robotic arm is a mechanical device that travels specifically in a designed pattern, with sensors ensuring that all moments are exactly of the similar pattern [1].

Robotic surgery uses robotic systems to aid in surgical procedures such as minimally invasive surgeries. In case of robotically assisted minimally invasive surgery, the surgeon uses one of the two methods to control the instruments; either a direct tele-manipulator or by computer control [2]. In addition to cost efficiency, robotic surgery has several other advantages over conventional surgery including enhanced precision and reduced trauma to the patient. Robotic telesurgery enhances the dexterity and sensation of regular and minimally invasive surgery [3]. By using robotic manipulator it is possible to perform remote surgery for

care in rural area where specialty care is unavailable. The technology being used for controlling the arm moments is known as haptic technology.

The da Vinci system is the greatest example of robotic surgery system which works on the principle of minimally invasive procedure where small incisions are made causing a reduced blood loss, reduced post surgical trauma to the patient and an improved recovery period [4]. The da Vinci System consists of a surgeon's console that is typically in the same room as the patient with four interactive robotic arms controlled from the console. Three of the arms are for tools that hold objects or instruments. The fourth arm is for an endoscopic camera with two lenses that gives the surgeon full stereoscopic vision from the console [5]. These robotic systems enhance dexterity and greatly enhance the surgeon's ability to manipulate instruments and thus the tissues [6]. These systems can scale movements so that large movements of the control grips can be transformed into micromotions inside the patient [7].

Haptic technology or haptics is a tactile feedback technology that takes advantage of a user's sense of touch for the enhancement of the remote control of machines and devices[8]. Haptics enable the manipulation of objects in the environment in the natural and effective way. Robotic surgery with haptic feedback would enable surgeons to identify different tissue structures, preventing tissue damage and also ensuring correct suture placement, thus reducing the completion time of surgery[9,10].

In order to make telesurgery a reality, expert surgeons have worked with the telecommunication industry to develop secure, reliable, high-speed transmission of data over large distances with imperceptible delays. A robotic surgery was successfully conducted where a patient, situated in a remote location, was operated for laparoscopic cholecystectomy. The console being controlled by the surgeon situated in some other location, gave control signals to the robotic arm to perform the surgery on the patient with the help of telesurgery [11].

The research deals with the design and implementation of the robotic arm which has 5 degrees of freedom including: base rotation, shoulder rotation, elbow rotation, wrist pitch and wrist roll and six motors are used to achieve these movements. The gripper has two claws, and it is attached to the sixth motor. The movements of the robotic arm are controlled by moving the user's arm using haptic technology.

II. SYSTEM DESIGN

The designed and developed robotic arm is a 5 axis robotic arm with servo gripper. It consists of six servo motors attached together by joints and linkages. The arm is designed such that the metal links are attached by each servo motor joint. The gripper consists of 2 claws. The arm uses metal gear servo motors with 15kg/cm torque and 2 servo motors with 7kg/cm torque. The robotic arm is designed by high grid machine/injection molded aluminum alloy [12].

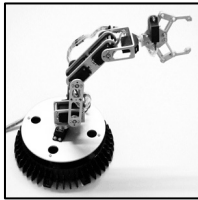


Figure1: Designed Robotic Arm

The control assembly consists of a potentiometric system which is given to the controller used to control the six servo motors attached to the robotic arm [13]. The potentiometers act as a joystick where the user makes the necessary movements and the arm depicts the changes.

Arduino is an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software. Arduino can sense the environment by receiving input from a variety of sensors and can affect its surroundings by controlling lights, motors, and other actuators. The hardware consists of a simple open hardware design for the Arduino board with an Atmel AVR processor and on-board input/output support. The software consists of a standard programming language compiler and the boot loader that runs on the board. Arduinoduemilanove used in this project is a microcontroller board based on the ATmega 328. It has 14 digital input/output pins, 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. [14,15].

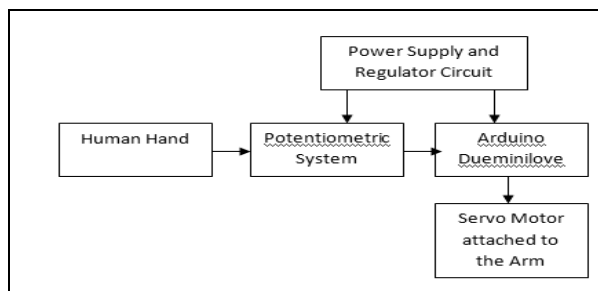


Figure 2: Block Diagram of Robotic Arm Controlled By Potentiometric System

A. Potentiometric transduction system:

The shaft of the potentiometers is attached to the assembly. Thus its position is constantly altered due to the changing hand movements. Since a voltage supply is given to the potentiometers, when the position changes there is also a voltage change. This changing analog voltage signal is fed to the arduino board for its digital conversion.

B. ArduinoDuemilanove:

The potentiometer are connected to the input side of the board. Arduino processes the signals internally, performs analog to digital conversions and gives corresponding PWM output signals.

C. Servo motors attached to the Arm:

The servo motors on the robotic arm receive the corresponding PWM signals from the arduino board and move according to the changing hand movements.

D. Regulated Supply:

It consists of a voltage regulator IC and two coupling capacitors. The regulator IC regulates the input supply of 9V to 5V which is then given to the servo motors and the potentiometers.

The flow chart gives the working of the program to control the servo motor by a potentiometric assembly using the arduino duemilanove controller.

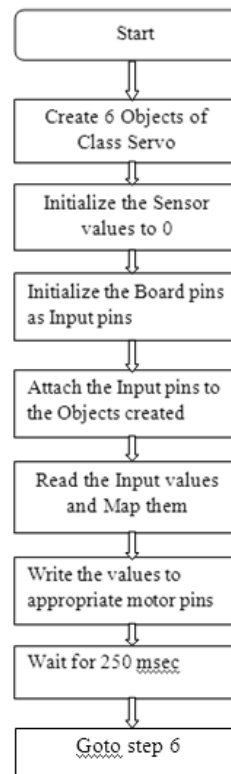


Figure 3: Flow Chart

III. RESULTS

The robotic arm was successfully controlled by using a potentiometric assembly and the arm could perform various motions such as pick and place of an object.

Figure 4(a) shows the original position of motor 1 and Figure 4(b) shows the final position of motor 1 which is achieved by moving the corresponding potentiometer.



Figure 4(a):Original position



Figure 4(b): Base Rotation

Figure 5 shows the final position of motor 4. Figure 6 shows the final position of motor 5



Figure 5: Shoulder Rotation

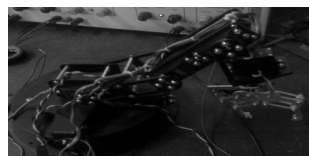


Figure 6: Elbow Rotation

Figure 7 and 8 shows the final position of motor 6.



Figure 7: Wrist Roll



Figure 8: Wrist Pitch

IV. CONCLUSION

Medicine is a field where extreme precision is necessary and the margin for error is very less. A robotic arm will enable the surgeons to perform many complicated operations with less fatigue. Telesurgery is the latest advancement in robotic surgery. Doctors can perform operations in rural communities or dangerous settings without having to be present.

In this project a robotic arm has been designed and controlled using haptics. The potentiometric transduction system gives control signals to the arm and the arm mimics the movement of the potentiometric assembly.

IV.FURTHER SCOPE

A number of advancements are possible for the current project.

- A. Feedback control: The arm currently does not have any feedback mechanism to control its movements. Pressure sensors can be attached to the arm to measure the amount of pressure applied to the patient and give a feedback to the robotic arm to limit it if it exceeds a certain preset value. Thus, it can be used to restrict the movement of the robotic arm to the required extent.
- B. Wireless transmission of control signals: The current stage of the project has input signals given to the arm through wired connections. Further improvement can be made to ensure wireless transmission of the control signal. This will be helpful in emulating telesurgical operation.

REFERENCES

- [1] Sackier, J. and Wang, Y. Robotically Assisted Laparoscopic Surgery: From Concept to Development, *Surg. Endosc.*, 8:63-66, 1994.
- [2] Bruno Siciliano, Lorenzo Sciavicco, Luigi Villani. *Robotics: Modelling, Planning and Control*, Advanced Textbooks in Control and Signal Processing, Springer 2011.
- [3] Joel Leroy, Michelle Smith. *Transcontinental Robot Remote Telesurgery-Feasibility and Potential Applications*, V.23(4), 2002.
- [4] Sung GT, Gill IS. Robotic Laproscopy surgery: A comparison of DA Vinci and Zeus systems, *Urology*, 58(6):893-898, 2001.
- [5] Rosen M, B. Hannaford, M Macfarlane, M. Sinnanan. Force Controlled and teleoperated Endoscopic grasper for minimally invasive surgery, *IEEE transactions on Biomedical Engineering*, 1999.
- [6] H. Fischer, B. Neisius, R. Trapp, "Tactile Feedback for Endoscopic Surgery", In "Interactive Technology and the New Paradigm for Health Care", 1995.
- [7] Abdulmotaleb El Saddik, Mauricio Orozco, Mohamad Eid, Jongeun Cha. *Haptics Technologies: Bringing Touch to Multimedia*, Springer Series on Touch and Haptic Systems, 2011.
- [8] Allison M. Okamura. Haptic Feedback in Robot Assisted Minimally Invasive Surgery, *Urology*, V -19, I-1, 2009.
- [9] Anthony R. Lanfranco, BAS, Andres E. Castellanos, MD, Jaydev P. Desai, PhD, and William C. Meyers, MD, "Robotic Surgery-Current Perspective, *Ann Surg.* V.239(1); Jan 2004
- [10] Wikipedia the free encyclopaedia, http://en.wikipedia.org/wiki/Da_Vinci_Surgical_System.
- [11] Kim VB, Chapman WH, Albrecht RJ, et al. Early experience with telemanipulative robot-assisted laparoscopic cholecystectomy using Da Vinci. *Surg Laparosc Endosc Percutan Tech.* 2002;12:34-40.
- [12] Rodney A Brooks. *New Approaches to Robotics*, Science, 1991.
- [13] Terry L. M. Bartelt, "Industrial Control Electronics: Devices, Systems, and Applications".
- [14] Arduino, (2004), "Arduino- An Introduction", <http://arduino.cc/en/> Dec 2011. Shiffman, Daniel, September 23, 2009.
- [15] Wikipedia the free encyclopaedia, 2010, "Arduino", <http://en.wikipedia.org/wiki/Arduino>, March 2012.

DESIGN AND IMPLEMENTATION OF PMBLDC MOTOR FOR SOLAR POWERED PUMPS

Sreedevi S Nair

Assistant Professor

Department of Electrical Engineering
Fr.C.Rodrigues Institute of Technology
Vashi, NaviMumbai
sreedevi_nair@rediffmail.com

Mini Rajeev

Associate Professor

Department of Electrical Engineering
Fr.C.Rodrigues Institute of Technology
Vashi, NaviMumbai
minirajeev1@yahoo.co.in

ABSTRACT

This paper presents the design and implementation of a Permanent magnet brushless direct current (PMBLDC) motor for solar powered water pumping application. The design of the entire system consisting of PV panel, dc-dc converter, inverter, pump and controllers are discussed in detail. The hardware implementation consists of a three phase inverter, 50W PMBLDC motor and a microcontroller (PIC 16F877A) for controlling the commutation sequence and thereby the operation of the motor. The experimental results are illustrated in the paper.

Index Terms - PV panel, Boost Converter, PMBLDC, water pumping.

I. INTRODUCTION

Solar powered pumps are the best choice for water pumping, in view of energy conservation and are widely used as an alternative energy source specifically for remote locations. Usually inverter fed ac motors or dc motors which are fed by power electronic converters are used for solar powered water pumps. But conventional dc motors suffer from disadvantages such as frequent maintenance, high cost and size and non-availability of high power ratings [1]. Hence there is a growing trend among the pump manufacturers to use them with Permanent magnet brushless DC motors (PMBLDC) for higher efficiency and low maintenance. However, the cost and complexity of these systems is significantly higher. The features of high operating efficiency, brushless construction, maintenance free operation, and increasing awareness about energy conservation have given a scope to employ PMBLDC motor in water pumping application operated by PV array.

Solar powered pump is a pump running on the power of the sun which is environment friendly and economical in its operation compared to pumps powered by an internal combustion engines. The different parts of a solar powered pump are the actual fluid pump that actually pumps liquids under pressure, the controller that adjusts

speed and output power, according to input from solar panels, the engine which is the electric motor, and the energy source which are the solar panels.

Two types of pumps are commonly used for water pumping applications: positive displacement and centrifugal [2]. Positive displacement types have the characteristic of being able to pump well at deeper pumping depths, but the flow rate is restricted. Centrifugal pumps have higher flow rates and efficiency but they are not good for pumping from deeper depths. Centrifugal pumps have added advantage that their characteristics are properly matched with the output of the solar panels. Hence centrifugal pumps are used for water pumping applications.

Fig.1 shows the block diagram of the solar water pumping system using PMBLDC motor. The solar water pumping system consists of a PV array, DC/DC Converter, inverter, and a PMBLDC motor coupled to a centrifugal pump. A DC/DC converter is used to interface the PV output with the inverter as shown in Fig.1. The PV output varies with changes in temperature, insolation etc. The DC/DC converter maintains the output dc voltage constant regardless of the variations in the PV output voltage. Fig. 2 [3] shows the motor and its drive considered for implementation. Three inbuilt Hall Effect sensors were used to sense the rotor position. This paper presents the design of the entire system and hardware implementation of the inverter and its control circuit of the PMBLDC motor.

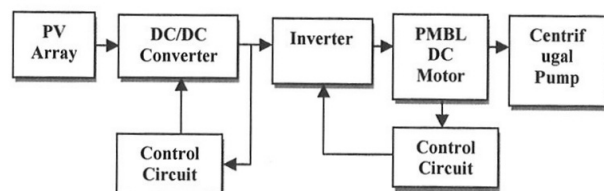


Figure 1. Block diagram of the system

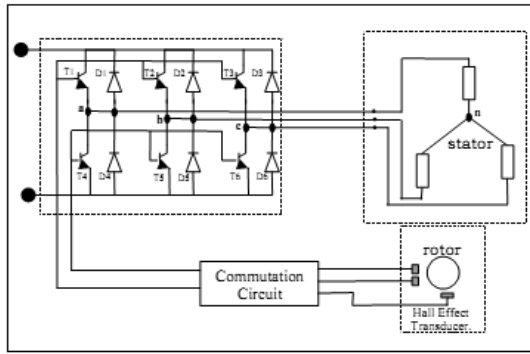


Figure 2. BLDC motor and its drive

II. BOOST CONVERTER

The DC/DC converter used is a boost converter, which is chosen based on the ratings of the PV array and the inverter voltage. Boost converter configuration shown in Fig.3 [4] steps up the PV output voltage to the voltage required by the inverter. Irrespective of the variations in the PV output voltage, the boost converter output voltage is kept constant by a control circuit which incorporates a PI controller. 10% of the input current is considered as ripple in inductor current (ΔI) and 1% of the output voltage V_o is taken as ripple in the output voltage (ΔV_o) for the boost converter [5]. A switching frequency of 20 kHz is chosen for the design and a MOSFET of appropriate rating is selected as the switch (S).

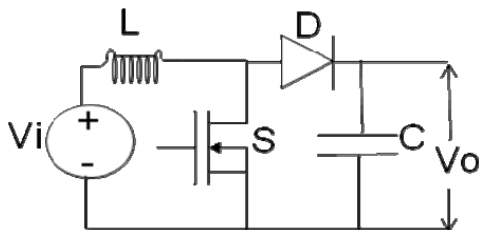


Figure 3. Boost converter circuit

$$L \geq V_i * D * T_s / (2 * \Delta I) \quad (1)$$

$$C = D * I_o * T_s / \Delta V_o \quad (2)$$

Here T_s and D are the switching time period and duty cycle of the switch S used in the converter. V_i is the output voltage of the PV array and V_o is the inverter input voltage. A value of $63\mu\text{H}$ of inductor and $1000\mu\text{F}$ capacitor was designed for this work based on "1" and "2" [5].

III. PMBLDC MOTORS

In a conventional DC motor, current polarity is altered by commutator and brushes. In the brushless DC motor, polarity reversal is performed by power transistor switching in synchronization with the rotor position. To accomplish this, the input of PMBLDC motor is

connected to inverter. Inverter is designed in such a way that, its output frequency is a function of instantaneous rotor speed and its phase control will correspond to actual rotor position [6].

Typically a BLDC motor is driven by 3 phase inverter with six step commutation as shown in Fig.4 [7]. Here the conducting interval of each phase is 120° electrical angle. In order to produce maximum torque, inverter should be commutated every 60° , so that current is in phase with the back emf. Commutation timing is determined by the rotor position and the sequence of commutation is retained in the proper order so that the inverter performs the function of brush and commutator in the conventional DC motor to generate a rotational stator flux. At one time instant, only 2 out of 3 phases are conducting current, and one winding is floating. Since input is dc, BLDC motor is a good choice for solar powered applications. Hence BLDC motor was chosen for this work.

PMBLDC motors find applications in diverse fields such as domestic applications, automobiles, transportation, aerospace equipment, sound equipment and healthcare equipment ranging from microwatt to megawatts. Advanced control algorithms and ultrafast processors have made PMBLDC motors suitable for position control in robotics, high precision servos, speed control and torque control in various industrial drives. PMBLDC motor's main disadvantage is higher cost, which arises from the fact that it requires electronic controllers to run [8].

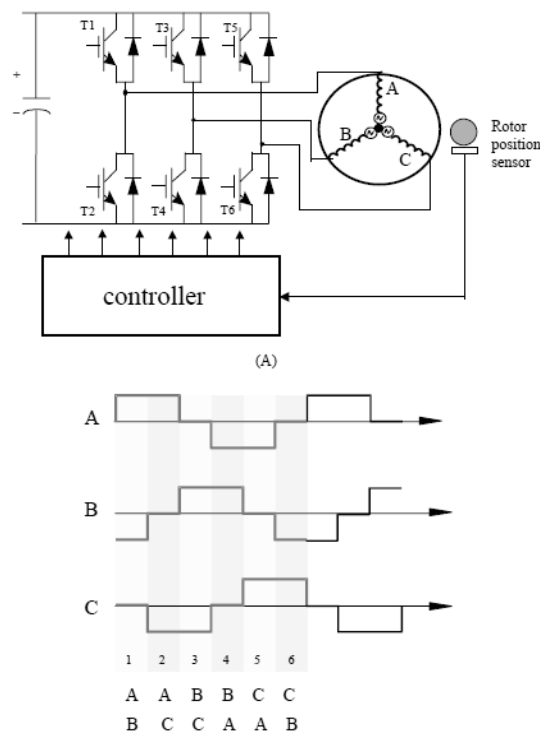


Figure 4. Driver of BLDC motor with its 3 phase current conducting intervals

IV. DESIGN DETAILS

Design aspects of various components of the system shown in “Fig. 1,” are discussed below.

A. Water Requirement

The first step in designing a solar-powered water pump system is to determine the overall water required for the operation.

Amount of water to be pumped = 10 m³/day.

B. Centrifugal Pump selection

Pump output or hydraulic horse power (hp) is the liquid horsepower delivered by the pump and is calculated by “3”

$$\text{Hydraulic power (hp)} = \frac{Q * TDH * \rho * g}{1000} \quad (3)$$

In “3” Q is the flow rate in m³/s, TDH is the total dynamic head in m, ρ is the density of water in kg/m³, and g is the acceleration in m/s². Table I shows data used for selecting the centrifugal pump.

TABLE I Data used for pump selection

Sl. No.	Parameters	Values
1	Q	$1.15 * 10^{-4}$
2	TDH	40
3	ρ	1000
4	g	9.8
5	Hydraulic power	0.045 hp = 34W

So the rating of the centrifugal pump used is selected as 0.05hp.

C. Rating of BLDC motor and inverter

Pump shaft power (P_s) is the actual horse power delivered to the pump shaft and can be calculated as follows:

$$\text{Pump shaft power } (P_s) = \frac{\text{Hydraulic power (hp)}}{\text{pump efficiency } (\eta_{\text{pump}})} \quad (4)$$

Considering pump efficiency as 69%, the pump output power is 48 W from “4”. This much power has to be supplied by the BLDC motor. So the rating of the motor selected for this application is 50 W. Considering the efficiency of the motor as 89%, the inverter selected has a power rating of 60W.

D. PV panel sizing

PV panel sizing depends on the total system load, which means determination of total wattage of the

load. Considering the efficiency of the inverter as 90% and boost converter as 90%, the PV panel should supply,

$$\text{PV Array size} = \frac{60}{0.9 * 0.9} = 74 \text{ W}$$

Considering the PV panel operating factor of 0.75 and mismatch factor of 0.85,

$$\text{PV Array size} = \frac{74}{0.75 * 0.85} = 105 \text{ W.}$$

Twenty two 5Wp, 6V solar panels needs to be used for this application. Since MPPT is not used, the current output from a single panel is taken to be 80% of the short circuit current which is 0.8 A.

V. HARDWARE IMPLEMENTATION

To verify the design, the entire circuit has been simulated in MATLAB-SIMULINK environment [9]. For the hardware implementation, 50W, 3000 rpm PMBLDC motor was used. Input to the motor was given from dc supply. The details of the hardware implementation of the inverter and the control circuit of the PMBLDC motor system used for water pumping is discussed here. The control circuit consists of the PIC 16f877a, the three driver ICs 2110, the six MOSFET switches IRF 260N and the 50W PMBLDC motor. This motor can be used for pumping 10m³ of water per day.

A. PMBLDC motor

The manufacturer’s specification of the PMBLDC motor selected for implementation is given in Table II.

TABLE II. SPECIFICATIONS OF PMBLDC MOTOR

Sl. No:	Parameters	Values
1	Motor Type	3 phase 12V delta connected PMBLDC motor
2	Rated Speed	3000 rpm
3	No: of poles	16
4	Stator resistance	84 m Ω
5	Stator Inductance	277 μ H
6	Voltage Constant	3V/Krpm

B. Three Phase MOSFET Bridge

A three phase full bridge arrangement using MOSFETs is chosen for the power interface between the motor and the controller. IRF 260N is chosen for driving the 12V PMBLDC motor.

C. Gate Drive

The gate drive forms the interface between the microcontroller and the power MOSFETs. The gate drive has two purposes. First it buffers the gate signals generated by the microcontroller. The second purpose is to generate the gate voltages required to activate the toside transistor.

The gate driver chosen for the application is IR2110 since it has got high output current capability to drive these MOSFETs [10].

D. BootStrap Capacitor Design

One of the most widely used methods to supply power to the high-sided drive circuitry of a gate driver IC is the bootstrap power supply. The bootstrap power supply consists of a bootstrap diode and a bootstrap capacitor [11,12]. The expression to find out the bootstrap capacitor is given by "5".

$$C \geq \frac{2 \left[2Q_g + \frac{I_{qbs}(t_{max})}{f} + Q_{ls} + \frac{I_{cbs}(t_{leak})}{f} \right]}{V_{CC} - V_f - V_{LS} - V_{MIN}} \quad (5)$$

Where,

Q_g , Gate charge of High Side MOSFET = 234 nC

I_{qbs} , Quiescent Current for high side drive circuitry = 230nA

Q_{ls} , Level Shift Charge required per cycle = 5nC

I_{cbs} , Bootstrap capacitor leakage current = 50μA

f , Frequency of operation = 400Hz

V_{CC} , Supply voltage, 12V

V_f , Forward voltage drop across the bootstrap diode = 1V

V_{LS} , Voltage drop across low side FET = 1.1v

V_{MIN} , The minimum voltage between V_b and V_s = 9.7V

The minimum value of capacitor is found out to be 10μF from "5" and the actual value selected is 22 μF.

E. Microcontroller

A PIC 16F877A controller is chosen for the task of performing the logic operation required to commute and control the motor. It is a 40-pin DIP which has got an operating frequency of 20MHz. Port E is used as the input sensor port and port C is used as the output port. Fig. 5 shows the entire hardware setup.

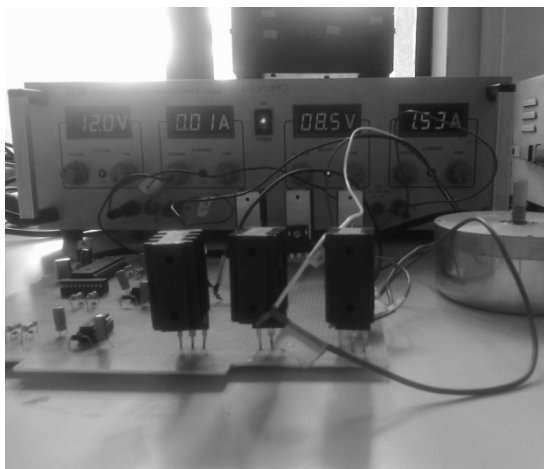


Figure 5. Full Hardware setup.

VII. EXPERIMENTAL RESULTS

The results of the hardware implementation of the PMBLDC motor is discussed here. The output of the inbuilt hall effect sensors were given to PIC controller. The controller produces PWM signals which were given to MOSFETs of 3Φ inverter.

The hall effect sensors which were there in the motor was tested by using pull up resistors of 1.2KΩ. Fig.6 shows the hall sensor outputs. The three phase inverter is tested with a three phase star connected resistive load of 50W, 50Ω and Fig.7 shows the three phase output voltage signals V_a , V_b , and V_c obtained.

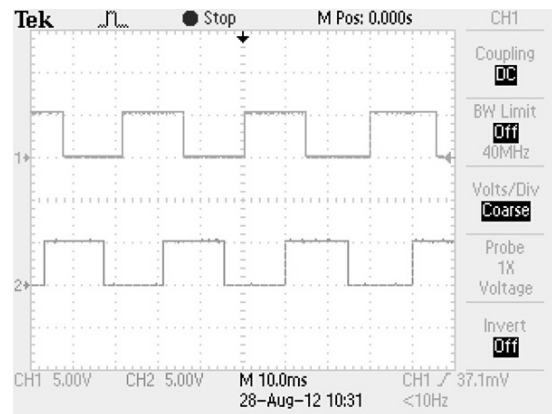


Figure. 6 Hall Effect sensor outputs

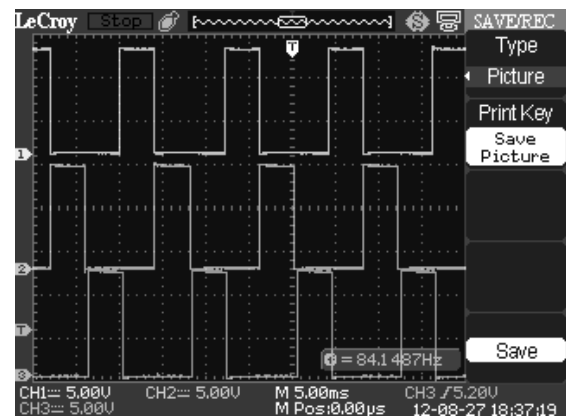
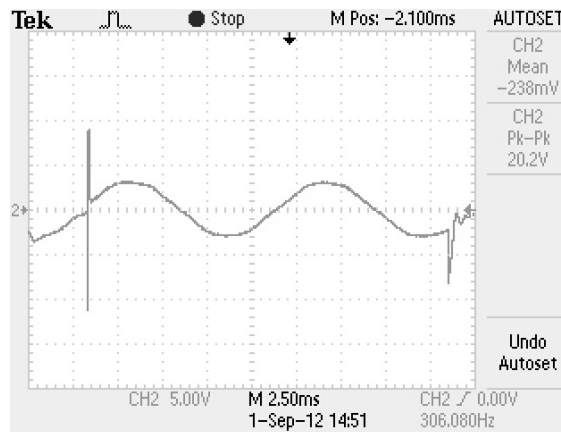


Figure 7. V_a , V_b , and V_c signals

After testing the three phase inverter, the entire hardware was tested with the BLDC motor. Fig. 8 shows the line voltage V_{ab} and Fig.9 shows the stator current i_a observed in DSO. Similarly, stator currents i_b and i_c which will be phase shifted by 120° from each other.

Figure. 8 Line Voltage V_{ab} Figure. 9 Stator current i_a

One common method of displaying motor characteristics graphically is by using speed- torque curves. From the motor data sheet, it is found that no load speed of the motor at 12 V is 3000 rpm and the stall torque is 2 N-m. One set of readings are taken with motor in loaded condition and the armature currents are measured. The proportionality constant for the relationship between motor speed and motor torque is the slope of the torque vs. speed curve. Fig.10 shows the speed vs torque characteristics of PMBLDC motor.

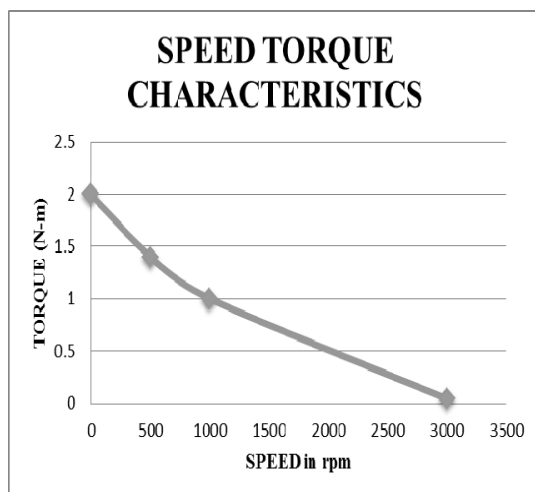


Figure 10. speed vs torque characteristics of PMBLDC motor

CONCLUSION

This paper presents the design of the solar powered water pumping system and hardware results of a PMBLDC motor used for pumping 10 m³ of water per day. Solar powered pumps are highly reliable and are often chosen because they offer the lowest life cycle cost especially when used in remote places where it is expensive to take grid electricity.

Nowadays the PMBLDC motors are the latest choice of researchers due to their high efficiency, maintenance free operation and compact size. Advancements in power electronics and digital signal processors have added more features to these motor drives to make them more prevalent in industrial installations.

REFERENCES

- [1] J. APPELBAUM, "Starting and Steady State Characteristics of DC Motor Powered by Solar Cell Generator", IEEE Trans. on EC, Vol. 1, No. 1, 1986, pp: 17-25..
- [2] Kyocera Solar Inc. Solar Water Pump Applications Guide 2001 (downloaded from www.kyocerasolar.com)
- [3] Dr. A. Moussi and T.Amal. "An Improved Efficiency PMBLDC Motor PV Pumping System". LARHYSS Journal, Nov 2001.
- [4] Mini Rajeev, Sreedevi .S. Nair, and Seema Jadhav, "Design Considerations of a Solar Powered Street Light for Stand -Alone PV Systems" CCPE 2010
- [5] Ned Mohan, "Power Electronics, Converters, Application and Design 3rd Edition."
- [6] T.J.Sokira and Jaffe, Brushless DC Motors: Electronic Commutation and Control, Tab Books USA, 1989.
- [7] Sreedevi S Nair, "A Survey of PV powered BLDC motor for Water Pumping", Zenith 2010.
- [8] S.L Putta Swamy, Bhim Singh, B.P Singh and S.Murthy, "Experimental Investigations on a PMBLDC motor fed by PV Array for Water Pumping", Journal of Solar Energy Materials and Solar cells, Vol 36, 1995, pp: 187-200.
- [9] Sreedevi S Nair, Mini Rajeev, "Design and Simulation of PV Powered PMBLDC motor for water pumping" NCNTE 2012.
- [10] Microchip Application Note AN 799.
- [11] International Rectifier Application Note AN 1123 "Bootstrap Network Analysis".
- [12] Fairchild Semiconductor Application Note AN 9052 "Design Guide for Selection of Bootstrap Components".

FACE RECOGNITION USING BLOCK SELECTION OF WAVELET TRANSFORM

Harshada Thakar

Biomedical Engineering Department
Thadomal Shahani college of engineering,
Mumbai University
harshadathakar@gmail.com

Soniya Kamath

Biomedical Engineering Department
Thadomal Shahani college of engineering,
Mumbai University
kamathsoniya@yahoo.com

Aparna Palbansh

Biomedical Engineering Department
Thadomal Shahani college of engineering,
Mumbai University
paparnapal@gmail.com

Dhananjay K. Theckedath

Guide, Assistant Professor
Biomedical Engineering Department
Thadomal Shahani college of engineering,
Mumbai University

ABSTRACT

In this paper we use a Block Selection technique of the Wavelet Transform for face recognition. The standard technique used for face recognition uses the Principle Component Analysis (PCA). The PCA algorithm functions by projecting the face image onto a feature space that gives significant variations among the known face images. The significant features are called eigenfaces since they are the eigenvectors of the set of faces. The projection operation characterizes an individual face by a weighted sum of the eigenface features and face is recognized by comparing these weights. In the wavelet transform approach discussed in the paper, we compute the wavelet transform of the input face image. The LL image of the wavelet decomposition is then partitioned into non-overlapping rectangular blocks. Similarity measurements are used to compare these blocks, which in turn gives us a robust face recognition algorithm.

Key Words : PCA, Eigen vector, Eigenfaces, Wavelet Transform

I. INTRODUCTION

Face recognition has been attracting intense research efforts due to its significance both as one of the main building blocks of natural human computer interfaces and as a biometric trait [1]. Moreover, it is expected that the solution of the face recognition problem would provide valuable insights about the other pattern recognition problems. The face is our primary focus of attention in social life, playing an inevitable role in conveying identity and emotion. Although the ability to infer intelligence or character from facial appearance is suspect, the human ability to recognize face is remarkable. We can recognize thousands of faces at a glance even after years of separation. This skill is quite robust, in spite of deviations in expressions, aging and distractions such as glasses or

changes in hairstyle or facial hair. As a consequence the visual processing of human faces has fascinated philosophers and scientists for centuries.

II. BIOMETRICS

Biometrics is one of the most critical emerging technologies of the 21st century. As the world becomes more volatile and potentially more vulnerable to fraudulent forces, our fate lies within our ability to secure and protect critical data and facilities [2]. A concise definition of biometrics is “the automatic recognition of a person using distinguishing traits.” A more expansive definition of biometrics is “any automatically measurable, robust and distinctive physical characteristic or personal trait that can be used to identify an individual or verify the claimed identity of an individual.” Measurable means that the characteristic or trait can be easily presented to a sensor, located by it, and converted into a quantifiable, digital format. This measurability allows for matching to occur in a matter of seconds and makes it an automated process. The robustness of a biometric refers to the extent to which the characteristic or trait is subject to significant changes over time. These changes can occur as a result of age, injury, illness, occupational use, or chemical exposure. A highly robust biometric does not change significantly over time while a less robust biometric will change. For example, the iris, which changes very little over a person’s lifetime, is more robust than one’s voice.

A. Face as biometric

Face recognition has a number of strengths to recommend it over other biometric modalities in certain circumstances, and corresponding weaknesses that make it an inappropriate choice of biometric for other applications. Face recognition as a biometric derives a number of advantages from being the primary biometric that humans use to recognize one

another [3]. Furthermore it is well-accepted and easily understood by people, and it is easy for a human operator to arbitrate machine decisions — in fact face images are often used as a human-verifiable backup to automated fingerprint recognition systems.

Face recognition has the advantage of ubiquity and of being universal over other major biometrics, in that everyone has a face and everyone readily displays the face. (Whereas, for instance, fingerprints are captured with much more difficulty and a significant proportion of the population has fingerprints that cannot be captured with quality sufficient for recognition.) Surveillance systems rely on passive acquisition by capturing the face image without the cooperation or knowledge of the person being imaged. The acquisition devices are cheap. The main drawbacks to face recognition are its current relatively low accuracy (compared to the proven performance of fingerprint and iris recognition) and the relative ease with which many systems can be defeated. Finally, there are many attributes leading to the variability of images of a single face that add to the complexity of the recognition problem if they cannot be avoided by careful design of the capture situation. Inadequate constraint or handling of such variability inevitably leads to failures in recognition. These include:

1. **Physical changes:** facial expression change; aging; personal appearance (make- up, glasses, facial hair, hairstyle, disguise).
2. **Acquisition geometry changes:** change in scale, location and in-plane rotation of the face (facing the camera) as well as rotation in depth (facing the camera obliquely, or presentation of a profile, not full-frontal face).
3. **Imaging changes:** lighting variation; camera variations; channel characteristics (especially in broadcast, or compressed images).

III. PCA

A. Eigen face-based

It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. The other main advantage of PCA is that once you have found these patterns in the data, you can compress the data, i.e., by reducing the number of dimensions, without much loss of information [4].

B. Working

The task of facial recognition is discriminating input signals (image data) into several classes (persons). The input signals are highly noisy (e.g. the noise is

caused by differing lighting conditions, pose etc.), yet the input images are not completely random and in spite of their differences there are patterns which occur in any input signal. Such patterns, which can be observed in all signals, could be – in the domain of facial recognition – the presence of some objects (eyes, nose, mouth) in any face as well as relative distances between these objects. These characteristic features are called eigenfaces in the facial recognition domain (or principal components generally). They can be extracted out of original image data by means of a mathematical tool called Principal Component Analysis (PCA)[5]. The objective is to extract the relevant information in a face image, encode it as efficiently as possible, and compare one face encoding with a database of models encoded in the same way. In mathematical terms, the objective is to find the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images. These eigenvectors can be thought of as a set of features which together characterize the variation between face images. By means of PCA one can transform each original image of the training set into a corresponding eigenface. An important feature of PCA is that one can reconstruct any original image from the training set by combining the eigenfaces. Eigenfaces are nothing less than characteristic features of the faces. Each face image in the training set can be represented exactly in terms of a linear combination of the eigenfaces. The number of possible eigenfaces is equal to the number of face images in the training set. The eigenfaces are essentially the basis vectors of the eigenface decomposition. Therefore one could say that the original face image can be reconstructed from eigenfaces if one adds up all the eigenfaces (features) in the right proportion. Each eigenface represents only certain features of the face, which may or may not be present in the original image.

If the feature is present in the original image to a higher degree, the share of the corresponding eigenface in the “sum” of the eigenfaces should be greater. In contrast, the particular feature is not (or almost not) present in the original image, then the corresponding eigenface should contribute a smaller (or not at all) part to the sum of eigenfaces. So, in order to reconstruct the original image from the eigenfaces, one has to build a kind of weighted sum of all eigenfaces. That is, the reconstructed original image is equal to a sum of all eigenfaces, with each eigenface having a certain weight. This weight specifies, to what degree the specific feature (eigenface) is present in the original image.

If one uses all the eigenfaces extracted from original images, one can reconstruct the original images from the eigenfaces exactly. But one can also use only a part of the eigenfaces. Then the reconstructed image is an approximation of the original image. However, one can ensure that losses due to omitting some of the eigenfaces can be minimized. This happens by

choosing only the most important features (eigenfaces). using only the “best” eigenfaces—those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images. The primary reason for using fewer eigenfaces is computational efficiency. The most meaningful M eigenfaces span an M -dimensional subspace—“face space”—of all possible images[6].

IV. WAVELET ANALYSIS

Wavelets are functions that satisfy certain mathematical requirements and are used in presenting data or other functions, similar to sines and cosines in the Fourier transform [7,8]. However, Wavelet Transform represents data at different scales or resolutions, which distinguishes it from the Fourier transform.

A. Discrete wavelet transform

Discrete wavelet transform (DWT) is a well-known signal analysis tool, widely used in feature extraction, compression and denoising applications. The two-dimensional wavelet transform is performed by consecutively applying one-dimensional wavelet transform to the rows and columns of the two-dimensional data [9]. The original image of resolution $N \times N$ is first filtered along the rows and down sampled by 2 yielding two $N \times N/2$ images that have high and low frequency contents, respectively. After this decomposition, the wavelet transform is applied to the columns of these $N \times N/2$ resolution images. In the final stage of the decomposition we have four $N/2 \times N/2$ resolution sub band images: the scaling component containing low-pass global information obtained by low-pass filtering the rows and columns; the horizontal details obtained by low-pass filtering the rows and high-pass filtering the columns; the vertical details obtained by high-pass filtering the rows and low-pass filtering the columns; the diagonal details obtained by high-pass filtering the rows and columns. The decomposition can be continued further by performing same processing steps on the scaling component [10].

In Figure, one level wavelet decomposition of a face image is shown.

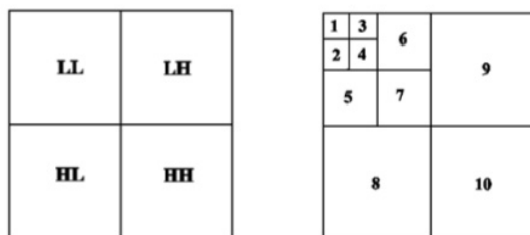


Figure 1: One level decomposition.

The band LL is a coarser approximation to the original image. The bands LH and HL record the

changes of the image along horizontal and vertical directions. The HH band shows the high frequency component of the image. This is the first level decomposition. Further decomposition can be conducted on the LL sub-band. After applying a 3-level wavelet transform, an image is decomposed into sub-bands of different frequency components as shown in Fig b. If the resolution of an image is 128×128 , the sub-bands 1, 2, 3 and 4 are of size 16×16 , the sub-bands 5, 6, 7 are of size 32×32 and the sub-bands 8, 9, 10 are of size 64×64 .

B. Overview of the algorithm

A two dimensional wavelet transform is derived from two one-dimensional wavelet transform by taking **tensor products**. The implementation of WT is carried out by applying a one-dimensional transform to the rows of the original image data and the columns of the row- transformed data respectively [11]. Three level wavelet decomposition is performed on an image as shown in the figure.

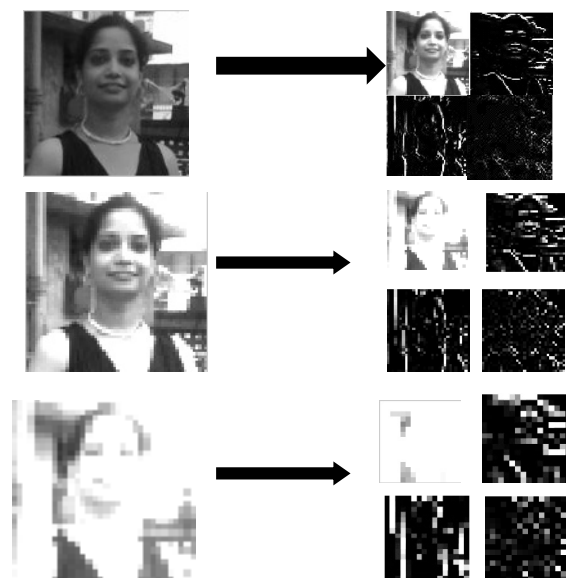


Figure2: 3 level wavelet decomposition.

Image of size 128×128 is read and three-level discrete wavelet decomposition is performed on the entire face image. The resulting third level scaling component is used for further processing. This 16×16 pixels resolution scaling component is divided into 4×4 pixels blocks, providing 16 non-overlapping blocks. From these blocks the important ones are selected using the scheme described in block selection. Schematic diagram is shown below.

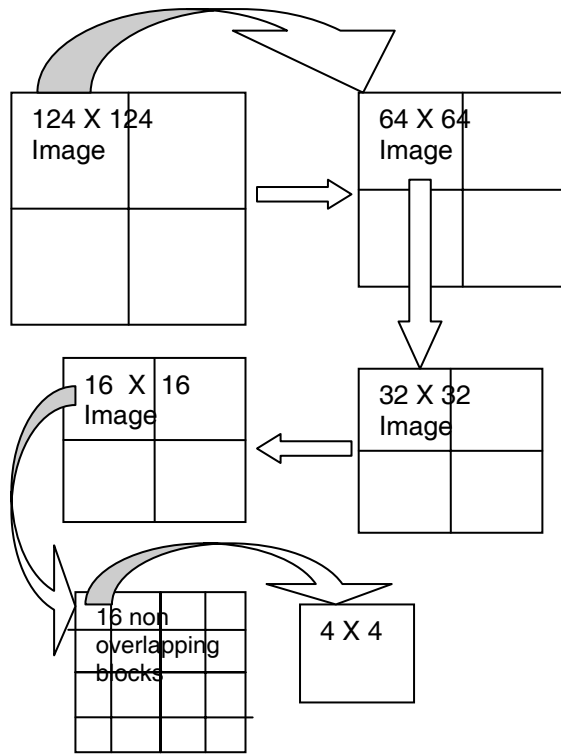


Figure 3 : Schematic diagram of three level wavelet decomposition.

C. Block selection

By measuring the similarity between the pixel values of each block in the average training image and the pixel values of the corresponding blocks in the test image block selection is done. This facilitates adaptive block selection. In other words, for each new test face image, different blocks can be selected [12]. According to reference paper [12] features are extracted from each selected block are concatenated to construct over all feature vector. But instead of forming a feature vector, we considered first block as shown in figure 3.

The average pixel values of the i th block of the training images can be calculated as

$$\bar{b}_{tra,i} = \frac{1}{L} \sum_{l=1}^L b_{tra,i,l} \quad (1)$$

where L is the number of images in the training set and $b_{tra,i}$ represents the vector that contains the average pixel values of the i th block. This K dimensional vector's mean value can be calculated as

$$m_{tra,i} = \frac{1}{K} \sum_{k=1}^K b_{tra,i,k} \quad (2)$$

By subtracting the vector's mean value from its pixel values, the $b_{tra,i}$ is obtained.

$$\bar{b}_{tra,i} = b_{tra,i} - m_{tra,i} \quad (3)$$

Similarly, the mean value of the i th block in the test image can be calculated as

$$m_{test,i} = \frac{1}{K} \sum_{k=1}^K b_{test,i,k} \quad (4)$$

Again, by subtracting the vector's mean value from its pixel values, the $b_{test,i}$ is obtained.

$$\bar{b}_{test,i} = b_{test,i} - m_{test,i} \quad (5)$$

Finally, the similarity between the i th block in the average training image and the corresponding block in the test image is calculated as [12]

$$s_i = \bar{b}_{tra,i} * \bar{b}_{test,i} / \|\bar{b}_{tra,i}\| * \|\bar{b}_{test,i}\| \quad (6)$$

RESULT

Database of ten different people with seventeen different expressions was used. The value of the S_i lies in the range of 0.9 to 1 for the blocks of the same person. Blocks having similarity can be used to form a feature vector. This feature vector can be used to compare the faces. Or else the blocks showing maximum S_i can be used to recognize faces from the database. To check the robustness of the system, the images were modified and checked for the similarity.

CONCLUSIONS

Early attempts at making computers recognize faces were limited by the use of impoverished face models and feature descriptions i.e. locating features from an edge image and matching simple distances and ratios. The eigenface approach to face recognition was motivated by information theory leading to an idea of basing face recognition on a small set of image features that best approximates the set of known face images without requiring information about facial parts and features. Wavelet transform which decomposes an image into different frequency bands is used widely in face recognition systems. The block selection technique used in the paper is simple and fast and robust. The block selection technique also appears to work well in a noisy environment. This technique can be used to check malignancy.

REFERENCES

- [1] Kaya Y & Kobayashi (1972). A basic study on human face recognition . In S. watanbe (Ed.), Frontiers of pattern recognition. newyork.
- [2] Bledsoe, W.W. (1966a). Manmachine facial recognition . Panoramic Research Inc. Palo Alto, CA. Rep. PRI:22, August.
- [3] Burt, P. (1988a). Algorithms and architectures for smart sensing. Processing of the image understanding workshop, April
- [4] Fischler, M. A, & Elschlager, R A (1973). The representation and matching of pictorial structures. IEEE Transactionson computers, c-22(1)
- [5] Flemming, M & Cottrell, G. (1990) Categorization of faces using unsupervised feature extraction. Proceedings of IJCNN 90, 2.
- [6] Matthew Turk & Alex Pentland. "Eigenfaces for Recognition". Journal of cognitive science pp-71-86, 1991.
- [7] Digital image processing by Gonzalez
- [8] Fundamentals of digital image processing by A.K. Jain
- [9] Digital Image processing (principles of image processing) by Dhananjay Theckedath
- [10] Kanade , T. (1973). Picture processing system by computer complex and recognition of human faces, Kyoto university.
- [11] The wavelet tutorial-Robi Polykar. www.site.uottawa.ca/~gingchen/wavelet/p2.htm.
- [12] Hazim Kemal Ekenel , Rainer Stiefelhagen. Local wavelet analysis for Face Recognition
- [13] Carey, S, & Diamond, R (1977). From piecemeal to configurational representation of faces. Science. 195, 312-313
- [14] Goldstein, Harmon, & Lesk (1971). Identification of human faces, proceedings IEEE, 59, 748.
- [15] Kohonen, T. (1981) Storage and processing of information in distributed associative memory.

BANDWIDTH ENHANCEMENT OF LOW PROFILE AIR FED MICROSTRIP ANTENNAS

Prathamesh Indulkar
ME EXTC
VIT
Wadala, Mumbai 400037
indulkar.pratham@gmail.com

Mithila Ghuge
ME EXTC
MIT
Nasik, Maharashtra
mithila.ghuge@gmail.com

Vaibhav Kshirsagar
ME EXTC
VIT
Wadala, Mumbai 400037
vaibsan@gmail.com

ABSTRACT

A compact, low profile air fed microstrip antenna is proposed and experimentally studied for its bandwidth improvement through simulation. A rectangular microstrip patch is modified on its edges in the form of small notches as perturbation elements with air as dielectric material. With proposed structure two closely spaced degenerating resonant frequencies are generated such that bandwidth over VSWR= 2 is enhanced.

The simulation result reported that, bandwidth with VSWR lower than 2 is 320MHz carried with Genesys, Agilent tool. Throughout this paper, simulation results of the impedance bandwidth, and parametric study of the antenna is followed. This antenna can be used in wireless communication system for its advantage of broad- bandwidth.

Index terms- Bandwidth enhancement, perturbation elements, degenerate frequencies

CONTENTS

I	Introduction	1
II	Antenna Geometry & Design	1
II-A	Element Width	1
II-B	Element Length	1
III	Results And Discussion	2
III-A	Effect of variation of feedpatch height	2
III-B	Effect of feed probe position	3
III-C	Effect of variation in feedpatch length	3
III-D	Effect of variation in feedpatch width	3
III-E	Effect of modification in patch	3
III-F	Effect of different notch size	4
III-G	Simulated results of modified antenna	4
IV	Future critical needs and further developments	5
V	Conclusion	5
	References	5

I. INTRODUCTION

Microstrip antenna technology has been the most rapidly developing topic in the antenna field in the last fifteen years, receiving the creative attentions of academic, industrial, and government engineers and researchers throughout the world. During this period there have

been over 1500 published journal articles, many books and innumerable symposia sessions and short courses devoted to the subject of microstrip antenna and arrays. As a result microstrip antennas have quickly evolved from academic novelty to commercial reality, with applications in a wide variety of microwave systems. In recent years, microstrip ultra wideband antennas have attracted more attention owing to their advantages such as simple structure, low profile, high data rate, easy integration with monolithic microwave integrated circuits (MMICs), and ease of fabrication. When compared to traditional antenna elements such as re-directors, horns, slots, or wire antennas, however, the electrical performance of the basic microstrip antenna or array suffers from a number of serious drawbacks; including very narrow bandwidth high feed network losses, poor cross polarization, and low power handling capacity. The purposes of this paper is to present a new approach of using Perturbation element.

II. ANTENNA GEOMETRY & DESIGN

A. Element Width

For an efficient radiator, width is given by

$$W = \frac{c}{2f_0 \sqrt{\frac{\epsilon_r + 1}{2}}} \quad (1)$$

B. Element Length

To choose the resonant length would also mean choosing the frequency of resonance since the resonant frequency of the patch is determined by the patch length. The length of the patch should be slightly less than half the dielectric wavelength since the actual patch is 'longer' due to the fringing fields. The length of the patch is given as,

$$L_{eff} = \frac{c}{2f_0 \sqrt{\epsilon_{eff}}} \quad (2)$$

Where f_0 represent the resonant frequency

$$\Delta L = 0.412h \frac{(\epsilon_{eff} + 0.3) \left(\frac{W}{h} + 0.264 \right)}{(\epsilon_{eff} - 0.259) \left(\frac{W}{h} + 0.3 \right)} \quad (3)$$

ΔL represent the line extension at the ends given by Hammerstad.

The effective dielectric constant ϵ_{eff} may be static or frequency dependent value where,

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + 12 \frac{h}{w} \right)^{-\frac{1}{2}} \quad (4)$$

Therefore length of the feed patch can be calculated as

$$L = L_{eff} - \Delta L \quad (5)$$

A low profile microstrip patch antenna is proposed as shown in Fig 1

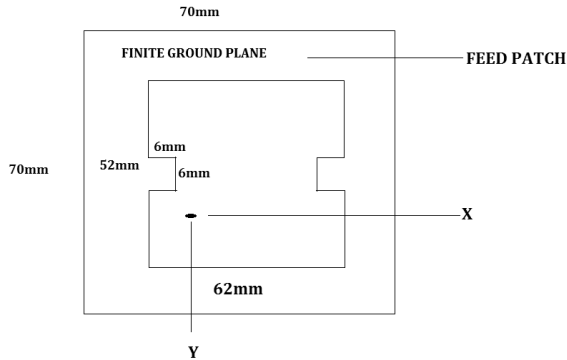
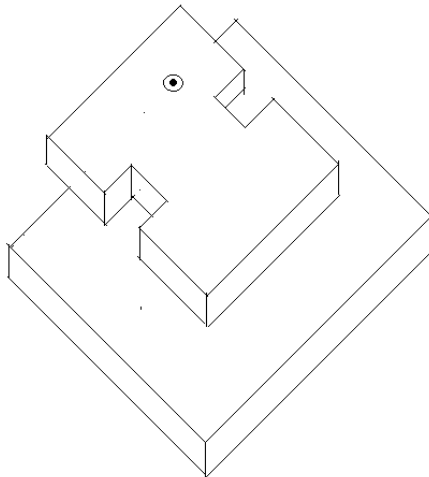


Fig. 1 Antenna Geometry



3D View of antenna

Rectangular geometry was simulated to optimize the performance, starting with rectangular patch antenna on the ground plane and ending with using perturbation structure. The conventional shape of rectangular microstrip patch antenna shown in Fig. 1 has the patch dimension of 62X52 mm and rectangular ground plane with dimension equal to 70X70 mm.

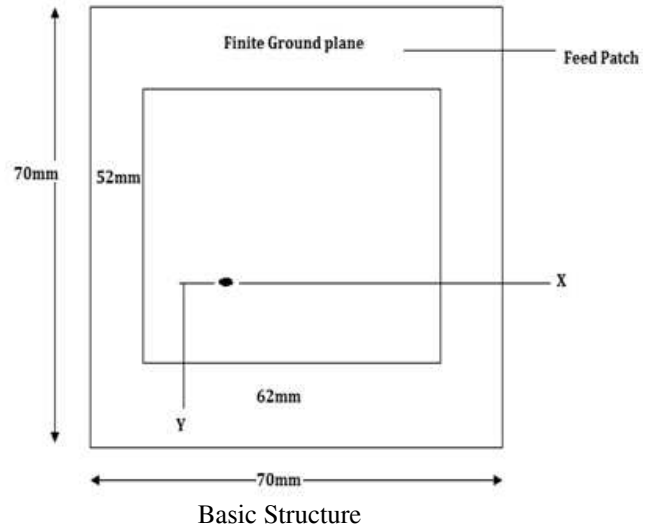
The patch is notch from both side on width equal to 6X6 mm from the center. The antenna is designed to operate band frequency range of 2.1GHz to 2.4GHz. The antenna fundamental frequency was found to be 2.4GHz. The feed patch is 0.5mm thick copper plate suspended in air with probe feed mechanism, and ground plane is also 0.5mm thick copper plate. We select design layout in GENESYS software and keep

grid spacing as 0.5 mm, box width (X) as 100-400 cells and box height (Y) as 100-400 cells. The coaxial feed probe is placed at x=20.5 mm and y=17 mm from left bottom corner. The antenna is printed on FR4 substrate with $\epsilon_r=1$

III. RESULTS AND DISCUSSION

The antenna performance was investigated numerically and experimentally. The simulation was done using commercial simulator GENESYS. Starting with the conventional rectangular shape for both radiating patch and ground plane. The dimension of the radiating patch are 52X62mm and for the ground plane dimensions are 70x70mm as shown in fig 2.

The coaxial feed probe is placed at x=20.5mm and y=17mm from left bottom corner.



A. Effect of variation of feedpatch height

The coaxial probe feed passing through air dielectric exhibits inductive shifts in impedance curve and capacitive curve. The inductive and capacitive reactance curve never passes through zero. The variations in the feed patch height changes the electrical thickness of the air dielectric below it. The decrease in feed patch height to 3 mm shifts the impedance curve and resonated frequency is shifted. The resonate frequency can be restored back by changing the length of feed patch. The cross polarization is reduced and efficiency is improved. The increased in the fed patch height to 6 mm shift the impedance curve clockwise on the smith chart with increase in impedance and resonant condition is altered. It can be again restored by changing the feed patch height. The fed patch with the height 5mm gives the best result because it gives the impedance matching. Fig. 4 shows the graph of variation in feed patch height.

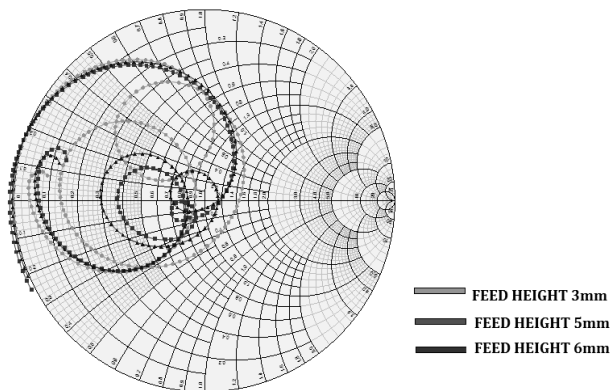


Fig. 4 Impedance Variation

B. Effect of feed probe position

The coaxial probe feed it effect the antenna characteristics it can be seen that by changing the position of the feed probe we can achieve the impedance matching for the antenna. Variation of the probe location along the X and Y axis varies the lobes of the on the smith chart which is shown below in Fig 5.

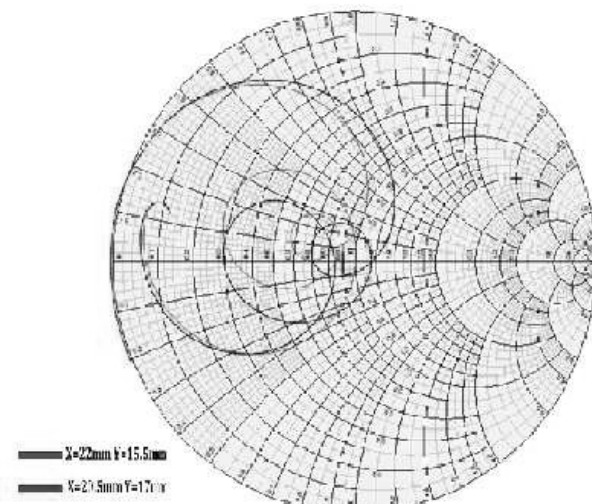


Fig. 5. Impedance variation

Here we can observe that the feed position of probe at X as 22mm and Y as 15.5mm is not gives perfect impedance matching. The position of feed position of probe at X as 20.5mm and Y as 17mm we get perfect impedance matching as it passes through the origin of the smith chart i.e matching point.

C. Effect of variation in feed patch length

For a patch antenna, the length is determining factor in deciding resonant condition in fundamental operating mode. The rectangular patch resonates in its fundamental TM₁₀ mode, where there is no variation along the width and field is constant while there is half wavelength variation along the length. The sinusoidal variation cancels in broadside direction and the radiation is only accomplished by the field along the width. The resonant frequency is detuned when the length of the

feed patch is varied. When the feed patch length is 60mm, the bandwidth we get is 18MHz is observed.

When we increase the length of the feed patch to 62mm we got the bandwidth as 114MHz for return loss - 9.54dB. These differences of bandwidth is shown in Fig 6

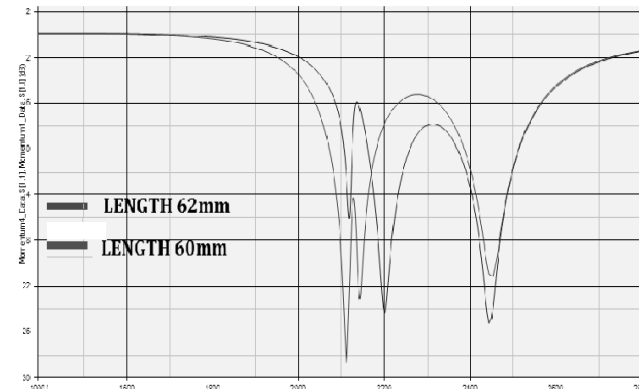


Fig. 6 Return Loss variation Vrs Different Length

D. Effect of variation in feedpatch width

For rectangular fed patch to resonate at 2.4 GHz, air dielectric. The theoretical calculated width is 62mm. The width of the 50 mm we got the return loss less than the width of 52mm. we got the approximately same bandwidth. It does not affect the resonant frequency of the antenna. The fig 7 shows the variation of width. By trial and error method we come to conclusion that, though the theoretical length and practical width both are different

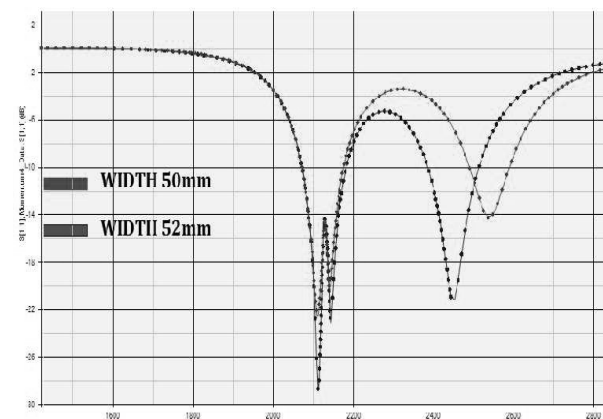


Fig. 7 Return Loss variation Vrs Different Width

E. Effect of modification in patch

Patch is notch on width from both side by 6X6mm from the center. The modified patch produced following results which is very improved compared to reference antenna. By changing the position of the feed probe location we can see the variations in the impedance on smith chart in fig. 8. It also varies the return loss of the antenna which affect the bandwidth of the antenna as shown in fig 8.

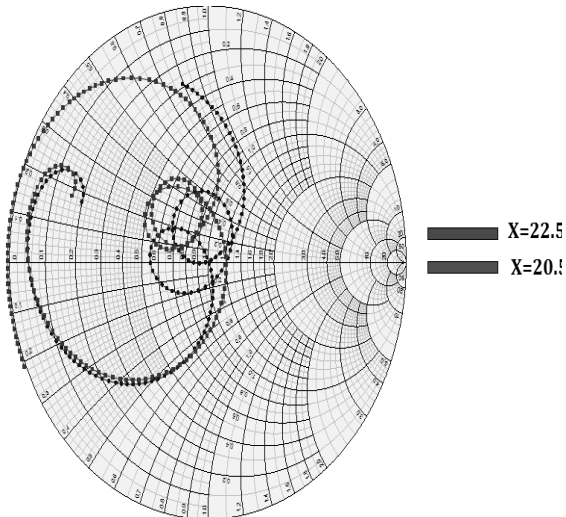


Fig. 8 Impedance Variation

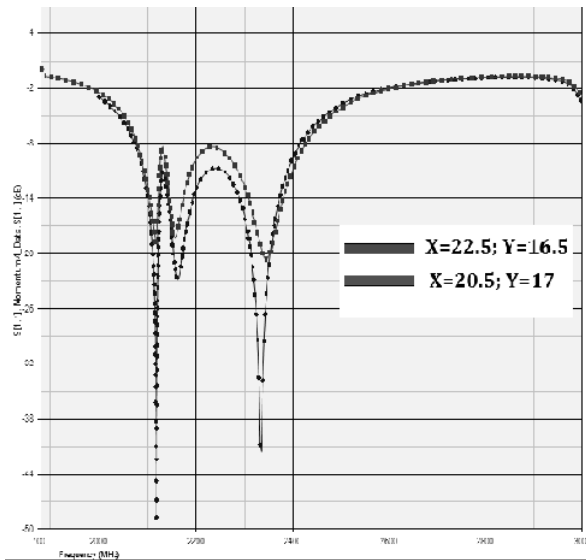


Fig. 9 Return Loss variation Vrs Different Feed

F. Effect of different notch size

The notch or shape which cuts in the patch affects the antenna parameter. Here we cut the square at both the radiating edge of the feed patch. The fig 10 and fig 11 shows the variation in return loss as well as variation in the impedance of the antenna respectively. Here we consider two type of square shape of 6mm x 6mm and 12mm x 12mm. It is find that bandwidth is more in the structure having the notch of the 6mm x 6mm rather than the 12mm x 12mm. And it also shows that the impedance matching is perfect for the notch of 6mm x 6mm rather than 12mm x 12mm. It is find that bandwidth is more in the structure having the notch of the 6mm x 6mm rather than the 12mm x 12mm. And it also shows that the impedance matching is perfect for the notch of 6mm x 6mm rather than 12mm x 12mm.

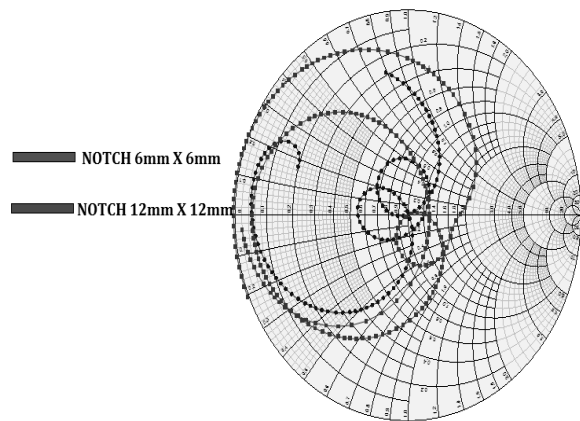


Fig. 10 Impedance Variation

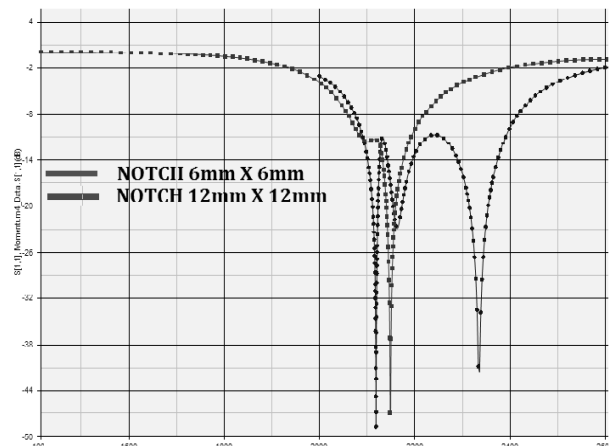


Fig. 11 Return Loss variation Vrs Different Notch Size

G. Simulated results of modified antenna

By comparing all the results, which are obtained by parametric study, the antenna is simulated with following specification. Ground plane is of 70X70 mm. For patch W=52mm, L=62mm. The probe radius is of 2mm and it is located at x=20.5 and y= 17. Patch is notched on width with 6X6mm dimension. The fig 12 shows the variation of impedance. Fig12 shows the return loss and fig 13 shows the VSWR curve.

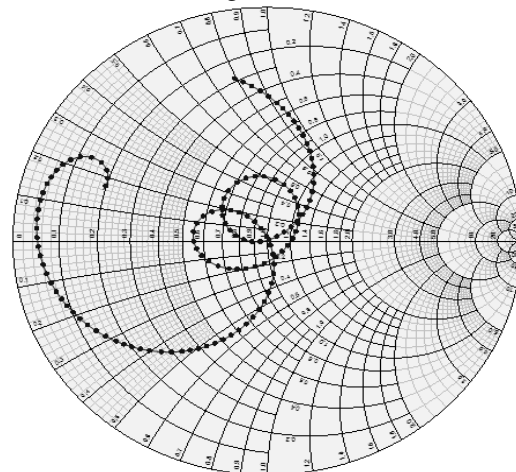


Fig. 12 Impedance variation

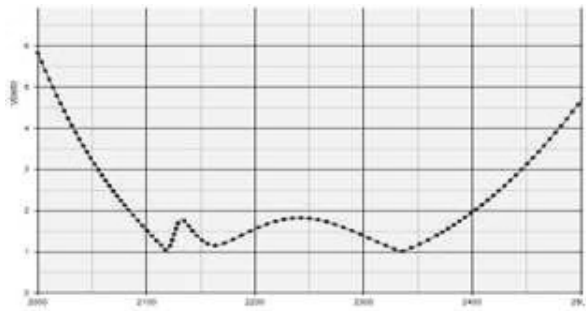


Fig. 13 VSWR

IV. FUTURE CRITICAL NEEDS AND FURTHER DEVELOPMENTS

The planar patch antennas have many advantages over the traditional high directive structures in the view of high efficiency, cost effective, low profile and conformability to the planar and non planar structures. These planar antennas are found to be having applications in military aircraft, missiles, rockets and satellites. Dual frequency operation is possible using either dual stacked patches or patch with loaded diode or a stub. However patch antennas suffer from the drawbacks of narrow bandwidth, low RF power handling capability, and larger ohmic insertion loss than other types of antennas of equivalent aperture size. The following further improvements and future development works as suggested in [6] are,

Design, analysis, finite ground array structures

- Design of dual band and circularly polarized arrays
- Analysis of exact calculation of mutual coupling between the array elements
- Analysis of phase variation and geometrical optimization of patch dimension in array configurations
- Analysis of excitation amplitude, diffraction effects, positioning errors and errors due to imperfect element matching and feed network isolation to improve the side lobe levels
- Modelling techniques of patch elements
- Design of multifunction (switched or multiple beams at single frequency), multiple frequencies or polarizations within single aperture.

CONCLUSION

In this paper a low profile, air dielectric, notched microstrip antenna is investigated for its bandwidth improvement. instead of notch any other modification in the structure can be used as a perturbation element to achieve two closely resonant frequencies which will enhance the impedance bandwidth.

REFERENCES

- [1] G. A.Deschamps, "Microstrip microwave antennas", presented at the 3rd USAF Symp. On antennas, 1953
- [2] H. Gutton and G. Baissinot, "Flat aerial for ultra high frequencies", French patent No. 703113, 1955
- [3] R. E. Munson, "Single slot cavity antennas assembly", U.S. patent No.3713 162, January 23, 1973
- [4] Reference data for radio engineers, 5th ed. Indianapolis, IN: Howard W.Sams, Oct. 1968, Ch-22, pp 25-27
- [5] K.P. Ray and G. Kumar, "Broadband Microstrip Antennas", Archtech House, ISBN: 1-58053-244-6, 2003.
- [6] V. C. Kshirsagar, "Design and development of stacked planar high gain antenna for wireless communication", Dissertation report, M. E., Mumbai University, July 2008.

Prathamesh S. Mestry

Vidyalankar Institute of Technology

ME -EXTC

Wadala, Mumbai: 400037

prathamesh.s.mestry@vit.edu.in

ABSTRACT

The project undertaken was to study antennas using fractal patterns in order to obtain desired performance properties such as compact size and multi-band behavior one of the prevailing trends in modern wireless mobile devices is a continuing decrease in physical size[4]. In addition, as integration of multiple wireless technologies becomes possible, the wireless device will operate at multiple frequency bands. A reduction in physical size and multi-band capability are thus important design requirements for antennas in future wireless devices. The use of fractal patterns in antenna design provides a simple and ancient method for obtaining the desired compactness and multiband properties. A compact, multi-band antenna based on the Sierpinski fractal was analyzed to operate at 2.4 and 5.0 GHz. The design process consisted of initial theoretical calculations followed by extensive numerical simulations. The design methodology used, simulation and test results are presented.

Keywords– Fractal, Sierpinski, Monopole, Microstrip.

I. INTRODUCTION

To design antennas which have properties beneficial to modern wireless communication receivers, Although fractal geometry has been known to mathematics for a century, fractal antenna engineering research is a relatively very recent development because considerable computing speed is required to complete their design. It has been discovered that fractal shapes radiate electromagnetic energy well and have several properties that are advantageous over traditional antenna types. They can be used as multi-band antennas, which can radiate signals at multiple frequency bands [4]. Another desirable property is that they are compact, meaning that they can occupy a portion of space more efficiently than other antenna types. Such antennas could be used to improve the functionality of modern wireless communication receivers such as cellular handsets. As already mentioned in the previous section, fractal antenna technology could be applied to cellular handsets.

Because fractal antennas are more compact, they would more easily fit in the receiver package. Currently, many cellular handsets use quarter wavelength monopoles

which are essentially sections of radiating wires cut to a determined length. Although simple, they have excellent radiation properties. It would be highly beneficial to design an antenna with similar radiation properties as the quarter-wavelength monopole while retaining its radiation properties. Other prevailing trends in wireless communications technology could also benefit.

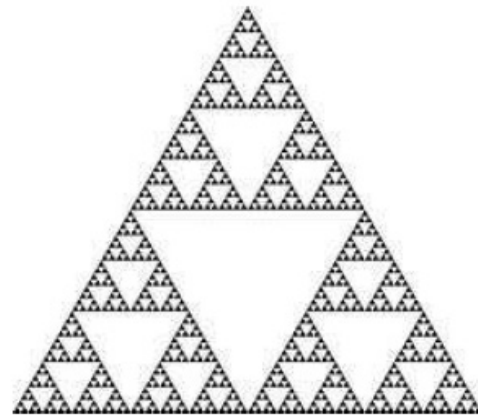


Fig. 1. Fractal Antenna using ISF method

More and more systems are introduced which integrate many technologies. They are often required to operate at multiple frequency bands and so they require antenna systems which accommodate that requirement. Using fractal patterns in antenna design can help solve the problems mentioned above.

II. LITERATURE REVIEW

A. Basic review

This section presents a literature review of the theory of fractal antennas [1]-[5]. An introduction to fractal theory is first discussed by giving example of Fractal Sierpinski Gasket. Micro strip patch antenna technology is then presented to outline its application to fractal antennas.

B. Fractal Theory

Fractals are classes of shapes which have no characteristic size. Each fractal is composed of multiple iterations of a single elementary shape. The iterations can continue infinitely, thus forming a shape within a finite boundary but of infinite length or area. This compactness property is highly desirable in mobile

wireless communication applications because smaller receivers could be produced.

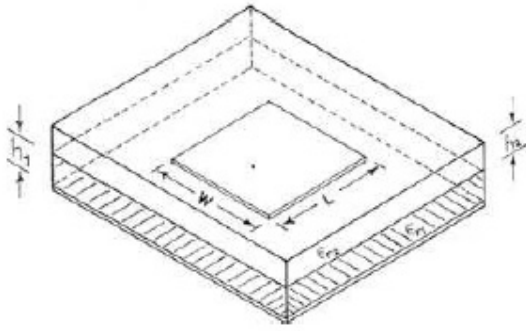


Fig. 2. Microstrip Patch Antenna Geometry

1) *Iterated Function :*

Many useful fractals can be generated by, Iterated Function Systems (IFS) [1]. IFS work by applying a series of affine transformations ω to an elementary shape over much iteration. The affine transformation $\omega(x,y)$, comprising rotation, scaling and translation in equation 1.

$$\omega(x,y) = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

The set of affine transforms $\omega(A)$, known as the Hutchinson operator is given by 2.

$$\omega(A) = \bigcup_{n=1}^N \omega_n(A) \quad (2)$$

The fractal can then be generated by applying the operator ω to the previous geometry for k th iterations [1],[5].

2) *The Sierpinski Gasket Fractal:*

The Sierpinski Gasket fractal is generated by the IFS method. A triangular elementary shape is iteratively scaled, rotated and translated, then removed from the original shape in order to generate a fractal. It is interesting to note that after infinite iterations of the fractal, the entire shape has an infinite area but is bounded by a finite perimeter [3].

C. Microstrip Patch Antennas

The microstrip patch antenna has become one of the most versatile radiating element solutions for a large variety of system. Microstrip patches are essentially conductive radiating elements etched on a thin dielectric layer as shown in Fig. 2. A ground plane is etched on the opposing side of the dielectric layer from the radiating element. Microstrip patches are resonant-type antennas, meaning that in order for the antenna to radiate, one of the dimensions of the radiating patch must be approximately half the wavelength of the electrical excitation signal being fed to the patch. An advantage of microstrip and printed antennas is that they can be easily

and inexpensively built using commonly available printed circuit board (PCB) technology. Microstrip patches can thus also be produced at small size and profile, allowing easy integration into the skins of various systems such as airplanes or cellular handsets. Because they are resonant type antennas, microstrip patches are efficient radiators. The efficiency of microstrip patches is generally between 95% and 99%.

Microstrip patches also have some disadvantages, most important of which is their limited impedance bandwidth. Because microstrip patches are resonators, its impedance will inherently be largely real over only a small percentage around the operating frequency for which it was designed. This characteristic prevents the use of microstrip patches to a number of very wide-band applications. However impedance bandwidths of up to 10% can commonly be achieved using microstrip patches, a value sufficient for many applications.

D. Fractal Microstrip Antennas

1) *The Sierpinski Gasket Monopole:*

The Sierpinski gasket monopole antenna was first presented by Puente et al. The Sierpinski gasket was chosen by researchers due to its resemblance to the well-known, triangular (bow-tie) monopole antenna. In addition, the triangular structure presents a simple way of feeding current to the antenna. Current is fed through a connector at the bottom tip of the monopole [3].

The current fed to an antenna will generally concentrate over a region in size comparable to the current's wavelength. The introduction of boundaries at such locations in the Sierpinski gasket allows radiation to occur. Thus, each resonant frequency is directly related to the dimensions of each sub-gasket. The resonant frequency band spacing is related to the similarity factor δ of the gasket. The gasket has $\delta = 2$, meaning a ratio of 2 would be seen between adjacent bands in the corresponding monopole. A general relationship between the resonant frequency of the monopole and gasket parameters is given in 3.

$$f_n = \frac{kc}{h} \cos\left(\frac{\alpha}{2}\right) \delta^n \quad (3)$$

Where c is the speed of light, h is the height of the monopole, α is the angle and n is the resonant band number. Thus, the resonant frequency band locations can be adjusted by varying α and δ .

III. PROBLEM DEFINITION

A. Basic Problem Definition

The purpose of this project is to design fractal antennas which have desired multiband resonance and compact profile.

B. Solution Overview

Since using fractals as an approach to antenna design is a relatively new development. The design plan includes the following phases:

- **Theoretical development:** Rough calculations of antenna parameters as well as develop a general idea of the physical implementation of the antenna.
- **Numerical simulation:** Perform software simulations in order to verify the theoretical design and adjust any parameters to predict the desired antenna performance.

1) *Theoretical Development:* Considerations for the following performance characteristics should be taken into account before proceeding with a design:

- **Resonant Frequency:** Narrow band at design frequency, as is inherent of resonant antennas. The Sierpinski Gasket monopole would exhibit a number of resonant bands.
- **Impedance Bandwidth:** Resonant-type antennas generally have low impedance bandwidth. The input impedance remains approximately constant in a narrow band of frequencies, generally between 5 to 10 percent of the center frequency.
- **Input Impedance:** Should be resistive and large in value, although the first resonance usually has low impedance and a noticeable reactive component.
- **Radiation Pattern:** As both antennas are most suitable for use in low power applications such as mobile receivers, their radiation patterns should be suitable for such use. This should be achievable since both the Sierpinski is monopoles, inherently having a very uniform radiation pattern in all directions
- **Radiation Efficiency:** Microstrip antennas have high radiation efficiency, between 90 to 99.9 percent. The thickness and relative dielectric constant ϵ_r of the dielectric substrate used affect radiation efficiency. For efficiency, substrates with low dielectric constants should be used.
- **Space Efficiency (Compactness):** Due to the space filling properties of Sierpinski fractals, the antenna designs should exhibit efficient space usage. Although the aim of the Sierpinski monopole design is multi-band operation, the antenna is expected to be compact, since the fractal structure of the antenna represents smaller antennas bounded by larger ones.
- **Gain:** The gain of the fractal antennas should be comparable to that of traditional antennas with similar radiation properties. This would allow the fractal antennas to be used in similar applications.

2) *Fractal Antenna Patterns:* After establishing the required resonant frequencies, initial calculations of the geometrical properties of the fractal radiators can be undertaken. The fractal pattern features must be comparable in size to the wavelength of the input signal. As a start, the height of each triangular structure in the Sierpinski gasket pattern should be approximately $\lambda/2$. The resonant frequency spacing can then be approximated by varying the parameters in equation 3.

C. Software Simulators

There are a number of commonly available software packages which allow the simulation of antenna parameters. Some of the best known are Genesis, ADS 2009, IE3D, HFSS are excellent professional design tools which offer a great deal of simulation analysis options.

IV. DESIGN THEORY OF THE SIERPINSKI GASKET MONOPOLE

A. Introduction

The purpose of building this monopole is to produce an antenna that can operate at multiple frequencies while retaining the compactness property.

B. Selection of Operating Frequency

The goal of the Sierpinski monopole design is to achieve multi-band resonance [1]. To achieve that goal, it is useful to design the antenna to operate at a set of frequency bands commonly used by wireless devices. Two bands of interest are the 2.4 GHz and 5.0 GHz unlicensed bands which, respectively, are used for applications such as 802.11a/b and 802.11g wireless networking [4]. Although lower frequency bands are also useful for many applications, microstrip patch technology is most advantageous at higher frequencies where required trace element sizes are physically smaller.

V. RESULTS AND DISCUSSION

Software simulations test was used in order to evaluate the performance of the antenna design.

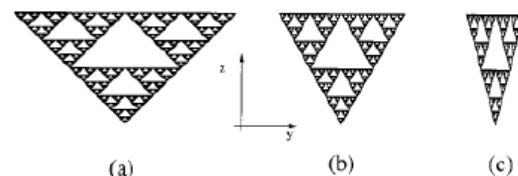


Fig. 3. Sierpinski gaskets with different α angle: (a) $\alpha = 90^\circ$, (b) $\alpha = 60^\circ$, (c) $\alpha = 30^\circ$.

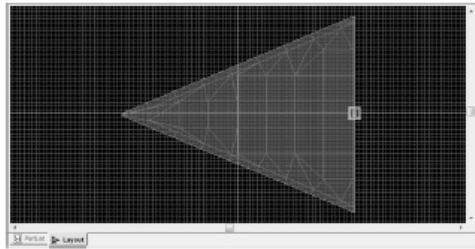


Fig. 4. First order fractal antenna geometry.

A. Sierpinski Gasket Monopole Results

First order Fractal Geometry & simulated S_{11} & VSWR of first kind pattern with $\alpha = 30^\circ$ value is given in Fig.4 &5 respectively. Second order Fractal Antenna Geometry & simulated S_{11} & VSWR of first kind pattern with $\alpha = 30^\circ$ value is given in Fig.6 &7 respectively.

CONCLUSION

Fractal Sierpinski Gasket Monopole antenna provides the good VSWR & response that is suitable for the Multiband frequency operation, Fractal geometry gives one advantage that because of the ISF original patch are made into smaller parts. This gives the multiple wavelength operation on single geometry. Every iterations wavelength is get divided. So on same geometry two frequency can be operated. This is continuing to nth iterations and which results n frequencies can be accommodated in single geometry.

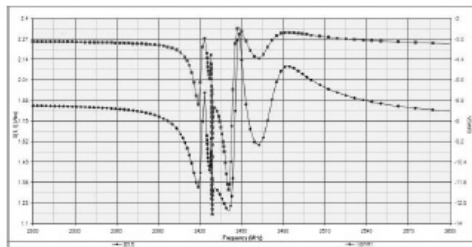
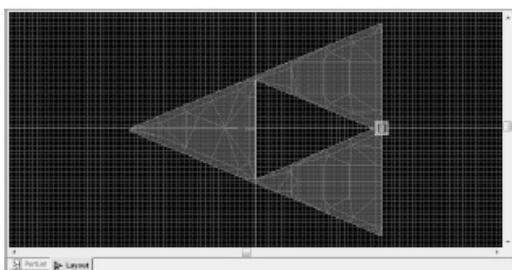
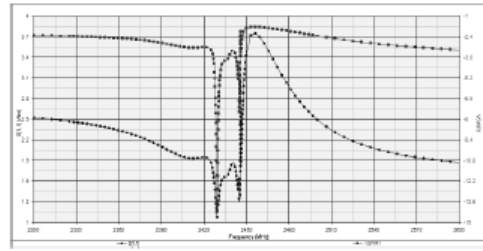
Fig. 5. Simulated result of S_{11} & VSWR of first kind pattern with $\alpha = 30^\circ$ 

Fig. 6. Second order fractal antenna geometry

Fig. 7. Simulated result of S_{11} & VSWR of second kind pattern $\alpha = 30^\circ$

REFERENCES

- [1] An Overview' of Fractal Antenna Engineering Research, Douglas H. Werner and Suman Gangul, IEEE Antennas and Propagation Magazine. Vol. 45, NO. 1, February 2003.
- [2] Fractal Antenna Engineering: The Theory and Design of Fractal Antenna Arrays, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, Vol. 41, No. 5, October 1999.
- [3] An Iterative Model for Fractal Antennas: Application to the Sierpinski Gasket Antenna, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 48, NO. 5, MAY 2000.
- [4] Behavior of the Sierpinski Multiband Fractal Antenna, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 46, NO. 4, APRIL 1998
- [5] A Modified Sierpinski Fractal Antenna for Multiband Application Kuem C. Hwang, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 6, 2007.

STUDY OF RELIABLE MPPT IN PV SYSTEM USING KALMAN FILTER

Sandhya Kumar

Vidyalankar Polytechnic
Electronics and Telecomm Engineering
sandhya.kumar@vpt.edu.in

Dr. Anjali Deshpande

Vidyalankar Institute of Technology
Electronics Engineering
anjali.deshpande@vit.edu.in

ABSTRACT

This paper presents study of maximum power point tracking (MPPT) method for photovoltaic (PV) systems using Kalman filter. MPPT is essential in PV cells in order to extract maximum power, but it is affected by different of noises. The Perturbation & Observation (P&O) method is widely used due to its easy implementation and simplicity. Dithering scheme can be used to reduce noise effects, but it slows the tracking response. Tracking speed is the most important factor for improving efficiency under frequent environmental changes. This paper presents study of Kalman Filter based technique for enhancing the reliability of PO method. An adaptive MPPT algorithm which uses an instantaneous power slope, along with Kalman filter is able to reduce tracking failures by the noises and also achieves fast tracking performance.

INTRODUCTION

A PV system is a power system that generates direct current (DC) from sunlight and processes it to generate alternating current (AC) electricity. As shown in Figure 1 the PV system consists of three stages: power generation, power conversion & conditioning, and power distribution.

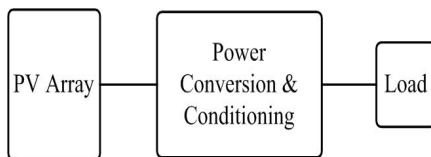


Figure 1: Block diagram of PV system

In the power generation stage, the PV array generates DC electricity directly from sunlight. The generated DC must be controlled and conducted in the power conversion & conditioning stage to suit distribution requirements.

The output current of PV panels varies depending on the status of the load. Over a wide range of I and V, it is very essential to find a point that maximizes the output power. The point is called the Maximum Power Point (MPP) and finding this point is called Maximum Power Point Tracking (MPPT).

Literature on MPPT techniques is abundant. There are many techniques to track MPP. These are Hill Climbing, Perturbation & Observation (P&O), Incremental Conductance (IncCond), Fractional Open-Circuit Voltage, Fractional Short-Circuit Current, Fuzzy Logic Control, Neural Network, Ripple Correlation Control (RCC), Current Sweep, DC-Link Capacitor Droop Control, Load Current or Voltage Maximization, dP/dV or dP/dI Feedback Control, and etc (Esram et al). Among these, the P&O and IncCond are most popular methods used in industry, for several reasons. The first is that the two methods can be easily applied to a microcontroller- or DSP-based design because both algorithms are very simple. Secondly, they are both independent of PV arrays; users have the flexibility to choose arrays. Both are expected to have accurate tracking performance under any environmental condition compared to other methods which are not tracking real MPP. Figures 2 and 3 show typical characteristics of PV solar cell for varying irradiances. It can be observed that the characteristics shift due to the effect of irradiance and maximum power is available when the irradiance is 100%. Moreover the characteristics are nonlinear and hence tracking MPP becomes critical under varying environmental conditions. In addition to this the Existing MPPT methods are also affected by the presence of noise in the sensors, thermal and noises in the power switches used.

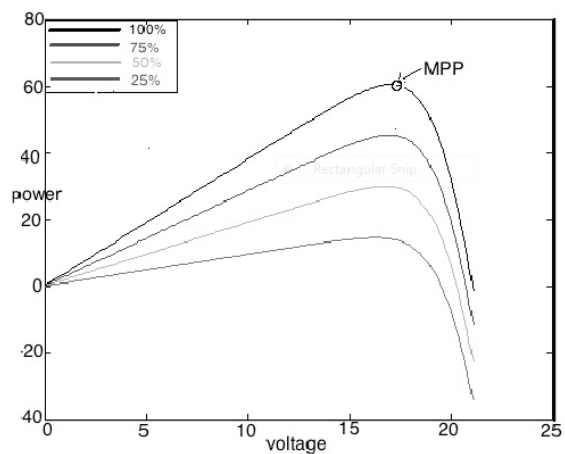


Figure 2: Power Voltage Characteristics of Solar Cell

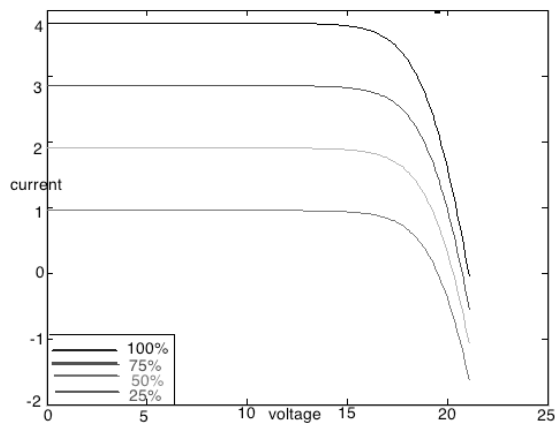


Figure 3: Current-Voltage characteristics of Solar Cell

As the above noises and errors may degrade the tracking performance, the MPPT controller is required to be robust against the noises and errors. This paper presents study of new MPPT method using Kalman filter in a PV system. Due to the estimation ability of the Kalman filter in the dynamic system with the noisy environment, the accurate MPP can be predicted by the Kalman filter without any drop of system dynamics. High computing amount of the Kalman filter may be a disadvantage. However, the present commercial microcontroller or DSP has from hundreds of megahertz (MHz) to a gigahertz (GHz). Those speeds may be assumed to overcome the Kalman filter's computations with several hundreds of hertz (Hz) speed of the MPPT controller.

PERTURBATION & OBSERVATION METHOD

The Perturbation & Observation (P&O) has been used mostly for its simplicity and easy implementation. It can be applied to both analog and digital designs. It is also compatible with any kind of PV arrays [13]. The P&O controller first measures voltage $V(k)$ and current $I(k)$ of a PV array and it calculates power $P(k)$, then controller compares between present voltage and previous voltage and then depending upon ΔP and ΔV it will take either positive and negative perturbation. The P&O sometimes shows tracking failure under Rapid Environmental Changes. This is shown in Figure 5.

The disturbances that affect the performance of MPPT control are caused by the thermal noise, switching noise, electromagnetic interference (EMI) etc. Switching noise is mostly generated by switching devices and it depends on a switching speed and size of the device. Also the voltage and current measurements will be corrupted due to the presence of sensors noise. Kalman Filter due to its excellent tracking capability and adaptive nature can be used to improve the performance of MPPT by reducing the effect of these disturbances. It is an optimal filter which uses recursive time and measurement updates over time to improve the quality of estimates [3].

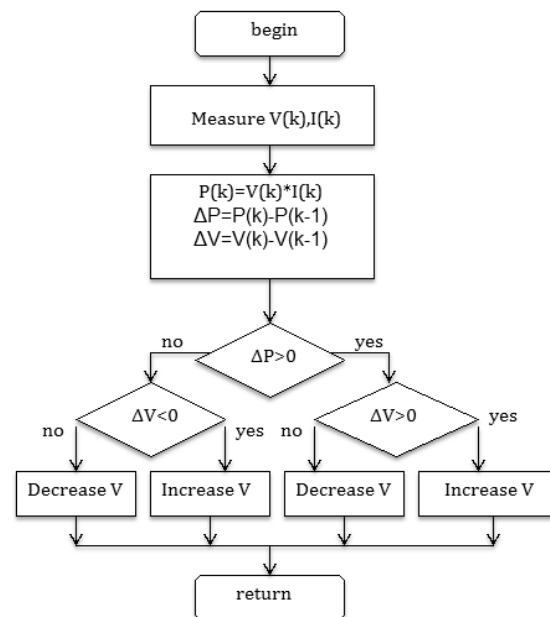


Figure 4: Algorithm for PO method

The P&O sometimes shows tracking failure under Rapid Environmental Changes. Figure 5 shows tracking failure of P&O method. This phenomena disturbs the controller to track towards the new MPP immediately when the environmental condition is changed. [1]

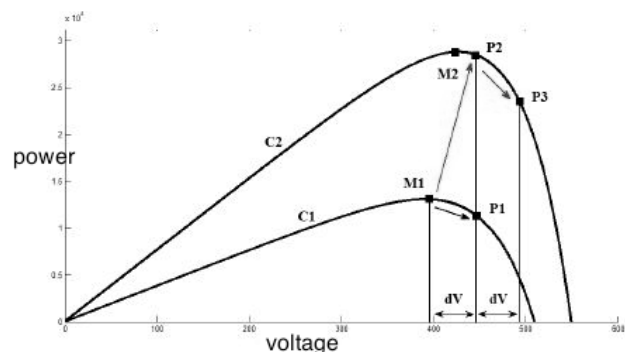


Figure 5: Tracking failure under rapid environmental change.

PROPOSAL OF NEW MPPT METHOD

Adaptive MPPT Algorithm

Figure 6 shows an adaptive MPPT algorithm. It uses instantaneous power slope as input to MPP controller. As shown in Figure 2, the P-V characteristic of a PV array is a simple convex function with one maximum peak point.

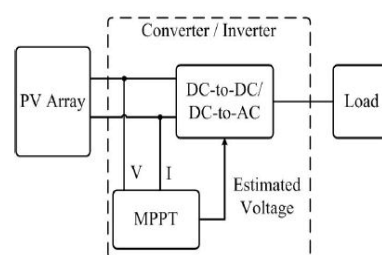


Figure 6: Block Diagram of Adaptive MPPT

The array power increases with a gradual positive slope until reaching the peak point, and decreases with a steep negative slope after passing the point. Based on the convex feature of the P-V characteristic, the tracking algorithm for a microcontroller-based design is the following.

$$V[k+1] = V[k] + M \frac{\Delta P[k]}{\Delta V[k]} \quad (1)$$

where M is a step size and $\Delta P[k]/\Delta V[k]$ is an instantaneous power slope at the solar array output. A converter or inverter controller samples voltages (V) and currents (I) of the array at every sampling time, and it updates the power slope in MPPT. Consequently, the estimated voltage $V[k+1]$ is derived from the updated slope at every sampling time. The instantaneous power slope can be approximately achieved by

$$\frac{\Delta P[k]}{\Delta V[k]} \approx \frac{P[k] - P[k-1]}{V[k] - V[k-1]}$$

The MPPT algorithm is adapted to the microcontroller- or DSP-based designs due to its easy application to a digital design. The algorithm is also capable of predicting the next voltage adaptively by the instantaneous power slope while the conventional methods move to the next voltage with the fixed value. Without errors in the system, an estimated voltage by MPPT is same as controlled array voltage. In this case, equation 1 is written with an actual voltage V_{act} as to:

$$V_{act}[k+1] = V_{act}[k] + M \left(\frac{\Delta P[k]}{\Delta V[k]} \right)_{true} \quad (2)$$

where $\Delta P[k]/\Delta V[k]$ true is the power slope calculated by the actual voltage and current.

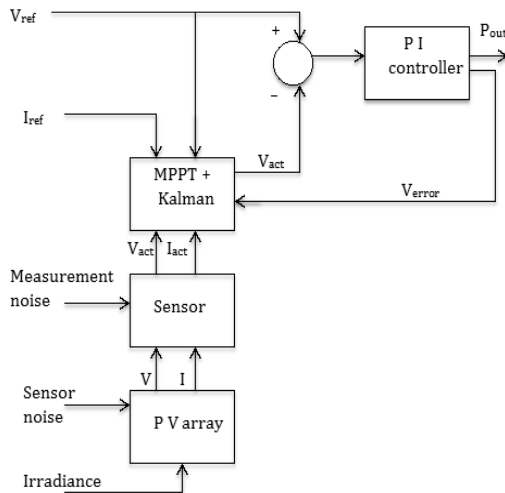


Figure 7:MPPT with Kalman Filter

In the real PV system, we must consider the effects caused by the mix of the disturbances and sensor errors. When the error by the mix of the disturbances is added as $w[k]$ and the error by the sensors is also considered to achieve the power slope, equation 2 changes to:

$$V_{act}[k+1] = V_{act}[k] + M \left(\frac{\Delta P[k]}{\Delta V[k]} \right)_{measured} + w[k] \quad (3)$$

where $\Delta P[k]/\Delta V[k]$ measured is the power slope measured by the sensors. $w[k]$ is assumed to be 0.01%, of the computed value of $V_{act}[k+1]$. In addition, the power $P[k]$ is achieved by the product of sensed voltages and currents as shown in Figure 8, and the calculation of measured $\Delta P[k]/\Delta V[k]$ contains the errors from both sensor accuracies. As illustrated in Figure 7, MPPT computes the input voltage of the converter, but there are errors due to difference between V_{act} and V_{ref} which is given by

$$V_{ref}[k+1] = V_{ref}[k] + M \left(\frac{\Delta P[k]}{\Delta V[k]} \right)_{actual} \quad (4)$$

and the difference between V_{act} and V_{ref} can be expressed by:

$$V_{ref}[k] - V_{act}[k] = \sum w[k]$$

where $\sum w[k]$ is the error between V_{ref} and V_{act} . Kalman filter is used to estimate the value of V_{act} . Instantaneous power slope $\Delta P[k]/\Delta V[k]$ is considered as an input signal of the system and therefore equation (3) can be assumed as one dimensional differential equation in linear state space representation and two known values V_{ref} and slope $\Delta P[k]/\Delta V[k]_{measured}$ can be used in kalman filter to estimate V_{act} in noisy environment. Thus it will help to improve tracking performance of MPPT controller by using optimal estimate of array voltage. PI controller is further used to remove the effect of other disturbances.

The recursive computation of the filter is divided by the measurement and time update steps.

MEASUREMENT UPDATE

From the error covariance $H[k]$ – from last time update, first of all, the Kalman gain $K[k]$ is computed in following:

$$K[k] = H[k] - (H[k] - R)^{-1}$$

where R is the measurement noise covariance. The measurement noise is assumed as disturbance which interfere an operation of both voltage and current sensors.

Next, the corrected $K[k]$ updates the estimate of $V_{act}[k]$ given by the sensor

$$\hat{V}_{act}[k] = \hat{V}_{act}[k] + K[k](V_{ref}[k] - \hat{V}_{act}[k])$$

In addition, the gain $K[k]$ also updates the error covariance $H[k]$ given by the $H[k]$ – in the following

$$H[k] = (1 - K[k])H[k] -$$

Time Update

Based on the updated estimate $V_{act}[k]$ and error covariance $H[k]$ of the measurement update, the succeeding estimate $V_{act}[k+1]$ is projected as:

$$\hat{V}_{act}[k+1] = \hat{V}_{act}[k] + M \frac{P[k] - P[k-1]}{V[k] - V[k-1]}$$

with a specific step size M. $H[k+1]$ – is also projected in the following:

$$H[k+1] = H[k] + Q \text{ where } Q \text{ is the process noise covariance of a plant.}$$

SIMULATION RESULTS

The simulation results given in [1] are obtained using MATLAB R2009b. Simulation result provided the comparison of the studied method and P&O method on transient response for sudden system restart and irradiance changes in different noise levels.

SUDDEN IRRADIANCE CHANGES

As we have seen PV arrays are very sensitive to rapid environmental changes so during sudden irradiance change tracking response of new method is good. Figure 8 shows the effect of irradiance on new method. It is assumed that the irradiance is suddenly reduced from 1 kW/m² to 0.5kW/m² with no temperature change.

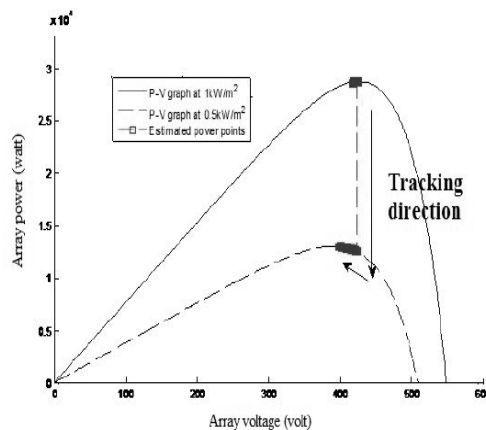


Figure 8 : Middle 200 estimation results of proposed method before and after irradiance change

Table 1
Generated Power Comparison of two methods over 1 minute in each scenario

Array Output power over 1 minute under 60dB SNR (percentage)		
Scenario	System restart	50% irradiance decrease
$\frac{\text{Kalman} - \text{P\&O}}{\text{P\&O}}$	+12.20%	+0.054%

Table 1 shows the comparison of accumulated output powers over a specific time. Table above shows that under two conditions system restart and irradiance decrease the new method generates more power than the conventional method. The environmental condition is limited only to 60dB process noise, 25°C, and 1 kW/m².

CONCLUSION

The P-V characteristic of the array is dependent of environmental conditions, especially temperatures and irradiances. Transient response of the MPPT algorithm during such changes is a key factor to improve the performance. The Kalman filter improves dynamic performance in the presence of such changes and also reduces the effect of measurement noise on the MPPT performance.

FUTURE WORK

Future work includes implementation of the algorithm in a real time microcontroller-based PV system, comparing performances of the two methods under a condition that irradiance and temperature are changed simultaneously. It can be further applied to mobile-based PV systems which need fast MPPT relatively due to movements,

REFERENCES

- [1] B O kang, maximum power point tracking method for PV cells using Kalman filter 2010,thesis.
- [2] R. Messenger and J. Ventre, Photovoltaic systems engineering: CRC, 2004.
- [3] G. Welch and G. Bishop, "An introduction to the Kalman filter," University of North Carolina at Chapel Hill, Chapel Hill, NC, 1995.
- [4] S. Julier and J. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," 1997, p. 26.
- [5] Math Works. R2010 Documentation - kalman [Online]. Available: <http://www.mathworks.com/help/toolbox/control/ref/kalman.html>
- [6] R. Kalman, "A new approach to linear filtering and prediction problems," Journal of basic Engineering, vol. 82, pp. 35-45, 1960.
- [7] L. Levy, "The Kalman filter: Navigation's integration workhorse," GPS World, vol. 8, pp. 65-71, 1997.
- [8] R. Engle and M. Watson, "The Kalman filter: Applications to forecasting and rational expectations models," 1987, p. 281.
- [9] R. Gencay, et al., An introduction to wavelets and other filtering methods in finance and economics: Academic Pr, 2002.
- [10] C. Chui and G. Chen, Kalman filtering: with real-time applications: Springer Verlag, 2009.
- [11] N. S. R. R. Laboratory. [Online]. Available: http://www.nrel.gov/midc/srnl_bms/
- [12] N. Mutoh, et al., "A method for MPPT control while searching for parameters corresponding to weather conditions for PV generation systems," IEEE Transactions on Industrial Electronics, vol. 53, pp. 1055-1065, 2006.
- [13] Y. Jung, et al., "Improved perturbation and observation method (IP&O) of MPPT control for photovoltaic power systems," 2005, pp. 1788-1791.
- [14] T. Markvart, Solar electricity: John Wiley & Sons Inc, 2000.

ROUTE OPTIMIZATION IN MOBILE MPLS

Trupti Patil

Electronics & telecommunication dept,
Sardar Patel institute of technology
Mumbai, India
trupti.bh87@gmail.com

Sukanya Kulkarni

Electronics & telecommunication dept.
Sardar Patel institute of technology
Mumbai, India
sk.extc@gmail.com

ABSTRACT

Integration of two technologies, Mobile IP and MPLS has gained popularity now days over dynamic mobile environment. This technology i.e. Mobile MPLS ensures to provide high Quality of service (QoS) along with mobility management. However Mobile MPLS inherits a problem of triangle routing from Mobile IP. Hence this research proposes a technique to improve the route optimization in Mobile MPLS network to improve the overall performance of the system. By adding a correspondent agent function to Mobile MPLS's edge routers, the mobility binding of a mobile node can be cached by the edge routers and the packet routing to the mobile node can be route-optimized. The simulation results indicate that the proposed route optimization has improved the performance of Mobile MPLS by reducing its end-to-end delay.

Keyword- *Mobile MPLS, Mobile IP, Multiprotocol Label Switching (MPLS)*

I. INTRODUCTION

The number of mobile users has rapidly increased in recent years. At the end of 2010, there are approximately 5.3 billion subscribers of mobile cellular worldwide, including 940 million subscribers of 3G services [1]. There is also a forecast that the mobile-only Internet users will grow from 14 million at the end of 2010 to 788 million by the end of 2015. This information indicates the trend towards more mobile Internet usage in everyday life. The demand for services through mobile networks has evolved from common voice communication to various Internet services. These services include email, web browsing, audio / video streaming, Internet phone, and real-time multimedia, which require different levels of quality of service (QoS). As mobile users demand for more of these services, the mobile networks must be able to provide appropriate level of QoS for each of these services.[1]

In view of this Mobile IP have been introduced which provides the services to the mobile users, such that any one from anywhere can have access to the services. Thus, providing Macro-mobility and Micro mobility as well. However, it does not guarantee the Quality of Service (QoS) provided to the services offered to the users. The packets used to travel through non-optimal

paths in terms of delay and bandwidth. Hence is not a successful technique to provide services in an efficient manner.

To overcome this problem, route optimization techniques have been introduced which helps to select optimal path from source to destination in terms of delay, bandwidth and jitter. However, the conventional IP tunneling of the packets is a time consuming process in which each intermediate router need to match the destination IP address. Hence a new technology was introduced namely Multi Protocol Label Switching (MPLS) which eliminates the use of conventional IP address matching and is faster routing technique. Hence, providing appropriate QoS to the applications [2].

II. MIP (MOBILE IP) AND MPLS IN PERSPECTIVE

A. Mobile IP

Mobile IP acts as core of mobility management mechanism for various networks, providing universal roaming solution. It support continuous and seamless connectivity, i.e. Mobile Node appears always on home network of the mobile device and can have access to any service at any time and any place. Hence with MIP, any device (node) willing to communicate with the mobile node can send the messages to mobile node's home address without any need to know about its current location. Each mobile node is identified by its permanent home address. While away from its home network, a mobile node is associated with a care-of address (CoA). MIP specifies how a mobile node registers with its home agent and how the home agent routes data packets to the mobile node through the tunnel. Such a network is inefficient in the terms that whenever any node will send the message to mobile node, the message needs to pass through Home Agent (HA) in its home network, which led to a triangle routing problem. Hence MIP alone is unable to provide adequate QoS to the network.[2]

Mobile IP components include home agent (HA), foreign agent (FA), mobile node (MN), and correspondent node (CN). MN has a permanent address called home address in the home network. When it moves to a new network, MN will get a temporary address called care-of address (CoA) in that foreign network. When a CN wants to

communicate with the MN, it will send packets to the MN's home address. If the MN is not in the home network, HA will intercept the packets and send them to MN in the foreign network through IP tunnel between HA and FA. Mobile IP helps make the movement of MN transparent to CN. While MN can send packets directly to CN, packets from CN must always pass through HA before reaching MN, which causes inefficiency problem called triangle routing problem.[4]

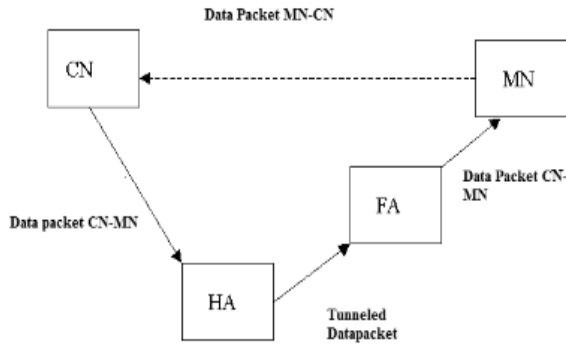


Figure 1: Triangle Routing Problem

B. MPLS Technology

To provide QoS along with mobility management, MIP has been integrated with MPLS (Multi Protocol Label Switching). MPLS provides for efficient routing, forwarding, switching and traffic flows through the network. It operates between the layer 2 (Data link Layer) and Layer 3 (Network layer) of OSI model. With MPLS packet-forwarding decisions are made solely on the contents of the label, without the need to examine the packet itself. Hence the latency is reduced here, as in conventional IP forwarding each router examines whole IP header in order to decide the next router.[2]

In MPLS, data transmission occurs on Label Switched Paths (LSPs). LSPs are a sequence of labels at each and every node along the path from source to destination. The labels are distributed using label distribution protocol (LDP). Each data packet encapsulates and carries the labels during the journey from source to destination in the MPLS domain. High speed switching of data is possible because the fixed length labels are inserted at the beginning of the packet and can be used by routers to switch packets quickly between the links.

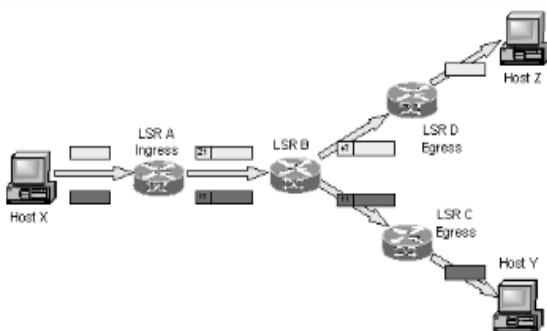


Figure 2: MPLS Network components

The routers are known as LSR and LER. LER i.e. Label edge routers which are at the edges of the network as shown (LSR ingress and egress) and LSR i.e. Label switching routers which are intermediate routers. Ingress LSR insert the label into the packet and egress pop up the label and forward it towards the intended destination.

C. Integration of Mobile IP & MPLS

By integrating the two technologies – Mobile IP and MPLS, a new technique have been developed known as MOBILE MPLS, which inherits the advantages of both the techniques and providing efficient system Mobile MPLS, the integration of Mobile IP and MPLS, can support both mobility and QoS as required by multiservice mobile networks. Future wireless networks are expected to provide IP based coverage and efficient mobility support with end-to-end QoS guarantees. Two enabling factors are considered as crucial: (i) maintaining the network connectivity during node mobility and; (ii) provisioning the network resources required by the Mobile Node (MN) in all the visited sub networks. PREVIOUS WORKS[9].

D. Mobile MPLS

Fundamentally, Mobile MPLS uses MPLS as a forwarding scheme for Mobile IP. The work in [2] proposed the integration of Mobile IP and MPLS by using MPLS label switching along LSP, instead of IP-in-IP tunneling, in the packet forwarding process from HA to FA. The packet forwarding is done by looking at MPLS labels without the need to go up to the IP header. Because label headers are much smaller than IP headers, the traffic overhead from HA to FA is also reduced.

The scheme proposed in [2] as described above has improved the scalability of Mobile IP and supports macro mobility in MPLS network. However, it is not efficient to support micro-mobility concerning frequent movements which require shorter handoff delay. To overcome this limitation, Micro Mobile MPLS [3] is proposed to improve the original Mobile MPLS to efficiently support micro-mobility by using the idea of Mobile IP Regional Registration (MIP-RR). In Micro Mobile MPLS, the hierarchical structure is added by including one LER Gateway (LERIGW) into each MPLS domain. When a mobile node has moved into an MPLS domain for the first time, it will register with the HA by using LERIGW address as the CoA. Subsequently, when the mobile node has moved to the new FA within the same domain under the same LERIGW, it will only need to regionally register with the LERIGW using the new FA as its new local CoA, and then the LSP from LERIGW to the new FA will be established as a path for sending packets between LERIGW and FA. New registration with the HA is not necessary for the intra-domain movement. The mobile node will need to register with its HA only when it has moved to a new MPLS domain.

The latest work [5] explained in details the operations of three mechanisms, namely, Fast Handoff (FH), Forwarding Chain (FC), and Master Forwarding Chain (MFC) Micro Mobile MPLS. Analytical models and simulation results indicated that the proposed mechanisms can substantially reduce registration update cost, handoff latency, and packet loss rate.

The work in [4] designed and implemented a generic Micro Mobile MPLS module for Network Simulator 2 (NS-2). The same paper also mentioned MiM-MPLS, which is another proposal on Micro Mobile MPLS. MiM-MPLS's architecture is also hierarchical. Its features include multicasting, buffering, and pre-registration. Additionally, it has ability to predict the mobile node's possible next location by detecting the mobile node's link layer movement. These capabilities help improve the handoff performance of the Micro Mobile MPLS.

E. Route Optimization

The Route Optimization in Mobile IP [2] was an IETF Internet draft proposed to solve the triangle routing problem in Mobile IP. According to the draft, CN may receive update for MN binding information from HA, and maintain this information in its binding cache. The draft defines Route Optimization messages which may be used to optimize the packet routing to the MN. These messages include Binding Warning, Binding Request, Binding Update, and Binding Acknowledge messages. However, there are some drawbacks for allowing CN to cache the binding information. First, CN will have to keep up with the MN current location.

In Profile based MPLS [6] is a scheme which helped to avoid the triangle routing problem. In this authors proposed that each mobile user maintains a profile that includes user regular behavior like user's mobility pattern, travel schedule etc. if the CN is able to obtain the profile of mobile user, it can predict the current location of user (to some extent of accuracy) and forward the traffic directly to the user, thus eliminating triangle routing problem. There are two approaches by which CN can access the profile of user. One is distributed approach, in which the mobile users itself distribute its profile after receiving the request from CN and performing authentication. But major drawback of this approach is that this increases the load on the mobile user. Second approach is centralized approach, in which there is a profile server which keeps the track of profiles of various mobile users and distribute them whenever receive a request. One major advantage of proposed scheme in [1] is that the CN can send all the packets to the mobile user whereas in other schemes like route optimizations some of the packets need to be forwarded to home agent in the beginning phase. However this scheme is not much successful because of increase in overhead signaling. Hence a lot of bandwidth resources get indulged in providing the nodes with the profile of users.

In Agent based MPLS While attempting to solve the triangle routing problem, the adopted route optimization places extra constraints on the network nodes. This work is motivated by the need to find an alternative solution that would effectively solve the triangle routing problem without sacrificing performance, transparency, and simplicity. The proposed solution is an agent-based route optimization, which moves the tasks of maintaining and updating binding caches and encapsulating messages away from individual correspondent nodes to the correspondent agents[7].

In this paper solved the weakness of related to having CN to be alert to MN location change by proposing an agent called Correspondent Agent into each network to maintain binding information for all the CN's in that network. This approach helps keep MN movement to be transparent to CN. However, the limitation of this work is that There must be one Correspondent Agent for each of the network in which CN is located. This is not practical and not scalable, because the network that CN is located can be any outside network and is not in our control to provide an agent for it. Moreover, this work was primarily proposed for Mobile IP, not Mobile MPLS.[8]

III. MOBILE MPLS WITH ROUTE OPTIMIZATION

A. The Proposed Mobile MPLS-RO

The route optimization technique introduced in [8], overcome the triangle routing problem along with providing requisite QoS. The authors introduced a correspondent agent (LCA) located at each Label edge router (LER) in each network. This LCA keeps track of current location of MN on behalf of CN. As these LCAs acts as LER, hence each incoming packet need to go through these agents to enter the intended network. LCA thus optimize the path of the packet without any need of additional agent since it already has the binding information of the MN.

The proposed LCA will have binding cache for keeping the current location of the contacted MN, so it can directly forward packets to the MN without passing through HA. Thus the triangle routing problem is solved, and the packet routing from CN to MN is route-optimized. In proposed route optimization protocol, three of the four Route Optimization messages defined in [4] are used, which include Binding Request, Binding Warning, and Binding Update messages. These messages and the binding cache at LCA are the main components of the proposed route optimization protocol.

Another advantage of having LCA as a correspondent agent at each edge router in the network is security consideration. As each LCA is part of the same network that LHA (or the home agent) is located, it is possible for LCA to establish security association with LHA in

advance and the binding update from LHA can be authenticated before being updated at the LCA's cache[10].

B. MobileMPLS-RO Hierarchical Architecture

The structure of Mobile MPLS-RO network topology as shown in Figure 1 is hierarchical to support micro-mobility. Its components are composed of LHA (label edge router / home agent), LF A (label edge router / foreign agent), LG (label edge router gateway), LSR (label switching router), MN (mobile node), CN (correspondent node), and LCA (label edge router / correspondent agent). As illustrated in Figure 1, the dotted lines indicate the boundary of the proposed Mobile MPLS-RO, while the networks outside the boundary are outside networks.

CN can be any computer or device wanting to communicate with MN in the Mobile MPLS-RO network. CN may be located in any outside network directly connected to the Mobile MPLSRO network, or it can be located in other distant networks connected to the core network.

The components of Mobile MPLS-RO network are almost similar to those in previous Micro Mobile MPLS research [5], except the addition of LCA at each edge router to work on behalf of the CN in caching mobility binding information of the MN and forwarding packets from the CN directly to the MN without the need to pass through the LHA as usual.

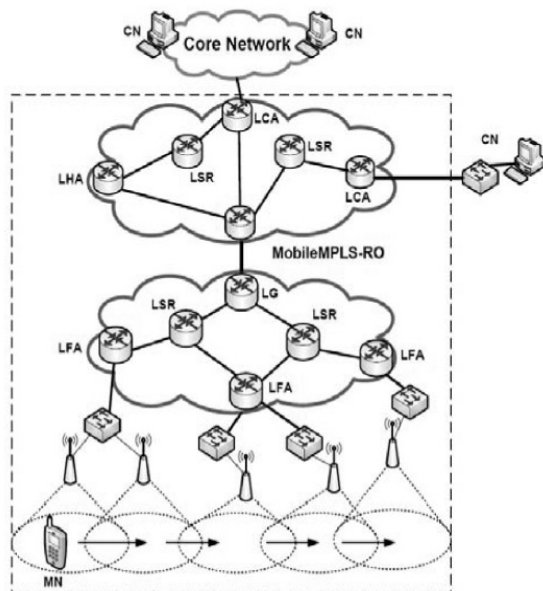


Figure 3: Mobile MPLS-RO hierarchical network structure

C. Route Optimization Protocol in MobileMPLS-RO

The main components of the proposed route optimization are (1) LCA and its binding cache, and (2) route optimization messages which are binding request (BR), binding warning (BW), and binding update (BU). These components work together to optimize the packet

routing. The operation of the proposed route optimization is as follows :

Phase 1: Pre-optimization Phase

In this beginning phase, the operation of Mobile MPLS-RO is the same as that of the regular Micro Mobile MPLS without route optimization. With LG as a gateway of each domain, when MN has moved into a domain for the first time, it will globally register with LHA by using IP address of the LG serving that domain as its CoA. And at the same time it will also regionally register with the LG by using IP address of the LF A at which the MN is currently located as its Regional CoA (RCoA). Afterwards, when the MN has moved to a new LF A in the same domain under the same LG, it will only have to update its current location by regionally register with the LG using the new LF A address as its new RCoA. The new registration with LHA is not necessary for the intra-domain.

Phase 2: Binding Update Phase

When the MN had registered to the HA with its new temporary CoA after moving to a foreign network, the HA keep its CoA in its cache. Thereafter, HA will forward this binding update information to all the CAs in its network. The CAs also stores the binding between the permanent IP address of the MN and temporary CoA, in its binding cache and starts a timer. When this timer runs out, that BU got expire and the CA either remove that BU from its cache or send a request to the HA for new update if cache is available. This passing of the binding information from HA to CAs is shown in figure 2. Thus every time whenever the HA receive any update from MN, it will forward this binding update to CAs such that the movement of MN is transparent to these agents and can provide optimized paths to the intended destination.

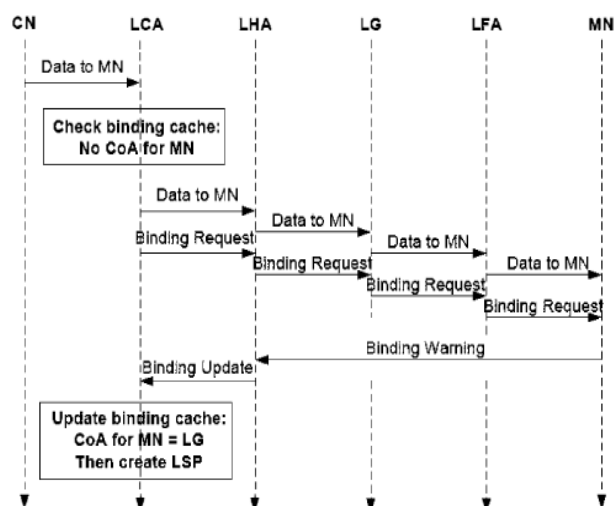


Figure 4: Binding update process before route optimization.

Phase 3: Route Optimization Phase

Now whenever any Correspondent node (CN) wants to send data to MN, it will send it to the home address of the MN towards the home network as it is not aware about the current location of the MN.

When the data packet enters the home network, it enters through CA as it is present at the edge of the network. The LCA after determining the destination of the packet, go through its binding cache to check the current location of the MN. Thereafter knowing the location, the LCA forwards the data packet directly towards the MN, thus optimizing the path. Hence the triangle routing problem is avoided. However, this is more efficient as compared to [1], as even in the beginning phase there is no need to send the data to HA. If the CA initially does not have the current location of the MN, it had to forward the data towards HA, which afterwards tunnel the data towards the MN. Thus the data had to follow a long path with a number of intermediary nodes to the HA, resulting into a significant delay which can reduce the performance of the whole network. Hence reducing the delay and improving the performance of the network.

Figure 5 illustrates the route optimization phase. During this phase if the MN moves to a new FA under the same LG, the route optimization will not be interrupted and work as the same before this inter-domain movement as the CoA for the MN remain the same gateway (LG).

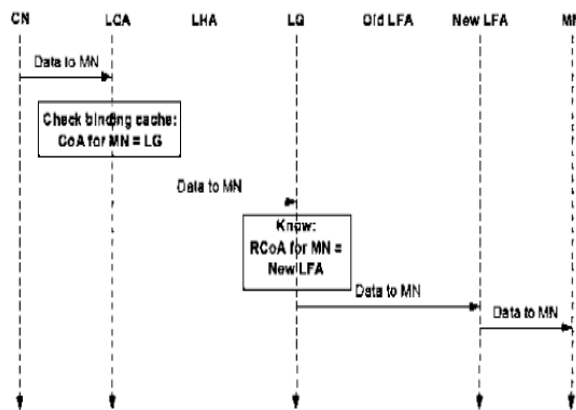


Figure 5: Forwarding the data from CN to MN with optimized route

Phase 4: Re- Optimization Phase

There may be probability of MN to move to a new domain under new gateway (LG). In such a case the MN is required to register its new CoA with its HA, such that the HA is aware of the movement of the MN.

After receiving the binding update from MN, HA will update its cache and simultaneously forwards the same to the CAs of its network, hence making them aware of the movement of the MN. The CAs thereafter updates their binding cache with new CoA of the MN and establishes a new LSP accordingly. Figure 6 shows the re-route optimization phase.

Next when the CA receives a packet destined towards the MN, it will check its binding cache and forwards the packet towards the new location of the MN with new CoA as new LG.

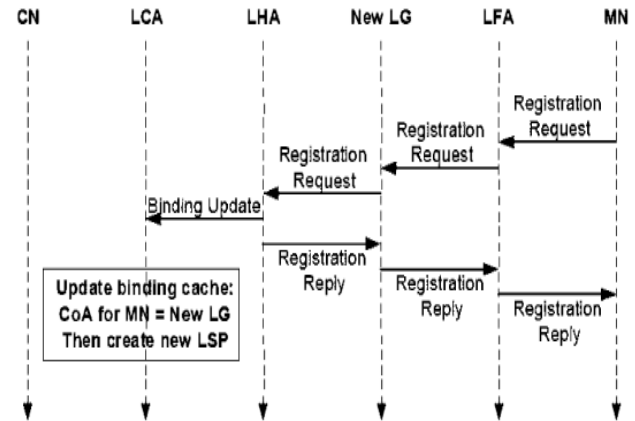


Figure 6: Re-optimization process after MN's inter-domain movement

IV. SIMULATION STUDY

In the simulation, the end-to-end delays of each packet in both cases, with and without route optimization (RO), are measured and compared. Figure 7 and Figure 8 show sending and receiving times of each packet in the case without and with RO respectively. The x-axis is time in seconds and the y-axis is packet sequence number. The upper line in each graph shows the packets sent from the source (CN), and the lower line shows the packets received at the destination (MN). So the sending and receiving times for each packet sequence number can be read from the graphs, and the end-to-end delay can be known.

In each graph, point A is the handover point at which MN has moved from LHA to LFA1, and point B is the handover point at which MN has moved from LFA1 to LFA2. Some packet losses during handover can be noticed from the graphs.

For the case of Mobile MPLS without RO, at the beginning when MN is at LHA, the sending and receiving times of each packet are close together (i.e. short delay). And when MN has moved to LFA1, the sending and receiving times of each packet are far apart (i.e. longer delay). This simulation results confirm that without RO, the end-to-end delay of each packet will get longer when MN is not in its home network due to the triangle routing problem.

In the other case of our proposed Mobile MPLS-RO as shown in Figure 7, when MN is at LHA, the delay will be the same as when it is at LHA in the former case, because the route optimization is not necessary and not in effect when MN is at home. Subsequently, however, when MN has moved from LHA to LFA1 the sending and receiving times of each packet will get closer together as appeared to be the same line in the graph in Figure 8 (i.e. very short delay). The results indicate that the proposed route optimization improves end-to-end delay of each packet when MN is not in its home network. This is because the proposed route optimization eliminates the triangle routing.

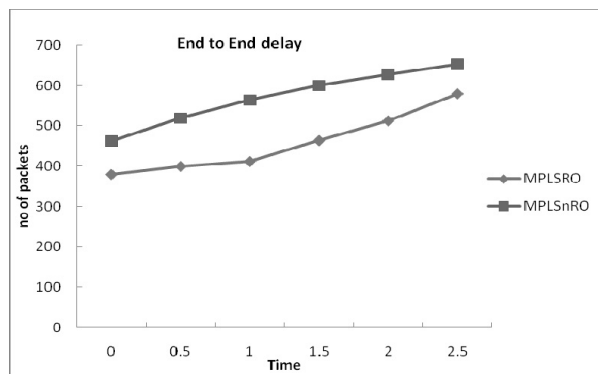


Figure 7: Packets sent from CN and received at MN for Mobile MPLS with and without RO.

From figure 8 graph we compare the end –to-end delay of packet in both cases with and without route optimization (RO), we get less delay using route optimization technique as compare to without route optimization technique.

From Figure 9 We come to know Bandwidth utilization is better in MPLSRO method as compare to MPLSnRO because due to route optimization technique triangular routing problem is overcome due to that maximum utilization of resource.

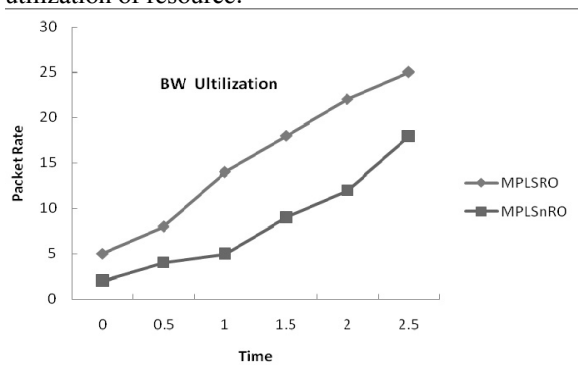


Figure 9: End to end delay of packet in both cases

From Figure 10 Packet delivery Ratio is better in MPLSRO method as compare to MPLSnRO. Optimization technique does not show better performance when MN is at home location. As MN move to LHA to LFA sending and receiving of packet is much better due to optimization of technique so we get better packet delivery ratio in route optimization technique.

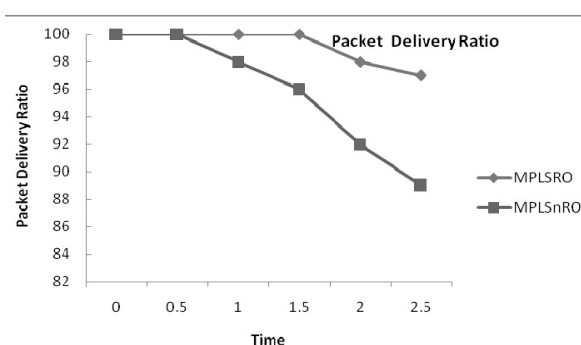


Figure 9: Packet delivery Ratio in MPLSRO and MPLSnRO

CONCLUSION

This paper proposes a route optimization protocol to overcome triangle routing problem in Mobile MPLS. The main components of the proposed route optimization are (1) correspondent agent and its binding cache at each of Mobile MPLS's edge router called LCA (label edge router / correspondent agent), and (2) route optimization messages which are binding request (BR), binding warning (BW), and binding update (BU). These components work together to provide the route optimization for Mobile MPLS, and consequently the triangle routing is eliminated. The simulation study is done to evaluate the proposed mechanism. The simulation results indicate that the proposed route optimization increases the Mobile MPLS performance by reducing its end-to-end delay when the mobile node is away from home.

REFERENCES

- [1] Savinya Prawit Chumchu, Somphol Sirisaingarn, Thawatchai Mayteevarunyou, (2011), "Performance Analysis and Improvement of Mobile MPLS", 2011 International Conference on Information Networking (ICOIN), pp 317-22 .
- [2] Z. Ren, C. Tham, C. Foo, and C. Ko, "Integration of Mobile IP and Multi-Protocol Label Switching," IEEE International Conference on Communications (IEEE ICC2001), Finland, June 2001
- [3] Rami Langar and Samir Tohme, "Micro Mobile MPLS: A Solution For Micro-Mobility Management In Next generation Networks", IEEE Wireless Communications Magazine, vol. 9, pp. 2-12, February 2004.
- [4] T. Sanguanwongthong , and P. Chumchu, "Design and Implementation of Micro Mobile MPLS for NS-2," the 3rd International Conference on Performance Evaluation Methodologies and Tools (ValueTools 2008), Greece, October 2008
- [5] R. Langar, G.L. Grand, and S. Tohme, "Micro Mobile MPLS Protocol in Next Generation Wireless Access Networks," The 9th CDMA International Conference 2004 (CIC 2004), 2004
- [6] Tingzhou Yang, Yixin Dong, Bin Zhoy Dimitrios Makrakis, (2002) "Profile Based Mobile MPLS Protocol", IEEE CCECE 2002 Canadian Conference On Electrical And Computer Engineering, Vol 3, pp 1352-6.
- [7] R. Vadali, J. Li, Y. Wu, and G. Cao, "Agent-Based Route Optimization for Mobile IP," IEEE 54th Vehicular Technology Conference, October 2001
- [8] Abd-Elhamid M. Taha, Hossam Hassanein and Hussein T. Moutfah, (2004) "Integrated Solutions For Wireless MPLS And Mobile IP: Current Status and Future Directions", 2004. Canadian Conference on Electrical and Computer Engineering, Vol. 3, pp 1463-66.
- [9] Shengling Wang, Yong Cui, Sajal Das, Mingwei Xu, (2008) "Optimized Mobile MPLS" , 2008 IEEE International Conference on Communications Workshops, pp 441-5.
- [10] Savinya Polvichai and Prawit chumchu, (2011) "MOBILE MPLS with Route Optimization: The proposed protocol and simulation study ", 2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE), pp 34-39.

IMAGE PROCESSING APPROACH ON DETECTION OF BRAIN TUMOR CELLS

Mohd. Ashfaque

Rizvi college of Engg. India
Information Technology, Mumbai University
mhdashfaque@yahoo.co.in

Varsha Bhonsale

Vidyalankar Institute of Technology,.India
Information Technology , Mumbai University
versha.bhosale@vit.edu.in

ABSTRACT

During past few years, brain tumor segmentation in magnetic resonance imaging (MRI) has become an emergent research area in the field of medical imaging system. The proposed method can be successfully applied to detect the tumor and its size. This technique can be handy tool for the practitioners especially the physicians engaged in this field. This paper describe segmentation method consists of two phases. In the first phase, the MRI brain image is acquired from patient and noise is removed. In the second phase (MR) image segmentation is to accurately identify the principal tissue structures in these image volumes. A new unsupervised MR image segmentation method based on C-Means clustering algorithm for the Segmentation is presented. At the end of the process the tumor is extracted from the MR image and its exact position and the shape also determined. The stage of the tumor is displayed based on the amount of area calculated from the cluster.

Key words: Tumor, k-means segmentation, thresholding, binarization.

I. INTRODUCTION

Tumor is an uncontrolled growth of tissues in any part of the body. Tumors are of different types and they have different Characteristics and different treatment. In general, tumors appear to occur when there is a problem with the dividing of cells in the body [1]. The tumor may be primary or secondary. If it is an origin, then it is known as primary. If the part of the tumor is spread to another place and grown as its own then it is known as secondary. Normally brain tumor affects CSF (Cerebral Spinal Fluid). It causes for strokes. The physician gives the treatment for the strokes rather than the treatment for tumor. So detection of tumor is important for that treatment. The lifetime of the person who affected by the brain tumor will increase if it is detected at current stage. That will increase the lifetime about 1 to 2 years.

Normally tumor cells are of two types. They are Mass and Malignant. The detection of the malignant tumor is somewhat difficult to mass tumor. For the accurate detection of the malignant tumor that needs a 3-D representation of brain and 3-D analyser tool.

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to visualize internal structures of the body in detail. MRI makes use of the property of nuclear magnetic resonance (NMR) to image nuclei of atoms inside the body [2]. MRI provides good contrast between the different soft tissues of the body, which makes it especially useful in imaging the brain, muscles, the heart, and cancers compared with other medical imaging techniques such as computed tomography (CT) or X-rays. Unlike CT scans or traditional X-rays, MRI does not use ionizing radiation.

II. EXISTING METHOD

The Segmentation of an image entails the division or separation of the image into regions of similar attribute. The ultimate aim in a large number of image processing applications is to extract important features from the image data, from which a description, interpretation, or understanding of the scene can be provided by the machine. The segmentation of brain tumor from magnetic resonance images is an important but time-consuming task performed by medical experts The digital image processing community has developed several segmentation methods[3], many of them ad hoc. Four of the most common methods are: 1) amplitude thresholding, 2) texture segmentation 3) template matching, and 4) region-growing segmentation. It is very important for detecting tumors, edema and necrotic tissues. These types of algorithms are used dividing the brain images into three categories (a) Pixel Based (b) Region or Texture Based (c) Structural Based. Several authors suggested various algorithms for segmentation.

III. WORKFLOW

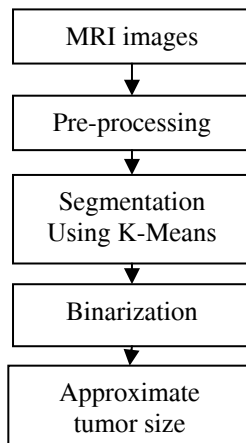


Figure 1: Basic Workflow

The workflow basically consist of five parts: (i) MRI images is the input (ii) pre-processing (iii) segmentation uses K-means clustering (iv) binarization method which uses threshold value for converting images into binary form (v) calculation of size of tumor and their stage

IV. IMAGING

MRI images are magnetic resonance images which can be acquired on computer when a patient is scanned by MRI machine. We can acquire MRI images of the part of the body which is under test or desired.

V. PRE-PROCESSING

In signal processing, it is often desirable to be able to perform some kind of noise reduction on an image or signal. The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise. The main idea of the median filter is to run through the signal entry by entry, replacing each entry with the median of neighbouring entries. The pattern of neighbours is called the "window", which slides, entry by entry, over the entire signal. For 1D signals, the most obvious window is just the first few preceding and following entries, whereas for 2D (or higher dimensional) signals such as images, more complex window patterns are possible (such as "box" or "cross" patterns). Note that if the window has an odd number of entries, then the median is simple to define: it is just the middle value after all the entries in the window are sorted numerically. For an even number of entries, there is more than one possible median. This filter enhances the quality of the MRI image [4].

VI. K-MEANS CLUSTERING

K-Means is the one of the unsupervised learning algorithm for clusters. Clustering the image is grouping the pixels according to the some characteristics. In the k-means algorithm initially we have to define the number of clusters k . Then k -cluster centre are chosen randomly. The distance between the each pixel to each cluster centres are calculated. The distance may be of simple Euclidean function. Single pixel is compared to all cluster centres using the distance formula. The pixel is moved to particular cluster which has shortest distance among all. Then the centroid is re-estimated. Again each pixel is compared to all centroids. The process continuous until the centre converges.

A. Flowchart of k-means clustering

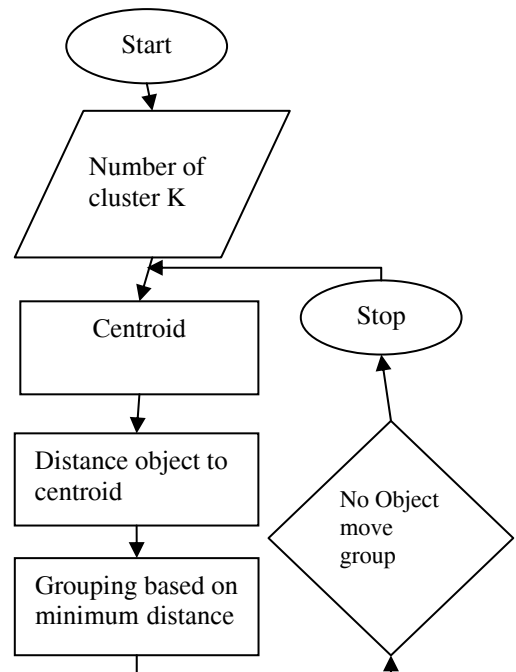


Figure 2: Flowchart of K-Means [5]

B. Algorithm of k-means clustering

- Step1: Define the number of cluster.
- Step2: K cluster centre numbers are chosen randomly.
- Step3: Calculate mean or centre of cluster.
- Step4: The distances between each pixel to each cluster centre are calculated. [6]

$$D(i) = \arg \min \|X_i - M_k\| \quad i=1, \dots, N$$

- Step5: Single pixel is compared to all cluster centres using distance formula.
- Step6: The pixel is moved to particular cluster which has shortest distance.
- Step7: Otherwise move to next cluster.
- Step8: Centroid is re-estimated.
- Step9: Each pixel is compared to all centroid.
- Step10: The process continues until centre converges.

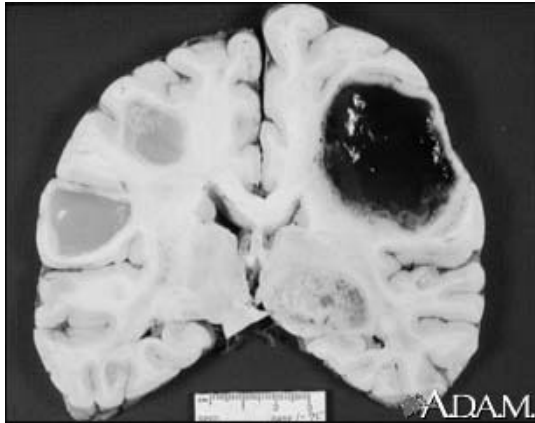


Fig. 3. Input image of k-means clustering

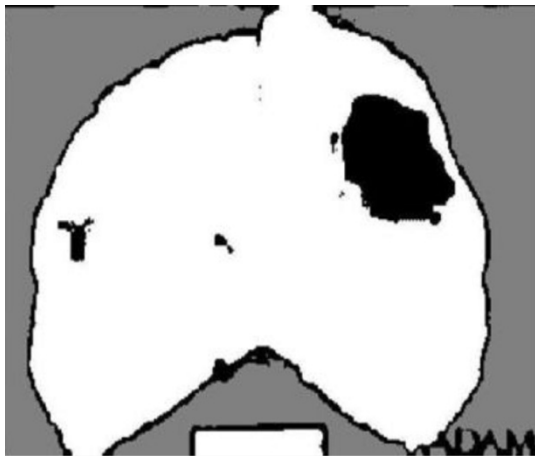


Fig. 4. Output image after k-means clustering

Fig. 3 is given as the input to the k-means clustering algorithm. The k-means clustering algorithm clusters the given image with some characteristics and based on their characteristics the tumor is extracted. Fig. 4 shows the output image of k-means clustering which has tumor shown with black color that will be helpful for detecting tumor in a brain.

VII. BINARIZATION

The extracted cluster is given to the thresholding process. It applies binary mask over the entire image. It makes the dark pixel become darker and white become brighter. In threshold coding, each transform coefficient is compared with a threshold. If it is less than the threshold value then it is considered as zero. If it is larger than the threshold, it will be considered as one. Suppose the given image is f then we should perform following operations on it [7].

$$f(x,y) = \begin{cases} 1, & \text{if } f(x,y) > T \\ 0, & \text{if } f(x,y) \leq T \end{cases}$$

C. Binarization Algorithm

Step1: compute mean or threshold of array (Say T)

Step2: for $i = 1$ to ht do
 for $j = 1$ to wd do
 $p \leftarrow X[i,j]$

Step3: if $(p \leq T)$ then begin
 $opim(i,j) \leftarrow 0$;
 else
 $opim(i,j) \leftarrow 255$;



Fig. 5. Output image after Thresholding

Fig. 5 shows the threshold image of the image shown in fig. 4 that is output image of k-means clustering, which clearly shows the tumor in the brain in the form of black shape.

VIII. ALGORITHM FOR BRAIN TUMOR SIZE CALCULATION

Step 1: Start the process.

Step 2: Get the binarization image as input.

Step 3: Calculate the number of black points in the image. [8]

$$\text{No. of black pixel } P = \sum_{W=0}^{255} \sum_{H=0}^{255} [f(0)]$$

Where p = no. of black pixel (width*height)
1 pixel = 0.264 mm

Step 4: Calculate the size of the tumor using the formula. [9]

$$\text{Size of tumor } S = [(\sqrt{P}) * 0.264] \text{ mm}^2$$

Where p = no. of black pixel.

Step 5: Display the size and stage of tumor.



Fig. 6. Output image after binarization

Fig. 6 is the binary output image of the input image shown in fig. 5. Here we are applying filtering techniques for removing the noise available in fig. 5 that is other than tumor and after that converting into binary form.

Here we are creating binary array of whole image and counting number of one's and zero's (white and black pixels) and then calculating the size of tumor that is collection of black tumor. The stage of tumor is based on the area of tumor. We considered that, if the area is greater than 5 mm^2 it will be the critical position.

CONCLUSION

The study was conducted to detect brain tumor using medical imaging techniques. The main technique used was segmentation, which is done using a method based on thresholding, k-means segmentation and filtering techniques, using this flow we can detect the size, stage of tumor. After performing all these operations on image as shown in fig. 3 of brain tumor we got the result accurately that shape of tumor is shown in fig. 6 after binarization and the size of tumor is to be calculated as 8.8035 mm^2 , which is greater than a critical size.

REFERENCES

- [1] <http://www.medhelp.org/medical-information/show/777/Tumor>.
- [2] http://www.diffen.com/difference/CT_Scan_vs_MRI.
- [3] Dunn, J.C., 1973." A fuzzy relative of the ISODATA process and its use detecting compact, well Separated clusters", Journal of Cybernetics, 3: 32-15
- [4] Brain Tumour Extraction from MRI Images Using MATLAB Rajesh C. Patil, Dr. A. S. Bhalchandra
- [5] <http://people.revoledu.com/kardi/tutorial/kMean/Algorithm.htm>
- [6] S.Mary Praveena ,Dr.Ila Vennila , June 2010, "Optimization Fusion Approach for [Image Segmentation Using K-Means Algorithm," International Journal of Computer Applications (0975 - 8887) Volume 2 - NO.7.
- [7] A. Suman Tatiraju, July-2008, "Image Segmentation using k-means clustering, EM and Nonnormalized Cuts," Symposium of Discrete Algorithms
- [8] D. Van De Ville, M. Nachtegaal, D. Van Der Weken, E.E. Kerre, W. Philips, I. Lemahieu (2003), "Noise reduction by fuzzy image filtering," IEEE Transactions on Fuzzy Systems .
- [9] H.R. Tizhoosh, G. Krell, and B. Michaelis, (1997) "On fuzzy enhancement of megavoltage images in radiation therapy," in: Proceedings of the 6th IEEE International Conference on Fuzzy Systems 3, pp. 1398-1404.

SEAMLESS STITCHING OF IMAGES USING NORMALIZED CROSS CORRELATION

Amey S. Gawde

Department of Electronics and
Telecommunication Engineering
VIT, Wadala East, Mumbai-37.
amey.gawde@vit.edu.in

Geetha Narayanan

Dept. of Electronics Engineering
VIT, Wadala East, Mumbai-37.
geetha.narayanan@vit.edu.in

Uday Pandit Khot

Dept. of Electronics & Tele. Engg.
Saint Francis Institute of Technology,
Borivali West, Mumbai
udaypandit@rediffmail.com

Nandesh Chiplunkar

Dept. of Electronics Engineering
VIT, Wadala East, Mumbai-37.
nandesh.chiplunkar@vit.edu.in

ABSTRACT

A method for seamless stitching of images with photometric inconsistencies in the overlapping region is presented. The method deals with first applying the cross correlation and normalizing the cross correlation to get the exact overlapping region. After getting the normalized cross correlation value the image is stitched at that point. Experimental results illustrate the method and show that it leads to seamless mosaic images despite differences in the overlapping region of input images.

I. INTRODUCTION

The combination of two or more images to get very big and high resolution image is called as Image Mosaicing. There are many applications where large view field images are of great importance. In areas ranging from medical imaging to computer graphics and satellite imagery, a computationally efficient and easy to implement method to produce high resolution wide angle images will continue to draw research interest. Two major steps were required namely image registration and blending.

Image registration, is typically [1] done by choosing one of the images as the reference and then finding the geometric transformations. Intensity as well as feature based registration algorithms are presented in [2] and in [3], [4] for evaluating the quality of images. Previously proposed image blending algorithms operate on image intensities or image gradient, at full resolution of the image or at multiple resolution scales. A well-known intensity based technique is feathering [5] where the mosaic is generated by computing a weighted average of the source images. In the composite mosaic image, pixels are assigned weights proportional to their distance from the centre of the image they come from, resulting in a smoother transition from one image to the other in the final mosaic. As it is discussed in [6], this method may generate blurring and ghosting artifacts when the images are not registered properly. The first intensity

based multiscale method is pre-sented in [7] and relies on pyramid decomposition. The source images are decomposed into bandpass components and the blending is done at each scale, in a transition zone inversely proportional to the spatial frequency content in the band. A similar approach in [8] replaces pyramid decomposition with wavelet decomposition. There are gradient based techniques for image processing applications such as image editing[9-11], high dynamic range imaging and compression[12], [13] and image mosaicking[14], [6], [15], [16]. In all these methods the gradients of the source images are modified and the modified gradients are used to obtain the final image. In general, these gradient modifications lead to a gradient which typically is not a conservative vector field, and image reconstruction from this gradient no longer has an exact solution, but can be formulated as an exact problem. A general framework for image reconstruction from non-integrable gradients is discussed in[17].

The methods used for image mosaicking in [18] had to de-compose the image, blend the image, find the overlapping re-gion curve, intensity matching and again Poisson and wavelet transform for smoothening of images. All these redundancies can be overcome by using the method proposed in this method and also the seam is eliminated. In this paper a new method for seamless mosaicking of images which are assumed to be registered is presented. Two cropped images of any size can be stitched seamlessly and the size of the final image is less as compared to the combined size of the two images when they are actually added. Chapter II discusses about the proposed technique and III discusses the results and analysis.

II. PROPOSED TECHNIQUE FOR SEAMLESS STITCHING OF IMAGES

This chapter deals with the proposed technique for finding the overlapping region using normalized cross correlation and the method of stitching the two images having overlapping region common in them.

A. Technique

The proposed method consists of inputting two images of same horizontal dimensions and stitching them. The two most important steps involved are image registration and image stitching. In this method, in image registration, two colour images are taken as input having same vertical dimensions and variable horizontal dimensions. Fast Fourier transform could also have given results but noise plays a major role in it. The noise appears as a bright spot in the Fast Fourier transform. To remove noise Band Reject filter or Notch filter is required and then Inverse Transform would be taken. But to avoid such redundancies and time complexities normalised cross correlation can be used.

The cross correlation is calculated between two images. The cross correlation values are then normalized to get the exact cross correlation values. The second image is cropped by 40% both length wise as well as height wise to get the exact cross correlation between the second image and the first one. The second image is made 40% of the original value in order to remove the time redundancies which would have been introduced if the size of the image is the whole image itself. Also if normalized cross correlation is used the size of the template should be smaller than the size of the image. If the cross correlation is used to find the overlapping region, it will give the overlapping region or the cross correlation between the two images but the region may not be exact because the correlation is calculated by multiplying the template which is the fixed part but the it is moved over the entire first image which is continuously changing. So to avoid these con_jictsthe cross correlation is normalized by dividing both the variable parts to get a value which is fixed. This value can be obtained by normalization of the cross correlation. The normalized cross correlation gives the exact overlapping region as compared to other techniques discussed above. Once the overlapping region is found out the position of the maximum normalized cross correlation value is found out.

Once registered, the images can be combined or fused in a way that improves information extraction. The images are then stitched at the point of the maximum cross correlation by eliminating common region from the first image. The images which are stitched have the same value as the origin al image. The final image is the stitched image without any seam.

B. Algorithm

The algorithm for the proposed method is as follows:

1. Input two colour images having same vertical dimensions and any horizontal dimensions having some overlapping region. Consider first image as $g(x, y)$.
2. Second image is made 40% of the original size and consider it as image $f(x, y)$.

3. Find the cross correlation between g and fusing

$$\int_{-\infty}^{\infty} f(x, y) \cdot g(x + u, y + v) dx dy$$

4. Finding value to normalised the cross correlation using the function

$$\int_{-\infty}^{\infty} [g^2(x + u, y + v) dx dy]^{\frac{1}{2}}$$

5. Calculate normalized cross correlation using the results obtained in 3 & 4 function.

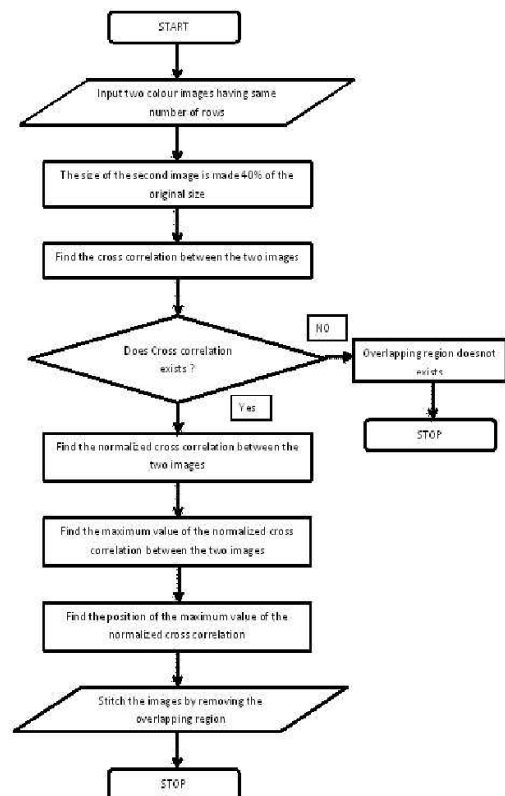


Fig. 1. Flowchart

6. Find the maximum value of cross correlation.

7. Find the position of the maximum overlapping region which is the maximum value of cross correlation.

8. Stitch the images at this point by eliminating the over-lapping region.

9. The final output is the mosaicked image without any seam.

C. Flowchart

The above Fig 1. describes the operations performed on the images to be stitched.

The two images having same number of rows but

different number of columns that is same vertical dimensions but different horizontal dimensions are taken as the input. The images which are taken as input can be colour images or grey scale images. Then the second image is cropped at the 40% row wise as well as column wise. This is done because the normalised cross correlation function requires the size of the template to be smaller than the size of the image to be compared with. Consider first image as $g(i,j)$ and second image as $f(i,j)$. At the beginning, cross correlation is calculated between the two images using equation (1) :

$$\int_{-\infty}^{\infty} f(x, y).g(x + u, y + v)dx dy \quad (1)$$

The cross correlation will also give the overlapping region but since in the above equation the second part is variable which changing with respect to the first image. To remove this and to make the values stable the cross correlation is divided by the variable part which is given by the equation (2)

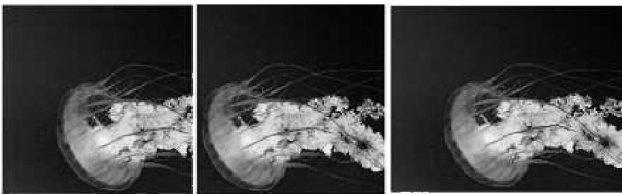


Fig. 2. Left to right (a) Input image 1, size: 512x512 pixels (b) Input image 2, size : 512x512 (c) Result obtained using the proposed method. Mosaic image : 512x644 pixels, overlap region: 80%

$$\int_{-\infty}^{\infty} f(x, y).g(x + u, y + v)dx dy \quad (2)$$

The cross correlation is divided by this value and the normalized cross correlation is calculated by the equation(3):

$$\int_{-\infty}^{\infty} f(x, y).g(x + u, y + v)dx dy \quad (3)$$

III. RESULT AND DISCUSSIONS

In this section the performance of the proposed method is illustrated with experiments. Given two partially overlapping, previously registered images the objective is to stitch them and produce a larger mosaic image which looks smooth and continuous without any noticeable artifacts such as seams or blurring.

The algorithm was implemented using Matlab 7.12.0.635(R2011a) 32bit (win32) and was running on 32 bit PC Architecture Intel^{IR} AtomTM CPU N2660, 1.60GHz. The technique is tested with three different overlapping regions having same size of 512x512

pixels. Different observations were noted, he time taken to stitch the two images with same size and different amount of overlapping as 80%, 50% and 25% was 15.63 seconds. Fig. 2a is one of the input images and Fig. 2b is another image to be mosaicked. The two images have the same size of 512x512 pixels which is kept same in order to check the time complexity involved during the computations. Fig. 2c is the mosaicked colour image having size 512x644. The maximum normalized cross correlation value in this case is 0.9998 and percentage of overlapping region is 80%. The two images are stitched at this maximum value.

Fig. 3a is one of the input images and Fig. 3b is another image to be mosaicked. The two images have the same size of 512x512 pixels which is kept same in order to check the time complexity involved during the computations. Fig. 3c is the mosaicked colour image having size 512x763. The maximum normalized cross correlation value in this case is 0.9998 and

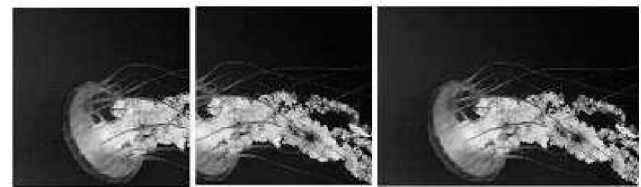


Fig. 3. Left to right (a) Input image 1, size: 512x512 pixels (b) Input image 2, size: 512x512 (c) Result obtained using the proposed method. Mosaic image: 512x763 pixels, overlap region: 50%.

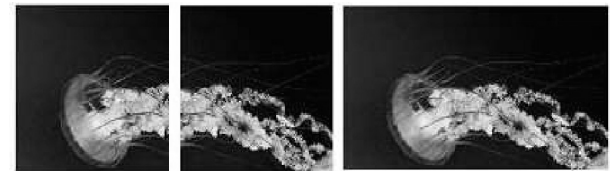


Fig. 4. Left to right (a) Input image 1, size: 512x512 pixels (b) Input image 2, size: 512x512 (c) Result obtained using the proposed method. Mosaic image: 512x884 pixels, overlap region: 5%.

percentage of overlapping region is 50%. The two images are stitched at this maximum value.

Fig. 4a is one of the input images and Fig. 4b is another image to be mosaicked. The two images have the same size of 512x512 pixels which is kept same in order to check the time complexity involved during the computations. Fig. 4c is the mosaicked colour image having size 512x884. The maximum normalized cross correlation value in this case is 0.9998 and percentage of overlapping region is 25%. The two images are stitched at this maximum value.

The technique is tested with three different overlapping regions having same size of 1024x1024 pixels. Different observations were noted, the time taken to stitch the two images with same size and different amount of

overlapping as 80%, 50% and 25% was 47.23 seconds. Fig. 5a is one of the input images and Fig. 5b is another image to be mosaicked. The two images have the same size of 1024×1024 pixels which is kept same in order to check the time complexity involved during the computations. Fig. 5c is the mosaicked colour image having size 1024×1152 . The maximum normalized cross correlation

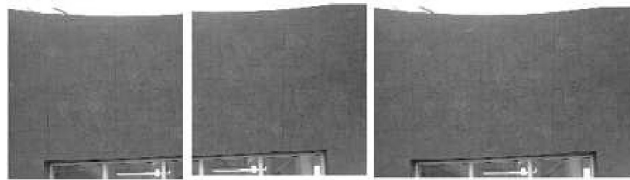


Fig. 5. Left to right (a) Input image 1, size: 1024×1024 pixels (b) Input image 2, size: 1024×1024 (c) Result obtained using the proposed method. Mosaic image: 1024×1154 pixels, overlap region: 50%.

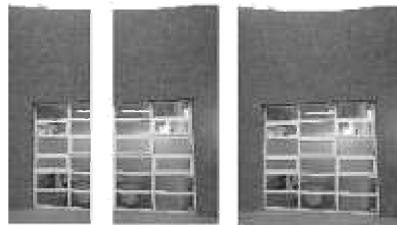


Fig. 6. Left to right (a) Input image 1, random size: 2048×768 pixels (b) Input image 2, random size: 2048×978 (c) Result obtained using the proposed method. Mosaic image: 2048×1536 pixels, overlap region: 50%.

value in this case is 0.9999 and percentage of overlapping region is 50%. The two images are stitched at this maximum value. The result for the remaining two overlapping regions of 25% and 80% required same amount of time.

The technique is tested with three different overlapping regions having same random size pixels. Fig. 6a is one of the input images and Fig. 6b is another image to be mosaicked. Fig. 6c is the mosaicked colour image having size 2048×1536 . The maximum normalized cross correlation value in this case is 0.9999 and percentage of overlapping region is 50%. The two images are stitched at this maximum value.

CONCLUSION

A method of seamless stitching of images by removing the overlapping region is presented. The method is based on finding the normalized cross correlation between the two images and finding its maximum value. The images are then stitched at this point by removing the overlapping region between them. The images stitched in such a way that the seam is not visible. Experimental results illustrate the method and show that it can lead to seamless mosaic images by eliminating the common region between the two.

REFERENCES

- [1] A., Ghoshtasby, 2 -D and 3-D image registration: for medical, remote sensing, and industrial applications, John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.
- [2] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys (CSUR)*, vol.24, no. 4, pp.325-376, Dec.1992.
- [3] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural, similarity," *IEEE Trans. on Image Processing*, vol.13, no.4, pp600-612, Apr. 2004.
- [4] J. Boutellier, O. Silvén, M. Tico, and L. Korhonen, "Objective evaluation of image mosaics," book chapter in *Computer Vision and Computer Graphics. Theory and Applications*, J. Braz et al. (Eds.), Springer Berlin, Heidelberg, pp. 107-117, 2009.
- [5] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, pp. 22-30, Mar. 1996.
- [6] R. Szeliski, "Image alignment and stitching: a tutorial," Technical Report MSR-TR-2004-92, Microsoft Research, Dec. 2004. Prausnitz, J., Lichtenhaler, R. and de Azevedo, E., *Molecular thermodynamics for fluid-phase equilibrium*, Prentice-Hall, Inc., 1986.
- [7] P. J. Burt, and E.H. Adelson, "A multi resolutions pyramid with application to image mosaics," *ACM Trans. Graphics*, vol.2, pp.217-236, Oct.1983.
- [8] C. T. Hsu, and J. L. Wu, "Multi resolution mosaic," *IEEE Trans. on Consumer Electronics*, vol. 42, pp. 981-990, Aug. 1996.
- [9] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *Proceedings of SIGGRAPH*, pp. 313-318, Jul. 2003.
- [10] A. Agrawal, R. Raskar, S. Nayar, and Y. Li, "Removing Photography Artifacts using Gradient Projection and Flash- Exposure Sampling," *ACM Trans. Graphics* vol. 24, no. 3, pp. 828-835, Jul. 2005.
- [11] G. D. Finlayson, S. D. Hordley, and M. S. Drew, "Removing shadows from images," *Proc. European Conf. Computer Vision*, pp. 823-836, 2002.
- [12] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *Proceedings of SIGGRAPH*, pp. 249-256, 2002.

-
- [13] J. Tumblin, A. Agarwal, and A. Raskar, "Why I want a gradient camera," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 103–110, 2005.
- [14] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," Eighth European Conference on Computer Vision, vol. 4, pp. 377–389, May 2004.
- [15] M.-S. Su, W.-L. Hwang, and K.-Y. Cheng, "Analysis on multi resolution mosaic images," IEEE Trans. on Image Processing, vol. 13, no. 7, pp. 952–959, Jul. 2004.
- [16] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, and B. Curless, "Interactive Digital Photomontage," ACM Trans. Graphics, vol. 23, no. 3, pp. 294–302, Aug. 2004.
- [17] A. Agrawal, R. Raskar, and R. Chellappa, "What is the range of surface reconstructions from gradient field?," in A. Leonardis, H. Bischof, and P. Pinz, editors, Computer Vision and ECCV 2006, vol. 3951 of Lecture Notes in Computer Science, pp. 578–591, Springer Berlin, Heidelberg, 2006.
- [18] Ioana S. Sevcenco, Peter J. Hampton, and Pan Agathoklis, "Seamless Stitching Of Images Based On A Haar Wavelet Based 2D Integration Method," 978-1-4577-0274-1/11/1c2011 IEEE.

MULTI-LEVEL DECISION SYSTEM FOR THE DATA COMPRESSION OF THE X-RAY C-ARM MEDICAL IMAGES

Shruthi Santhosh

Dept. of Electronics and Telecommunication,
Vidyalankar Institute of Technology
Mumbai, India
shrurao21@gmail.com

Dr. Kasi Viswanathan Agilandam

Director, Imaging,
Skanray Technologies Pvt Ltd,
Mysore, India
kasi@skanray.com

ABSTRACT

Images are stored on the computer as a collection of bits representing pixels, forming the picture elements. Medical Imaging is one of the key ingredients in the medical domain for clinical research, diagnosis, and radiographic investigation of diseases. These X-Ray images need to be captured, compressed, stored and transferred to the required destination effectively, for proper diagnosis and various other clinical operations by the radiologists. As the number of medical image data grows, the need for storing, communicating and manipulating large amounts of digital data increases and hence a good compression and a storage mechanism is required, without degrading the quality of the data.

This paper discusses the different parameters and the preprocessing operations which play a very important role in analyzing the medical image size before and after compression, and hence a multi-level decision system is designed which helps in obtaining a good compression based on the exhaustive sets of experimental results obtained. The various parameters considered in this paper for our analysis are Signal to Noise ratio values for a set of test and reference images, Correlation coefficient obtained for a sequence of source and target images, and the preprocessing operations done are Creating a group or sets of images for a number of averaging factors, Subtracting the images in a sequential order and then compressing these resultant images to observe if these operations will improve the compression results compared to the compression results got for the unprocessed original images. JPEG2000 lossless compression is chosen to avoid any loss of data. The Correlation and SNR calculations have been done using the ImageJ software.

Keywords-*Medical X-Ray Images; Image Pixel depth; JPEG2000 Lossless Compression; Correlation; Signal to Noise Ratio; Averaging; ImageJ; Water-shed algorithm; Subtraction and compression; Compression ratio; Image size; DICOM; PACS*

I. INTRODUCTION

Image compression is the fundamental factor to the efficient and cost-effective use of digital medical imaging technology and applications. Medical services today rely heavily on imaging technology, including X-ray computed tomography, magnetic resonance imaging, ultrasound and nuclear medicine examinations. As these examinations become more common, there will be a steep rise in these medical equipments being used in the

hospitals, and hence the volume of the acquired data increases to the point of becoming extremely excessive.

Studies have shown that the radiology department of a large hospital can produce more than 20 terabits of image data per year. Overall, the amount of digital radiologic data, that is generated every year in the USA alone is of the order of petabytes and is increasing rapidly. For example, a CT (Computed Tomography) exam produces an average of 12 Mbytes of imaging data which is 30 images, 512x512 pixels, 12 bits/pixel. A digitized chest X-ray, which is 4096x4096 pixels, 12 bits/pixel, requires an average of 24 Mbytes of storage space per image. This is the result of high image resolution used in radiology, and the large number of images required for each examination. This stretches the capabilities of digital storage systems, and imposes exceedingly high requirements on the bandwidth of communication networks. To represent such large medical images with the smallest possible number of bits, the digital data compression is essential and plays a very important role in minimizing the need for large storage space and helps in speeding the transmission across low bandwidth channels. Digital Image Compression can address these problems by reducing the data storage and transmission requirements. Several reasons which can be pointed to stress the importance of compression are [1]:

- Digital medical image databases are normally very large repositories.
- Patient data must be stored for long periods resulting in a continuous growth of medical imaging databases.
- High resolution and many bits per pixel result in large volumes of data.
- The image transmission time is volume driven. The usefulness of a PACS (Picture Archiving and Communication Systems) depends on appropriate transfer waiting times.

The two types of compression are lossless and lossy. In lossless compression, an original image is perfectly recoverable from the compressed format. However, its gains are limited, ranging from a typical 2:1 compression to an optimistic 4:1. Serious compression of 10:1 or more is lossy compression in which only an approximate image can be obtained [5].

But, this loss is not desired, as data loss could lead to a wrong diagnosis during patient analysis, due to lack of clear data. Also, the compression procedure has to satisfy the medical imaging standards. Hence,

JPEG2000 lossless compression has been opted for this work.

This paper discusses the various parameters and preprocessing operations which play a very important role in analyzing the medical image size before and after compression and hence a multi-level decision system is designed which decides if we would obtain a good lossless compression based on these results. The parameters considered for our analysis are Signal to Noise Ratio and Correlation coefficient, whereas the preprocessing operations done are preparing the images in the required format, averaging the images for varying averaging factors, subtracting the images in a sequential order, and later compressing these preprocessed images to compare the results with the original images which did not undergo any preprocessing operations. Plenty of images of varying types, resolution, pixel depth and size were considered for our analysis and research was done to understand the way compression ratio would vary based on the preprocessing operations and parameter calculations considered for this work. All these operations are discussed in depth further in the paper. Thus, this work helps in saving an enormous amount of storage space for more medical image examinations which is a boon to the medical imaging world.

This paper is divided into the following sections. Section II describes the basic definitions and parameters required for the understanding of this system. Section III involves the working of this system. Section IV comprises the images used for our experiment and the results obtained. Section V is the conclusion of the experiment and the observations made. Section VI is acknowledgements and Section VII gives the list of reference papers which was referred for this paper work.

II. BASIC DEFINITIONS AND PARAMETERS

This section describes the different parameters and the preprocessing operations required for the analysis of this work and also a few definitions with their respective formulae which are used for the numerical calculations.

A. DICOM Image

DICOM or DCM stands for Digital Imaging and Communications in Medicine and is a standard for handling, storing, printing, and transmitting information in the medical imaging world. DCM images differ from the other grayscale images, as each image will have a header which has a group of tags containing the complete information about the patient and the image. This means that the X-ray image actually contains the patient details within the image as part of the header, so that the image can never be separated from this information by mistake. An example for a DCM image is as shown in Fig. 1 and the Table 1 shows the entire header of this image.



Figure 1. Example of a DICOM Image

Table 1. Header of the DICOM image which contains patient details

```

0002,0002 Media Storage SOP Class UID:
1.2.840.10008.5.1.4.1.1.12.1
0002,0010 Transfer Syntax UID: 1.2.840.10008.1.2.1
0002,0012 Implementation Class UID: 1.3.6.1.4.1.5962.99.2
0008,0008 Image Type: DERIVED
0008,0016 SOP Class UID: 1.2.840.10008.5.1.4.1.1.12.1
0008,0020 Study Date: 2012/12/05
0008,0030 Study Time: 14:44:54
0008,0060 Modality: XA
0008,0068 Presentation Intent Type: FOR PRESENTATION
0008,0070 Manufacturer: SKANRAY TECHNOLOGIES
0008,0090 Referring Physician's Name: DR. SANTHOSH KUMAR
0008,1030 Study Description: Orthopaedic - Short Bone - Tissue
0008,103E Series Description: FLUORO
0010,0010 Patient's Name: MRS. SHRUTHI
0010,0020 Patient ID: 1
0010,0030 Patient's Birth Date: 1988/01/21
0010,0040 Patient's Sex: F
0018,0040 Cine Rate: 12.5
0018,0060 kVp: 51.25
0018,1063 Frame Time: 80
0018,1066 Frame Delay: 0
0018,1150 Exposure Time: 1000
0018,1151 X-ray Tube Current: 0.75
0018,1155 Radiation Setting: SC
0018,1164 Imager Pixel Spacing: 0.22\0.22
0018,1500 Positioner Motion: STATIC
0018,1510 Positioner Primary Angle: 0.01
0018,1511 Positioner Secondary Angle: 0.21
0018,1600 Shutter Shape: RECTANGULAR
0018,1604 Shutter Right Vertical Edge: 1024
0018,1608 Shutter Lower Horizontal Edge: 1024
0018,7004 Detector Type: DIRECT
0020,0010 Study ID: SKAN_C
0020,0011 Series Number: 1
0020,0013 Image Number: 1
0020,0020 Patient Orientation: LAF
0020,4000 Image Comments: NO COMMENTS
0028,0002 Samples per Pixel: 1
0028,0004 Photometric Interpretation: MONOCHROME2
0028,0008 Number of Frames: 19
0028,0009 Frame Increment Pointer: 1577059
0028,0010 Rows: 1024
0028,0011 Columns: 1024
0028,0030 Pixel Spacing: 0.22\0.22
0028,0100 Bits Allocated: 16
0028,0101 Bits Stored: 16
0028,0102 High Bit: 15
0028,0103 Pixel Representation: 0
0028,0106 Smallest Image Pixel Value: 768
0028,0107 Largest Image Pixel Value: 11277
0028,1040 Pixel Intensity Relationship: LIN
0028,1050 Window Center: 6022
0028,1051 Window Width: 10509
0028,2110 Lossy Image Compression: 00
0040,0244 Performed Procedure Step Start Date: 20121205
0040,0245 Performed Procedure Step Start Time: 000000.00
0040,0253 Performed Procedure Step ID: 1
0040,0254 Performed Procedure Step Description: X-RAYS
7FE0,0010 Pixel Data: 1400

Title: 1.2.840.10008.2.11.3670118518064.12002012112469.DCM
Width: 225.28 mm (1024)
Height: 225.28 mm (1024)
Depth: 19.00 mm (19)
Resolution: 4.545 pixels per mm
Coordinate origin: 0,0,0
Bits per pixel: 16 (unsigned)
Display range: 767 - 11276
Image: 1/19
Magnification: 0.50
Size: 2MB/frame

```

Fluorography provides real time X-Ray images. All the DCM images used for our analysis in this paper are either Fluoro or X-Ray grayscale images of various modalities like XA (X-Ray Angiography), CT (Computed Tomography), MR (Magnetic Resonance) [4]. When we use Fluoroscopy images, each image is separated into individual frames, as compression was done for single image or groups of single images [2]. It does not support multi-frame image compression. To solve this problem, and support compression for any type of DCM image, the Visual Basic coding was done to break the continuous fluoro frames to individual frames without any degradation in the quality of the image or any loss of data, to make the processing easier. The steps are shown in Fig. 2.

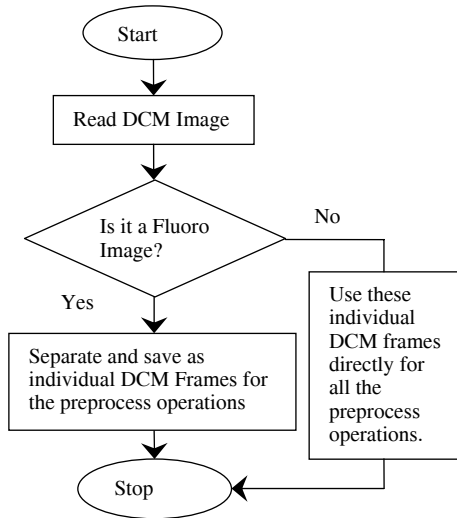


Figure 2. Steps to separate the DCM image into individual frames

The image shown in Fig.1 is a Fluoro image of a Hand made up of 19 frames, (1024x1024) pixels, 16-bit pixel depth, and size 2MB/frame. The complete information of this image is given in the table 1, from which plenty of details about the patient, the medical procedure done, type of modality, physical properties of the image, C-Arm system details, and much more information is obtained.

B. X-Ray and C-Arm Machine

X-ray is a discrete packet of electromagnetic wavelength 10^{-8} m to 10^{-11} m and it is a radiation containing a defined amount of energy. X-ray is used in radiography to view a non-uniformly composed material such as the human body. Images with good amount of precise information are obtained, using a few physical properties of X-Rays which are:

- Invisibility; they cannot be perceived by the senses.
- They travel in straight lines and at the speed of light.
- Diffracted by a crystalline grid.
- They can pass through matter and are partly absorbed during transmission.
- Fluoroscopy provides real time X-ray images.

The C-Arm is a specialized X-Ray imaging system which can scan the entire human body and produce very high resolution and precise X-Ray and Fluoro Multi-Frame images. It is so called, due to the special arc semi-circular design and is used to improve the outcomes during pre, intra and post operative operations. The directions of movement of this device are vertical, horizontal, swivel and rotation [3].

In the C-Arm system, the rays are produced in an X-ray tube when high-energy electrons interact with target atoms. The variable fraction of its energy is converted into an X-ray photon, and the remainder is dissipated as heat. The patient is illuminated with a beam of X-rays and different organs in the patient's body absorb different fractions of the incident radiation, modulating the beam intensity. This modulated beam that exits from the patient is detected by an image receptor, and the image is obtained and is viewed by the radiologist. Further image processing is done, to store the data.

C. Correlation Coefficient Calculation

Correlation coefficient, 'R' is used to exploit the similarities in the image pixels for two images considered, and is given by equation (5).

$$(1) \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n} ; \quad \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

$$(2) \quad Var(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$(3) \quad Var(Y) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}$$

$$(4) \quad Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

$$(5) \quad R_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Means of variables 'X' and 'Y' is given using (1) as ' $\bar{X}$ ' and ' $\bar{Y}$ ' respectively. From (2) and (3), Variance is obtained. Covariance of 'X' and 'Y' is calculated using (4). Consider an image with total number of frames 'n'. The correlation coefficient is calculated for the sequence of frames 'F(n)' in the increasing order, that is between F(1) and F(2), F(2) and F(3), F(3) and F(4), F(4) and F(5),.....,F(n-1) and F(n). The results obtained were named as R(1), R(2), R(3), R(4),...,R(n-1) respectively.

The correlation value ranges from +1 to -1. If the value is almost equal to 1, then the images are said to be well correlated, and this value states that there is no much difference between the pixels in the two images being compared. In this case, the set of images can be reduced to a single image, instead of storing all the similar images, thereby decreasing the size of the entire image.

But, if the images are nowhere similar to each other, then they are badly correlated and the correlation coefficient value decreases. In this case, all the frames cannot be reduced to a single image, as there is a possibility of loss of data, even if a single frame is lost. Thus, correlation coefficient plays a very important role. In this paper the correlation coefficient has been calculated and tabulated, using the ImageJ software. For all the calculations, the statistic that was correlated was the average, with the pixel local region size set as '1', and decimal value has been considered for 8 decimal places.

D. Signal to Noise Ratio Calculation

It is defined as the ratio of signal power to the noise power. A ratio higher than 1:1 indicates more signal than noise. In our work here, we have analyzed that the Signal to Noise ratio or SNR plays a very important role in deciding if we would obtain a good compression. All the SNR calculations were done using the ImageJ software, using (6) as shown below.

$$SNR = 10 \cdot \log_{10} \left[\frac{\sum_{x=0}^{n_x-1} \sum_{y=0}^{n_y-1} [r(x,y)]^2}{\sum_{x=0}^{n_x-1} \sum_{y=0}^{n_y-1} [r(x,y) - t(x,y)]^2} \right] \quad (6)$$

In (6) the parameter $[r(x,y)]$ is the reference image and $[t(x,y)]$ is the test image. The images considered for the calculations had the same size denoted as $[n_x, n_y]$. For an image with total number of frames ' n '. These ' n ' frames are averaged with an averaging factor ' $j=n$ ' using (7) to obtain a single image. This resultant image is chosen as the reference image 'O'. The SNR is then calculated for the sequence of frames 'F(n)' in the increasing order, that is SNR is calculated between 'O' and I(1), I(2), I(3), I(4),.....,I(n). Results are tabulated for all example images.

Here, SNR values are obtained using the water-shed algorithm. This method, segments an image by directly flooding the gray level images. The grayscale watershed segmentation is useful to segment particles in contact when the model of shape is unknown a priori. The changes in the results of compression depending on the SNR, for different cases of preprocessed images have been described in the later sections.

E. Averaging Of Images

To check if calculating the average for a set of images for different averaging factors would improve the compression results, sets of images were averaged and the resultant images were compared with the unprocessed original images. However, for practical medical applications averaging the medical images is not allowed. In this work, to understand if smoothing the sets of images would bring improve the compression results, this procedure was executed. Results have been discussed in detail using examples later in this paper.

Thus, for an image with ' n ' frames, the average of images 'A' for varying averaging factors ' j ' was calculated using (7), where ' j ' = prime factors of ' n '. These resultant images were tabulated and used for comparison and analysis of results.

$$A(j) = \left[\frac{A1 + A2 + \dots + A(n)}{j} \right] \quad (7)$$

F. Subtraction of Images

In this project, further analysis was done by subtracting a sequence of frames, and later compressing them, to observe the changes if any, in the compression results obtained due to this operation. Coding was done using Visual basic software for this preprocessing operation.

For example, if two perfectly identical images are subtracted, the resultant image obtained would be an image with pixel value as zero, and if there is very less noise, then it will give a very good compression. Thus, storage can be reduced by storing only one of the images, instead of two. Similarly for a large sequence of frames, a set of similar frames can be sequentially subtracted and only the resultant images can be stored, thus drastically reducing the count of frames being stored and in turn increasing the storage space. In order to retrieve the original image, the single reference image has to be added to the subtracted image. This procedure was extended to all the DCM images considered for our experiment and the resultant images were compared with results

The subtraction was calculated for an image with ' n ' frames as F(1)-F(2), F(2)-F(3),.....,F(n-1)-F(n) and the resultant images obtained were named S12, S23,.....,S(n-1)(n) respectively. These subtracted images obtained were compressed and changes in file size obtained were observed, and the cause for this change was analyzed.

G. JPEG2000 Lossless Compression

JPEG or Joint Photographic Experts Group is a compression standard that is used universally for digital photographic images. In addition to the benefits of scalability, JPEG 2000 delivers about 20-40% better compression than JPEG. At more extreme compression ratios, JPEG 2000 delivers significantly better quality compression results. A few advantages of JPEG2000 lossless compression over the other compression techniques are as follows:

- High compression efficiency
- Progressive by resolution and quality
- ROI coding (static and dynamic)
- Error resilience capabilities
- Object oriented functionalities (coding, information, embedding)
- Superior low bit-rate performance
- Accuracy and resolution, robustness to bit-errors.
- Lossless transformations

JPEG2000 lossless compression is used here, in order to make sure that there is no loss of data, as in medical imaging it is very important to make sure that, there is no ambiguity in the image data as it may lead to a wrong clinical diagnosis [7].

The generalized block diagram for the JPEG2000 lossless compression code is as shown below in Fig. 3 and the flow of operations are given in the steps below:

Tiling: The image is split into different smaller components, and these components are decomposed into rectangular tiles.

Forward Transform: The wavelet transform is applied to each tile, decomposing it into different resolution levels, which are made up of sub-bands of coefficients that describe the frequency characteristics of local areas of the tile-component, than across the entire tile-component.

Quantization: The coefficients are scalar-quantized to reduce the number of bits to represent them.

Entropy encoding: Bit-planes of the coefficients of a "code-block" are entropy coded.

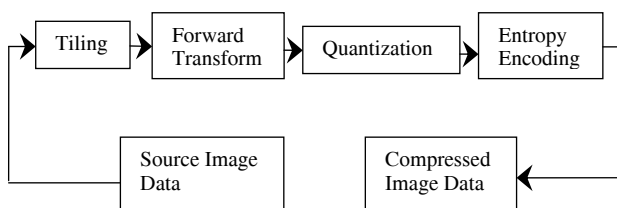


Figure 3. JPEG2000 Lossless Compression Code

JPEG2000 carries out a codification that is independent from the code-blocks; which is an advantage in error resilience, as each error in the bit-stream that corresponds to a code-block will remain inside it. This helps wireless communications as they are subject to random channel errors, while Internet communications are subject to a loss of data due to the congestion in traffic. The quality of a compressed image is evaluated by analyzing the difference between the original and the compressed one.

III. MULTI-LEVEL DECISION SYSTEM

The proposed multi-level decision system has been worked upon as described briefly in this section. DICOM medical images of various modalities are chosen for this work, to analyze and understand the changes in the compression results, caused by the parameter calculations and preprocessing operations. If the image is a fluoro image with 'n' frames, then it is separated into 'n' individual frames and then processed.

Using the ImageJ software, Correlation values 'R' and the SNR values will be calculated for the sequence of 'n' frames, in the sequential order F(1) and F(2), F(2) and F(3), F(3) and F(4),.....,F(n-1)F(n). These values will be tabulated for further investigations. Correlation values will indicate the amount of similarity between the target and source images, with the values ranging from +1 to -1. SNR would indicate the signal to noise ratio in decibels.

Now, using the Visual Basic coding, these individual frames shall be grouped into different sets, with varying averaging factors 'j'=Prime factors of 'n'. These will also be correlated and SNR values shall be calculated for each set.

Subtraction is done to these 'n' frames in a sequence, and the resultant subtracted images 'S' are stored. This process is done for all sets of averaged images. The coding for this process is done using Visual Basic software.

Finally, all the original and preprocessed images are compressed using JPEG2000 lossless compression and their initial and final sizes are tabulated. The flow of operation of this system is given below, in Figure 4. Once, these sets of results is obtained for a variety of medical images, a further analysis will be done and a conclusion will be obtained on the parameter which acts as a deciding factor for the compression results.

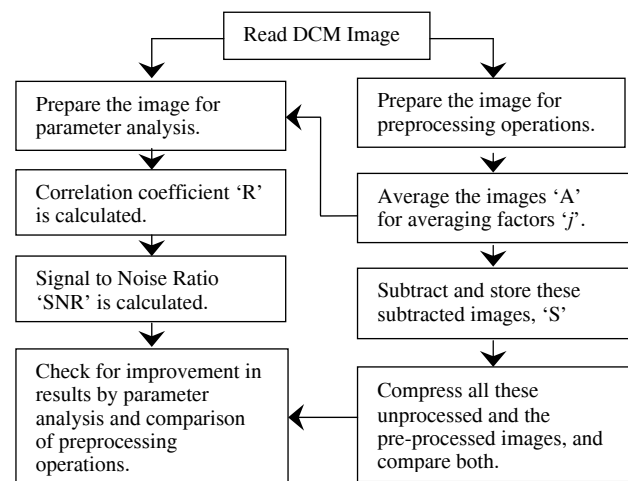


Figure 4. Multi-Level Decision System

IV. EXPERIMENTAL RESULTS

Various medical images were processed using our system, and the experimental results were analyzed. Two examples are:

Fig. 5(a), (b), (c) are a few frames of a CT image of size of 514KB/frame with 9 frames, resolution (512x512) pixels, and pixel depth 12-bit. Fig 5(d) and Fig 5(e) shows the subtracted image. Tables 2 and 3 shows the tabulated results. Fig 5(f) depicts the averaged image with $j=3$.

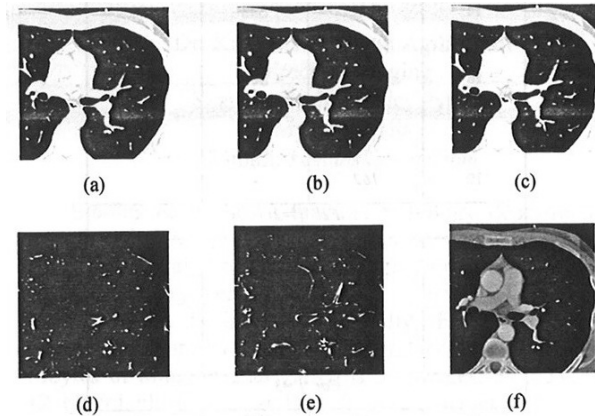


Figure 5 :

- (a). Original image frame 6 (b). Original image frame 7
(c). Original image frame 8 (d). Frame 6 and 7 Subtraction
(e). Subtraction between frames 7 and 8
(f). Averaged image with $j=3$

Table 2. Calculation of Correlation coefficient 'R' & 'SNR' for CT Image.

Target Frame	Source Frame	'R' for $j'=1$	Reference Image	Test Image	SNR (dB) for $j'=1$
1	2	0.98030174	Org	1	13.0802537
2	3	0.99366409	Org	2	15.867306
3	4	0.99355805	Org	3	17.88400101
4	5	0.99355638	Org	4	19.92258849
5	6	0.99366122	Org	5	20.89272991
6	7	0.9935993	Org	6	19.88172661
7	8	0.9933067	Org	7	17.96041487
8	9	0.99384594	Org	8	15.86479297
			Org	9	14.12833955
Target Frame	Source Frame	'R' for $j'=2$	Reference Image	Test Image	SNR (dB) for $j'=2$
1	2	0.97654426	Org	1	14.86705968
2	3	0.98056811	Org	2	19.35635209
3	4	0.98073792	Org	3	21.20669016
			Org	4	17.15873264
Target Frame	Source Frame	'R' for $j'=3$	Reference Image	Test Image	SNR (dB) for $j'=3$
1	2	0.96498299	Org	1	16.30104604
2	3	0.96694785	Org	2	22.50398575
			Org	3	16.3963476
Target Frame	Source Frame	'R' for $j'=4$	Reference Image	Test Image	SNR (dB) for $j'=4$
1	2	0.9533267	Org	1	17.9017028
			Org	2	20.72136421

Table 3. Compression results for the original and pre-processed images.

For $j'=1$			
Original Frame, 'I'	Size of the Compressed Original Frame (KB)	Subtracted Frame, 'S'	Size of the Compressed Subtracted Frame (KB)
I1	163	S12	167
I2	164	S23	149
I3	163	S34	147
I4	163	S45	150
I5	164	S56	148
I6	162	S67	147
I7	163	S78	150
I8	164	S89	149
I9	162	-	-
For $j'=2$			
I1	156	S12	160
I2	160	S23	159
I3	160	S34	160
I4	160	-	-
For $j'=3$			
I1	153	S12	160
I2	156	S23	161
I3	156	-	-
For $j'=4$			
I1	151	S12	161
I2	153	-	-

Fig. 6(a) is a frame of a Joint image of size of 2MB/frame with 5 frames, resolution (1024x1024) pixels, and pixel depth 16-bit. Fig 6(b) and Fig 6(c) shows the subtracted image and averaged image respectively. Tables 4 and 5 shows the tabulated results.

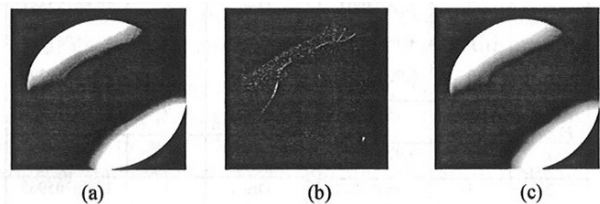


Table 4. Calculation of Correlation coefficient 'R' & 'SNR' for CT Image.

Target Frame	Source Frame	'R' for $j'=1$	Reference Image	Test Image	SNR (dB) for $j'=1$
1	2	1.00086999	Org	1	19.40814744
2	3	0.9924196	Org	2	20.03558891
3	4	0.9914943	Org	3	24.86884097
4	5	0.99731314	Org	4	27.65182994
			Org	5	22.76453324
Target Frame	Source Frame	'R' for $j'=5$	Reference Image	Test Image	SNR (dB) for $j'=5$
1	2	0.9915964	Org	1	29.73722187
			Org	2	26.16655404

Table 5. Calculation of Correlation coefficient 'R' & 'SNR' for CT Image.

<i>For 'j'=1</i>			
<i>Original Frame, 'I'</i>	<i>Size of the Compressed Original Frame (KB)</i>	<i>Subtracted Frame, 'S'</i>	<i>Size of the Compressed Subtracted Frame (KB)</i>
I1	990	S12	1030
I2	1000	S23	1040
I3	990	S34	1030
I4	980	S45	1030
I5	990	-	-
<i>For 'j'=5</i>			
I1	940	S12	943
I2	933	-	-

V. ANALYSIS AND CONCLUSION

We understand from the results that SNR is a very important parameter that decides the compression factor. In the CT image, the images are very well correlated, and SNR is good. Hence, the subtracted images give better compression. In the Joint image, even though the images are well correlated, SNR is bad, and hence compared to the original images the subtracted images gives a bad compression. Also, averaging and subtraction operations help in improving the compression ratio. Similar results were obtained for many other images.

ACKNOWLEDGEMENTS

The medical images used in this paper were provided by Skanray Technologies which provided the platform for us to understand the various Medical Imaging techniques.

REFERENCES

- [1] A.W. Wong, R.K. Taira, and H.K. Huang, "Implementation of a digital archive system for a radiology department", in Proc. SPIE Conf. on Medical Imaging VI: PACS Design and Evaluation 1645, pp. 182-190, 1992.
- [2] S. Wong, L. Zarembo, D. Gooden, and H.K. Huang, "Radiologic Image Compression - A Review", Proc. IEEE, vol. 83, no. 2, pp. 194-218, Feb. 1995.
- [3] W.J. Chimiak, "The digital radiology environment", IEEE Journal on Selected Areas in Commun., vol. 10, no. 7, pp. 1133-1144, Sep. 1992.
- [4] J.R. Cox, E. Muka, J.G. Blaine, S.M. Moore, and G.R. Jost, "Considerations in Moving Electronic Radiography into Routine Use", IEEE Journal on Selected Areas in Commun., vol. 10, no. 7, pp. 1108-1120, Sep. 1992.
- [5] Komaz Karadimitriou, John M. Tyler, "Min-Max Compression methods for Medical image databases", IEEE Journal, 2000
- [6] B. R. Sanders, J. H. Shannon, "Telemedicine: Theory and Practice", Springfield, Illinois, 1997.
- [7] A. N. Skodras, C. A. Christopoulos and T. Ebrahimi, "JPEG2000: The Upcoming Still Image Compression Standard"

A NEW METHOD FOR OBTAINING AN OPTIMAL SOLUTION FOR TRANSPORTATION PROBLEMS

Pritam Warke

Assistant Professor, MCA Department
Navinchandra Mehta Institute of Technology and
Development, Dadar West
pritamwarke123@yahoo.com

Ankita Kore

Research Scholar, MCA Department
Navinchandra Mehta Institute of Technology and
Development, Dadar West
anki.kore32@gmail.com

ABSTRACT

In this paper a new method is proposed for finding an optimal solution for a wide range of transportation problem. This method is easy to understand and use. The main feature of this method is that it requires very simple arithmetical and logical calculation. This method is very efficient for those decision makers who are dealing with logistics and supply chain related issues. This method can easily adopt among the existing methods.

Keywords— *Transportation problem, Optimal solution, Cost Matrix.*

I. INTRODUCTION

Transportation problem is a special variety of classical linear-programming problem. The objective of the transportation problem to provide the following information to the decision makers: What quality should be transported from a manufacturing unit to all possible destinations? And What would be the cost for this allocation?

Transportation problem is a particular class of Linear Programming, which is associated with day-to-day activities in our real life and mainly deals with logistics. It helps in solving problems on distribution and transportation of resources from one place to another. Transportation problems deal with the transportation of a single product manufactured at different plants (*supply origins*) to a number of different warehouses (*demand destinations*). The objective is to satisfy the demand at destinations from the supply constraints at the minimum transportation cost possible.

The objective of the transportation model is to determine the amount to be shipped from each source to each destination so as to maintain the supply and demand requirements at the lowest transportation cost.

II. PROPOSED METHOD

Step 1 :

Construct the transportation table from the given transportation problem.

Step 2 :

Subtract each row entries of the transportation table from the respective row minimum and then subtract each column entries of the transportation table from the respective column minimum.

Step 3 :

Now there will be at least one zero in each row and in each column in the reduced cost matrix. Select the first zero (row wise) occurring in the cost matrix. Count the total number of zeros excluding the selected one in the corresponding row and column. Repeat the procedure for all zeros in the matrix.

Step 4 :

Now choose a zero for which the number of zeros counted in step 3 is minimum and supply maximum possible amount to that cell.

If tie occurs for any cell then first check for the corresponding value if demand and if again tie occurs then check the corresponding values in the rows and column and select the minimum one.

Step 5 :

After performing step 4, delete the row or column for the further calculation.

Step 6 :

Check whether the resultant matrix possesses at least one zero in each column and in each row. If not repeat step 2, otherwise go to step 7.

Step 7 :

Repeat step 3 to step 6 until and unless all the demands are satisfied and all the supplies exhausted.

III. NUMERICAL EXAMPLE

Consider the following cost minimizing transportation problem with three origins and four destinations.

Step 1 :

Construct the transportation table from the given transportation problem :

	D1	D2	D3	D4	Supply
S1	13	20	25	8	8
S2	50	20	25	45	10
S3	25	6	55	10	11
Demand	4	7	6	12	29 (Total)

Step 2 :

Subtract each row entries of the transportation table from the respective row minimum and then subtract each column entries of the transportation table from the respective column minimum.

(i) Subtracting row minimum :

	D1	D2	D3	D4	Supply
S1	5	12	17	0	8
S2	30	0	5	25	10
S3	19	0	49	4	11
Demand	4	7	6	12	29 (Total)

(ii) Subtracting column minimum :

	D1	D2	D3	D4	Supply
S1	0	12	17	0	8
S2	30	0	0	25	10
S3	19	0	49	4	11
Demand	4	7	6	12	29 (Total)

Step 3 :

Now there will be at least one zero in each row and in each column in the reduced cost matrix. Select the first zero (row wise) occurring in the cost matrix. Count the total number of zeros excluding the selected one in the corresponding row and column. Repeat the procedure for all zeros in the matrix.

Position	No. of zeros
(1,1)	1
(1,4)	1
(2,2)	2
(2,3)	1
(3,2)	1

Step 4 :

Now choose a zero for which the number of zeros counted in step 3 is minimum and supply maximum possible amount to that cell.

If tie occurs for any cell then first check for the corresponding value if demand and if again tie occurs then check the corresponding values in the rows and column and select the minimum one.

The tie occurs at position (1,1) and (1,4) then check the values in the Demand.

Demand of (1,1) = 4

Demand of (1,4) = 12

Hence select the minimum demand....

The tie occurs at position (2,2) and (3,2) then check the values in the Demand :

Demand of (2,2) = 7

Demand of (3,2) = 7

Again tie occurs then check the corresponding values in the rows and column :

For (2,2) : - Row 1 : 12 and column 1 : 30 + 25 = 55 Total = 67

For (3,2) :- Row 1 : 12 and column 4 : 19 + 49 = 68 Total = 80

Select the minimum total.

Position	No. of zeros	Supply	Total (No. of zeros + supply)	
(1,1)	1	8	9	(1)
(1,4)	1	8	9	(2)
(2,2)	2	10	12	(4)
(2,3)	1	10	11	(3)
(3,2)	1	11	12	(5)

Step 5 :

After performing step 4, delete the row or column for the further calculation.

Position (1,1) :

	D1	D2	D3	D4	Supply
S1	0	12	17	0	8 - 4 = 4
S2	30	0	0	25	10
S3	19	0	49	4	11
Demand	4 - 4 = 0	7	6	12	29 (Total)

Demand of D1 is zero. Hence delete the column D1.

And assign 4 to the (1,1)

Next position is (1,4) :

	D2	D3	D4	Supply
S1	12	17	0	4 - 4 = 0
S2	0	0	25	10
S3	0	49	4	11
Demand	7	6	12 - 4 = 8	29 (Total)

Supply of S1 is zero. Hence delete the row S1. And assign 4 to the (1,4)

Next position is (2,3) :

	D2	D3	D4	Supply
S2	0	0	25	10 – 6=4
S3	0	49	4	11
Demand	7	6 – 6=0	8	29 (Total)

Demand of D3 is zero. Hence delete the column D3. And assign 6 to the (2,3)

Next position is (2,2) :

	D2	D4	Supply
S2	0	25	4 – 4=0
S3	0	4	11
Demand	7 – 4=3	8	29 (Total)

Supply of S2 is zero. Hence delete the row S2. And assign 4 to the (2,2)

Next position is (3,2) :

	D2	D4	Supply
S3	0	4	11 – 3=8
Demand	3 – 3=0	8	29 (Total)

Demand of D2 is zero. Hence delete the column D2. And assign 3 to the (3,2)

	D4	Supply
S3	4	8
Demand	8	29 (Total)

Step 6 :

Check whether the resultant matrix possesses at least one zero in each column and in each row. If not repeat step 2, otherwise go to step 7.

Delete the row minimum from the cell and then column minimum from that cell.

	D4	Supply
S3	0	8
Demand	8	29 (Total)

	D4	Supply
S3	0	8-8=0
Demand	8-8=0	29 (Total)

Step 7 : Repeat step 3 to step 6 until and unless all the demands are satisfied and all the supplies exhausted.

Solution :

	D1	D2	D3	D4	Supply
S1	4	20	25	4	8
S2	50	4	6	45	10
S3	25	3	55	8	11
Demand	4	7	6	12	29 (Total)

The total cost associated with this allocation is $(13*4)+(8*4)+(25*6)+(20*4)+(6*3)+(10*8) = \text{Rs.412}$

CONCLUSIONS

Thus it can be concluded that the proposed method provides an optimal solution in less number of iterations, directly for the given transportation problem. As this method required less number of time and is very easy to understand and apply. So it will be very helpful for decision makers who are dealing with logistic and supply chain problem.

REFERENCES

- [1] G.B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, N J, 1963.
- [2] Taha. H.A., Operations Research- Introduction, Prentice Hall of India (PVT), New Delhi, 2004.
- [3] Sharma. J.K., Operations Research-Theory and applications, Macmillan India (LTD), New Delhi, 2005.
- [4] V. J. Sudhakar, N. Arunsankar and T. Karpagam, A New approach for finding an Optimal Solution for Transportation Problems, *European Journal of Scientific Research*, vol 68, pp. 254-257, 2012
- [5] Chandrasekhar Putcha, Aditya K. Putcha, MD Rohul Amin Bhuiyan and Nasima Farzana Hoque, Development of New Optimal Method for Solution of Transportation Problems, Preceding of World Congress on Engineering 2010.
- [6] H. Arsham, Post optimality analyses of the transportation problem, *Journal of the Operational Research Society*, vol 43, pp. 121-139, 1992.
- [7] A.Henderson and R. Schlaifer Mathematical programming: Better information for better decision-making, *Harvard Business Review*, vol 32, pp. 73-100, 1954.
- [8] Koopmans T. C., Optimum Utilization of Transportation System, *Econometrica*, Supplement vol 17, 1949.

H.264 VIDEO ENCRYPTION ALGORITHM FOR HIGH-END MOBILE DEVICES

Ajay Kulkarni
Third Year,
Vidyalankar Institute
of Technology
Mumbai, India

Saurabh Kulkarni
Final Year
Vivekanand Education Society's
Institute of Technology
Mumbai, India

Ketki Haridas
Third Year,
Vidyalankar Institute
of Technology
Mumbai, India

ABSTRACT

Securing multimedia data has become of utmost importance especially in the new generation mobile phones and PDA's. With the rise in development in computer and internet technology, multimedia service has become the fastest growing technology. An innovative H.264 video encryption algorithm is proposed in this paper for systems having limited processing power, memory and bandwidth. To maintain a balance between security and computational time, the proposed algorithm will first shuffle the video frames along with the audio, and then AES is used to selectively encrypt the sensitive video codewords. Using this approach we can prevent unauthorized viewing of the video file and hence this algorithm provides a high level of security.

Keywords -- H.264, encryption, shuffle, video codewords, AES.

I. INTRODUCTION

With the fast growth of multimedia technology, wireless mobile systems, such as PAD and smart phones, they are no longer used just for calls and messaging, nowadays they are especially popular for transmission and storage of multimedia data, such as images and videos, which is the private data of users. To ensure their security, such sensitive data has to be protected either in transmission or storage. One possible way to protect multimedia information is to stop unauthorized access. But this approach cannot make sure that the multimedia information is physically secure. Another easy approach is to encrypt the complete bit stream with a cryptographic algorithm, such as DES or AES. However videos generally possess a large amount of data and require real-time operations. Moreover, in the case of the wireless mobile systems, there is limited processing power, memory and bandwidth, and is rarely able to handle the heavy encryption processing load. Therefore, taking into consideration the specific characteristics for resource-limited systems, new video encryption algorithms need to be developed.

For real-world applications, a video encryption algorithm has to take into account various parameters like security, computational efficiency, compression efficiency and so forth. Different types of video applications require different levels of security. For example, for Video on Demand, low security is will just be fine, whereas for military purposes or financial information, high level of security is required to completely prevent unauthorized access. Computational efficiency means that the encryption or decryption process should not cause too much time delay, so that the requirements of real-time applications are met. Video compression is employed to reduce the storage space and save bandwidth, so that the encryption process should have the least impact on the compression efficiency. All in all, a well-designed video encryption algorithm should provide sufficient security, high computational efficiency; impose little impact on the compression efficiency.

In this paper, we propose a video encryption algorithm in H.264 compressed domain for resource-limited systems. In order to optimize the tradeoff between security and computational complexity, first the frames are shuffled along with the audio and then only the codewords which are most critical or sensitive are encrypted.

The rest of this paper is organized as follows. Section II gives overview of H.264. Section III explains former encryption algorithms of H.264. Details of the proposed video encryption algorithm are described in Section IV. Section V explains AES key generation. The algorithm's performance, such as computational efficiency and memory requirement, is discussed in Section VI. Security issues are explained in Section VII and Section VIII concludes the paper.

II. OVERVIEW OF H.264

H.264 is state of the art video coding standard of ITU-T and ISO/IEC. It has enhanced compression. In H.264, an input video frame is divided into blocks of 16x16, called MB and each of them is encoded separately. Each video frame can be encoded as intra or inter. In intra frame,

the current MB is predicted spatially from MBs which have been previously encoded, decoded and reconstructed (MB at top and left). In inter mode, motion compensated prediction is done from the reference frames. The purpose of the reconstruction in the encoder is to ensure that both encoder and decoder use identical reference frames to create the prediction. If this is not the case, then the predictions in encoder and decoder will not be identical, leading to an increasing error or drift between the encoder and decoder.

The difference between original and predicted frame is called residual. This residual is coded using transform coding followed by quantization and zigzag scan. In the last step, entropy coding comes into action. First, quantized transform coefficients are run-length encoded. These runs and levels are then encoded using either CAVLC or context adaptive binary arithmetic coding (CABAC). On the decoding side, compressed bit stream is decoded by entropy decoding module, followed by inverse-zigzag scan. These coefficients are then rescaled and inverse transformed to get the residual signal which is added to the predicted signal to get the original signal back.

doing this, we achieve an audio stream that is impossible to decrypt and understand.

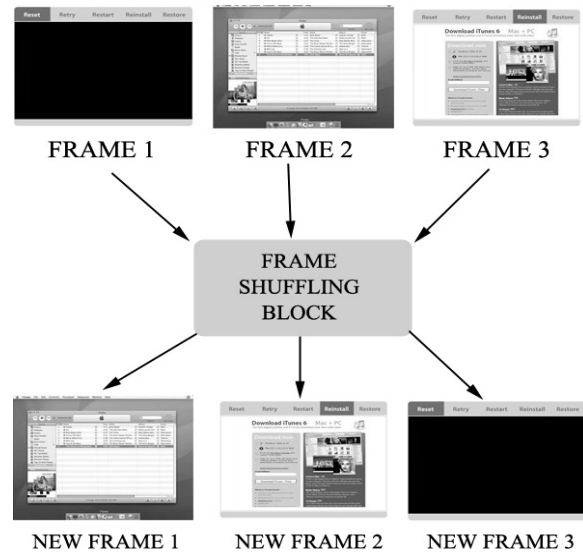


Fig 1: Frame shuffling block

III. FORMER ENCRYPTION ALGORITHMS OF H.264

In the past decades, a number of video encryption algorithms have been proposed. According to their association with video compression, these algorithms can be classified into joint compression and encryption algorithms and compression-independent encryption algorithms. For the former one, the encryption algorithms are embedded in a certain step of the video compression process. For example, some algorithms permute or scramble the DCT coefficients after the Discrete Cosine Transform (DCT). Differently, for compression-independent encryption algorithms, the compression and encryption process are carried out separately, which usually refers to the encryption of the compressed bit stream. It is also known as compressed domain encryption. In this, the odd indexed bytes in compressed bit stream are firstly encrypted with a conventional cryptographic algorithm, and then used as keys to XOR with the even indexed bytes.

IV. THE PROPOSED ENCRYPTION ALGORITHM

In the encryption block the video is first divided into frames by using a video cutter. These frames also contain the audio information. Here the individual frames are in video format instead of being an image because image files cannot store audio information. The shuffling block then shuffles these frames and then these frames are passed on to the frame stitching block. These frames that are in random position form the video. By

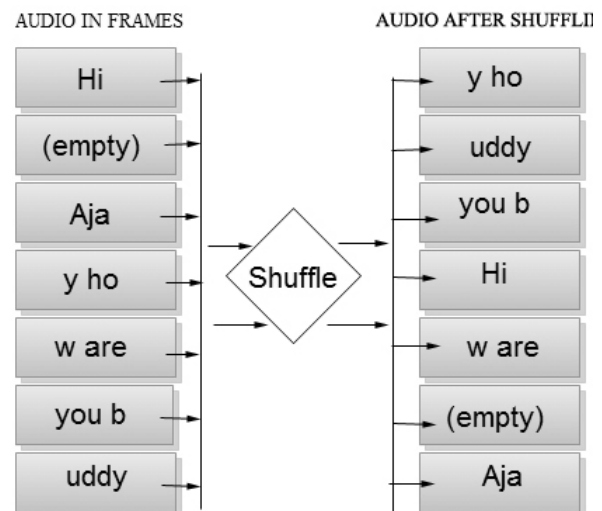
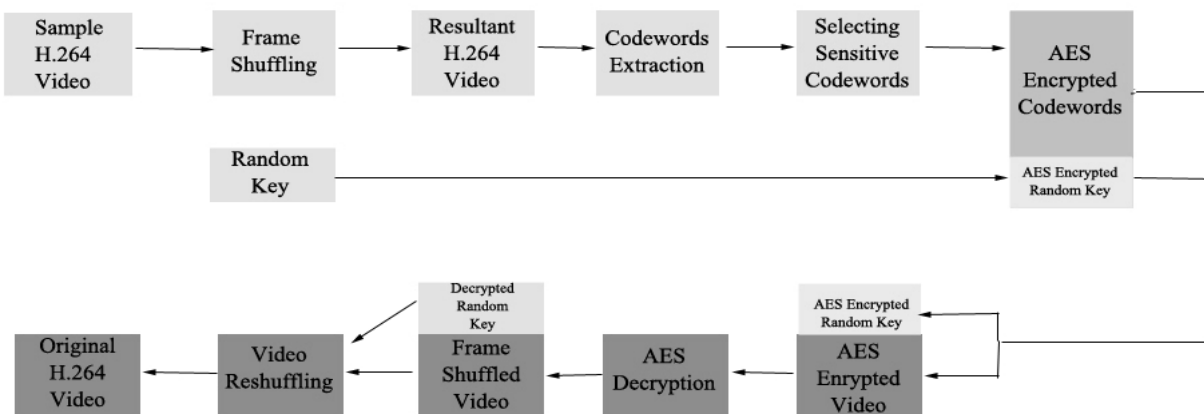


Fig 2: Audio after frame shuffling

Thus it is clear from the above diagrams that the audio is very difficult to decode unless the shuffling methodology is known. The shuffling algorithm uses a random key generation function implemented in java; this function is termed as the 'Shuffling Key' and is sent along with the video to the destination decryption block. The shuffling key is encrypted along with the video using AES. Hence by doing this we have ensured that the audio is impossible to understand and the video just shows random frames. But this is not enough, the frames maybe extracted from the video and maybe rearranged to form the actual video. Even though this

ENCRYPTION



DECRYPTION

Fig 4: Block diagram for encryption and decryption using proposed algorithm

may look to be a herculean task, it is still possible as now days Brute-force attacks have become more sophisticated. Groups of expert video analysts may sit together and analyze the entire video frame by frame and may bring together the original video. Hence there is a need to increase the security further to prevent such Brute force attacks. To do this we would require an advanced encryption algorithm like AES.

AES is used to encrypt the codewords extracted from the MVDs, DCs and ACs. These codewords are a stream of digital bits. The AES algorithm will be used along with the individual encoding algorithms of ACs, DCs and MVDs. Only the important or sensitive codewords are extracted and encrypted so that computation time is saved. After encryption, the codewords are jumbled up and the video will look like layers of scattered colors. The blocks on each frame will be the same but their position will be changed. Thus the video is beyond human interception.

Finally after the video is transferred to the client, the decoder will first run the AES algorithm over the codewords to decode and obtain the clean video. The decryption block will also decode the random key and use it to reshuffle the frames to its original position. The frame stitching block once again comes into existence at the decoding side to stitch the frames and output the original video. The complete process of encryption and decryption is explained in Fig 4.

V. AES KEY GENERATION

Multimedia compressed bitstream is often transmitted over wireless channel, in which bit-error or packet-loss often occurs. Furthermore, the lost-packets may be encrypted. Thus, unless the encryption algorithm takes enough steps to achieve synchronization between the bitstream and the cipher, the encrypted bitstream will fail to be decrypted, even if the original key is known. In order to solve this problem, the H.264 bitstream is encrypted slice by slice in the proposed algorithm. AES is the next generation encryption algorithm which will replace DES. It is capable of using cryptographic keys of 128, 192 and 256 bits in encryption. Due to its security, simplicity and flexibility, AES has been widely used in various kinds of platforms. In addition, AES can be operated as a stream cipher, and the key expansion in AES is utilized to generate multiple keys instead of the original key. Thus, AES is very suitable for the proposed encryption algorithm.

We propose to encrypt slice by slice with multiple keys. Therefore, the packet-loss affects only the current slice, but not the slices that are yet to come.

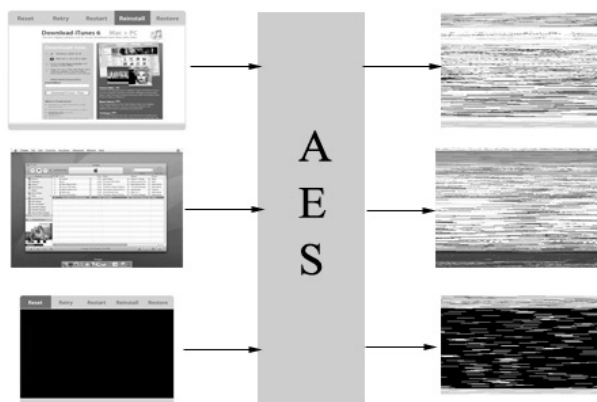


Fig 3: Frames after AES Encryption

VI. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

To evaluate the performance of this algorithm we have used a sample .mp4 video, 108 seconds long, which was encoded using H.264. Using MediaCoder we split the video into smaller videos containing one frame each. Using random key generation function in Java we shuffled the video files. Using PhotoLapse we stitched all videos together to make a video with shuffled frames. The codewords of the resultant video was then completely encrypted using VirtualDub. The table below shows the time in ms spent on each stage. All these computations were done on a standard computer.

Table I. Computation time of each task during encryption

Task	Computation Time (ms)
Video Shredding	1210
Shuffling	113
Video Stitching	1096
AES Encryption	1580

The total time required for encryption of a 108 second video was found to be 3999 ms which is approximately 4 seconds. That means approximately 3.7% of the video length is required to encrypt the video. Decryption takes exactly the same amount of time as it is just the reverse process.

All in all it took just took around 8 seconds to encrypt and decrypt a 108 second video. Thus this algorithm is efficient enough when it is used in new generation smart phones. Although the encryption time is slightly higher, the type of data protection it provides is unmatched.

VII. SECURITY ISSUES

A. Perceptual Attacks

The video encryption algorithms have to withstand not only classical cryptanalytic attacks, such as known-plaintext attacks, chosen-plaintext attacks and ciphertext attacks, but also the perceptual attacks, such as error-concealment-based attacks and replacement attacks. The error-concealment-based attacks means that attackers treat the encrypted parts as the loss caused by bit-error or packet-loss and recover them using error-concealment techniques. The replacement attacks means that attackers replace the encrypted parts with arbitrary data and makes the encrypted video recognizable. Figure 5 shows the encrypted videos recovered by error-concealment-based attacks and replacement attacks. As can be seen, all of the recovered videos are still unrecognizable. Thus, the proposed algorithm is immune to the perceptual attacks.

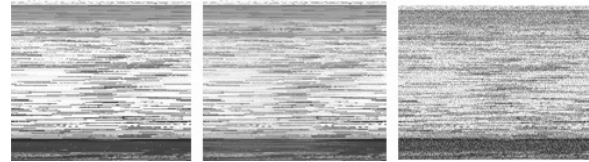


Fig 5: Original frame and frames after Perceptual Attacks

B. Brute-force attacks

Since the AES key lengths support 128, 192 and 256 bits, the minimum brute-force space for each slice is 2^{128} , which is too large for attackers to break. If we use 192 or 256 bit encryption security is further improved but computational time is increased, thus 128 bit AES encryption is sufficient enough to provide security from Brute-force attacks.

CONCLUSION

In this paper a novel H.264 video encryption algorithm is proposed which will keep a balance between computation time and security. Encryption is easily carried out using methods like shuffling of frames and AES. This algorithm provides high security at a low cost of your system resources. This algorithm is safe from all kinds of attacks and can practically be used to encrypt videos in the present generation.

REFERENCES

- [1] Data Encryption Standard (DES) (1999, Oct. 25). FIPS PUB 46-3
- [2] Announcing the Advanced Encryption Standard (AES) (2001, Nov. 26). FIPS PUB 197 J. Wen, M. Severa, W. Zeng, *et al.*
- [3] F. Liu, H. Koenig, "A survey of video encryption algorithms," *Computers and Security*, vol. 29, pp. 3-15, 2010.

ASIC VERIFICATION OF LEARNING AND AGING PROCESS IN ETHERNET BRIDGE

Sanjay C P

Department of Electronics and
Communication Engineering
Channabasaveshwara Institute of Technology
Gubbi, Tumkur – 572 216. Karnataka, INDIA
cpsanju@gmail.com

Vinay H M

Department of Electronics and
Communication Engineering
Channabasaveshwara Institute of Technology
Gubbi, Tumkur – 572 216. Karnataka, INDIA
vinayhulikal@gmail.com

ABSTRACT

In the last few decades, ethernet has become the most widely used LAN technology because of its speed, low cost and relative ease of installation. A large ethernet LAN with consistently heavy traffic is divided into smaller ethernet LANs, and connected in some manner. The easier way to do this is via a bridge. A bridge is a LAN interconnection device which operates at the data link layer of the OSI reference model. Learning and aging are the two important stages in the design of Ethernet bridge. In this paper both learning and aging stages of Ethernet bridge is implemented using verilog HDL and the design has been verified for different functionalities using industry standard Synopsys VCS tool and also the quality of the verification is checked through coverage results.

Keywords: - Ethernet, Bridge, Learning, Aging, Filtering, Learning Table, Predictor, Synthesize, Code coverage,

I. INTRODUCTION

Bridging is one of the easiest ways to connect network segments. A bridge is attached to two or more networks and simply forwards frames between them. The contents of these frames are of no concern to the bridge, frames are forwarded unchanged. Physical networks can be interconnected with bridges. It is a device which is able to connect these networks and to forward packets in between. A bridge (often better known as a two-port-switch) operates on layer 2, forwarding packets on the link layer based on the MAC address. These are the devices which analyze incoming frames, make forwarding decisions based on information contained in the frames, and forward the frames toward the destination. Learning stage includes extraction of source address and port number from Ethernet frame and maintains the database of all the nodes connected to LAN. The address database is capable of holding 64

entries. Learning stage is followed by aging, where will be removal of stale entries from the address table. The address which is least recently used is considered as stale entry. Those entries are found through search mechanism and aged out of address table.

II. ETHERNET BRIDGE

When a frame is received on any port, the bridge extracts the destination address from the frame, looks it up in the table, and determines the port to which that address maps. If the port on which the frame is received is the same port where the target destination resides, the bridge can simply discard the frame because it can be assumed that the target will have received the frame through the normal LAN delivery mechanisms. In the first case shown in Figure 1, the bridge receives a frame on port 1 being sent from station 08 - 00 - 60 - 00 - 00 - 46 to 08 - 00 - 60 - 00 - 00 - 09. Then the bridge will discard this frame because the address table indicates that station 08 - 00 - 60 - 00 - 00 - 09 is on port 1 and the frame was received on port 1. This is called *filtering* the frame.

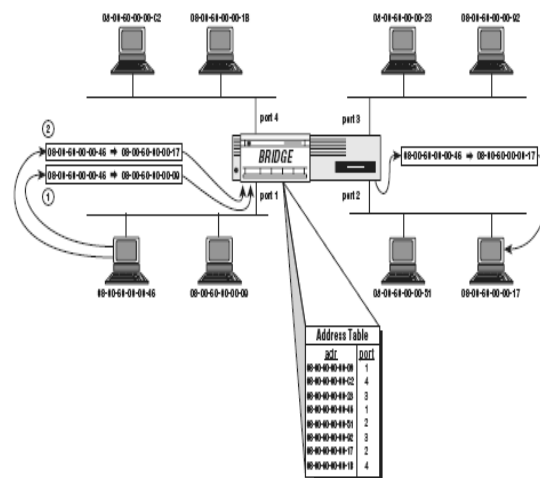


Fig. 1. Simple bridged LAN configuration.

The bridge filters all frames whose destination resides on the same port as that on which the frame arrived. It is important that the bridge not retransmit the frame onto port 1 (the port of arrival). This would cause a duplication of the frame that could cause problems in the end station(s), not to mention doubling the traffic on every port.

If station 08 - 00 - 60 - 00 - 00 - 46 now sends a frame to station 08 - 00 - 60 - 00 - 00 - 17, the bridge will receive this frame on port 1 (being in promiscuous mode, it is receiving all frames) and look up station 08 - 00 - 60 - 00 - 00 - 17 in the address table. The table indicates that the target destination is reachable through port 2. In order for the target to properly receive the frame, the bridge must now *forward* the frame onto port 2. Note that in the process of forwarding the frame.

III. DESIGN ARCHITECTURE

Design includes following Modules:

- I. Memory Model Design
- II. Design of Learning
- III. Design of aging

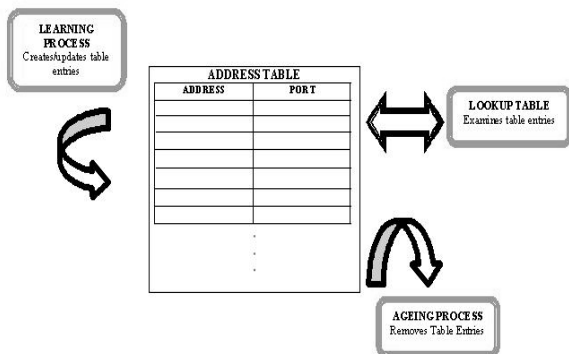


Fig. 2. Design Architecture & Flow for designing Ethernet Bridge

IV. TOP LEVEL DESIGN

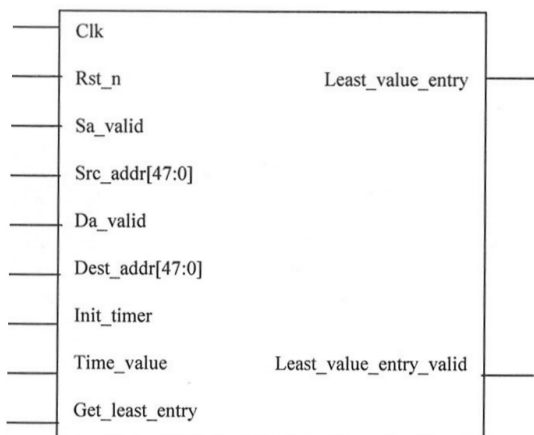


Fig. 3. Top Level Block Diagram

Input Output Definitions:

- a) Clock : Designed to function at rising edge of clock at 300MHz Frequency.

- b) Rst_n : Designed for active low reset signal.
- c) Sa_Valid: Active high Control input signal. Source address extracted from Ethernet frame only when it is made high signal.
- d) Src_addr: 48 bit source address.
- e) Da_Valid: Active high Control input signal. Destination address extracted from Ethernet frame only when it is made high signal.
- f) Dest_addr: 48 bit Destination address.
- g) Init_timer: Active high control input signal. The time values extracted for aging.
- h) Time_value: Input Time value to indicate Aging.
- i) Get_least_entry: Active high input signal, Request to find out the more aged entries.
- j) Least_value_entry: Output which indicates more aged entries.
- k) Least_value_entry_valid: Gives active high signal after finding the more aged entry.

V. SRAM MODEL

A. Need for SRAM Model

In designing a Learning Bridge, there should be a storage area, where all Source address, Destination address, Port number are stored. Hence, SRAM is designed.

B. Description of SRAM Model

The SRAM model is as shown below in figure 4. It consists of two ports, PORT A and PORT B. It is designed such that PORT A is used for Write operation and PORT B for Read operation.

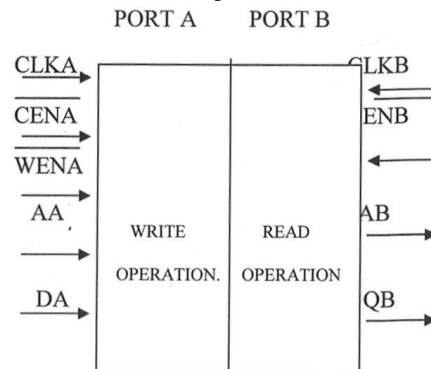


Fig. 4. SRAM Model

PORT A consists of:

1. CLKA: Input clock signal.
2. CENA: Chip Enable for PORT A active low input signal.
3. WENA: Write Enable active low input signal.
4. AA : Input Address for Write operation.
5. DA : Data Input.

PORT B consists of:

1. CLKB: Input Clock signal.
2. CENB: Chip Enable for PORT B active low input signal.
3. AB : Address where data can be read.
4. QB : Data can be read here.

VI. DESIGN OF LEARNING STAGE

There are three operations that need to be performed on the Bridge Address Table: Destination Address lookup, Source Address learning, and entry aging.

If the address table is sorted in numerical order, it becomes possible to perform a binary search for addresses, as depicted in Figure 5.

Manual address table administration can be avoided; the address table can be built automatically by considering the Source Address in received frames. Bridges perform a table lookup on the Destination Address in order to determine on which port(s) to forward a given frame.

The bridge can also perform a table lookup for an entry corresponding to the Source Address, indicating the station that sent the frame.

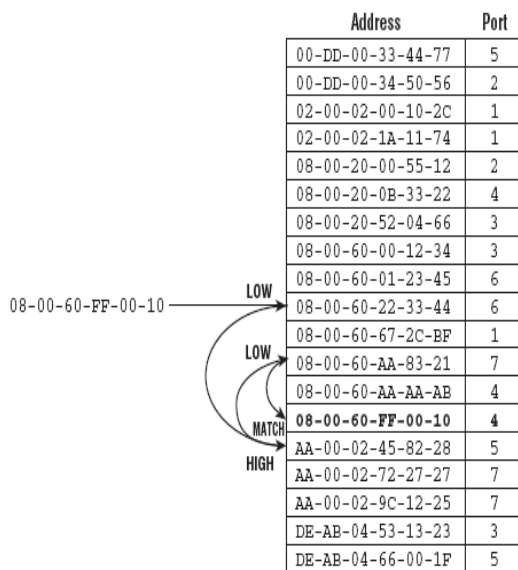


Fig. 5. Address search in Learning Table

If an entry is not found (i.e., the bridge has not heard from this station previously), the bridge creates a table entry for this newly learned address, with the port mapping indicating the port on which the frame arrived.

If an entry is already in the table, the port mapping is updated to reflect the port on which the latest frame arrived. This allows the bridge to properly map stations that have moved from one LAN segment to another. Over time, as stations send frames, the bridge will learn the address-to-port mapping for all active stations.

VII. AGING ENTRIES IN THE TABLE

The aging time parameter controls the pruning of old entries from the Bridge Address Table. The standard allows a range of aging times from 10 seconds to 1 million seconds (11.6 days), with a recommended default of 300 seconds (5minutes). Most commercial bridges use this default value or a value close to it.

Another Implementation of aging designed in this project is:

1. Append the counter value along with the Source address.
2. If the table is full, then least count value is found in the Learning table. It can be removed to take new entry.

The Verilog coding for the design is done using state machine concept. The State machine designed is as shown below in figure 6.

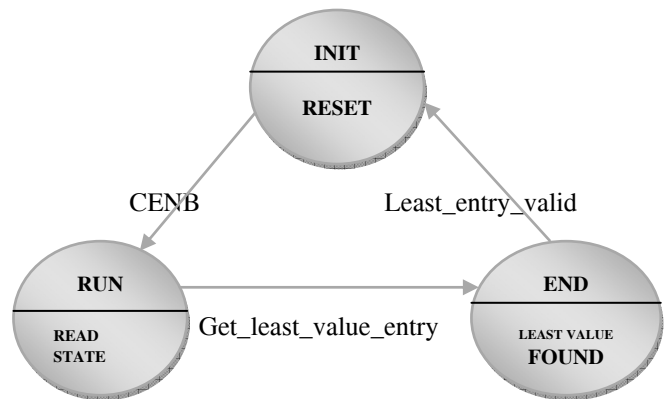


Fig. 6. State machine to find out least valued entry in learning table.

INIT: Indicates reset condition, where Address (AB), Chip enable for read (CENB) is made reset.

RUN: Indicates Read operation, Chip enable signal (CENB) made active low & all entries are read from the Learning table from address AB.

END: Indicates, the state where the least count value is found out. Input Get_least_value_entry initiates the operation and least_entry_value_valid gives the signal high, which indicates the least value is at least_value_entry register.

VIII. TEST BENCH ARCHITECTURE

Figure 7. Shows the block diagram of test bench which has been modeled for the design verification of this thesis. The block with red outline is the DUT, and boxes in blue outline are the test bench components. Coverage is a separate block which gets events from the input and output monitors. It is the same as the checker, but does something more.

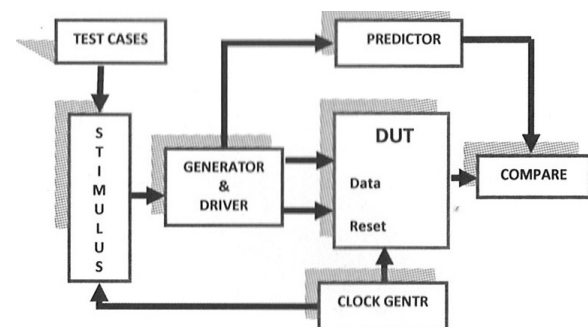


Fig.7. Block diagram showing various components

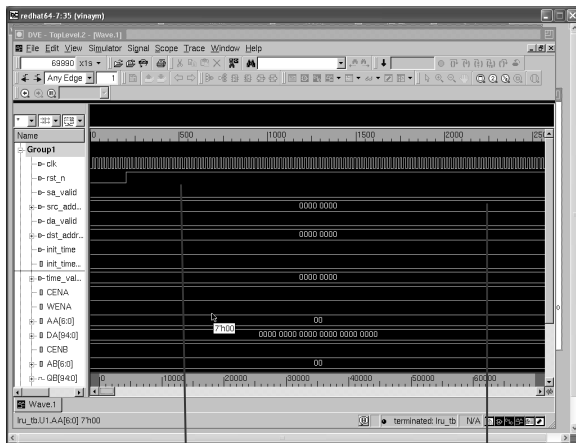
IX. CODE COVERAGE

After the completion of the verification with numerous tests we can check whether the test bench has satisfactorily exercised the design or not. Code coverage is used to measure the efficiency of verification implementation. It provides a quantitative measurement of the testing space. It describes the degree to which the source Code of a DUT has been tested. It is also referred as structural coverage.

These are the different types of coverage covered after the verification of the system design.

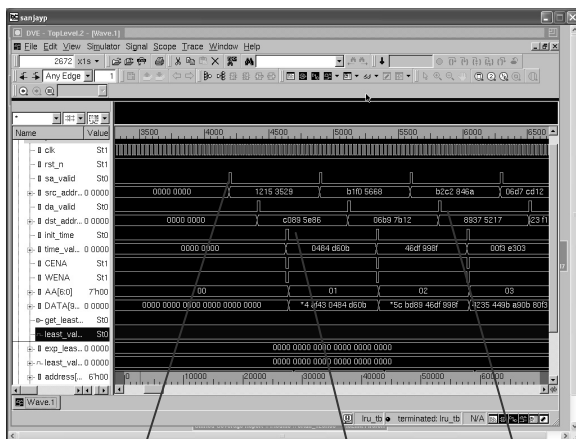
- Statement coverage/line coverage
- Conditional
- Branch coverage
- Toggle coverage

X. SIMULATION RESULTS



Screen shot 1
Rst_n active low, reset state.

Since sa_valid, da_valid are not activated, no extraction of source & destination address



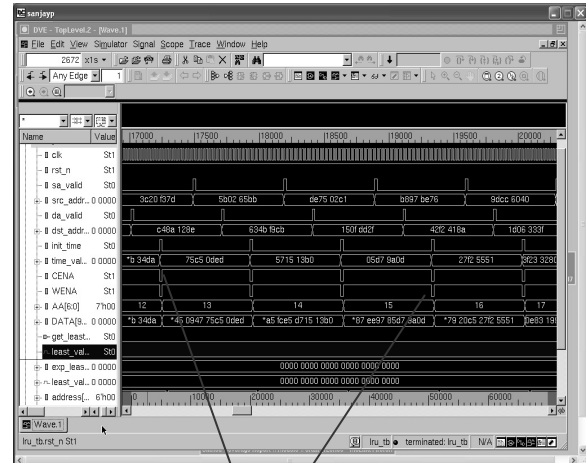
Screen shot 2

SA_Valid

init_time

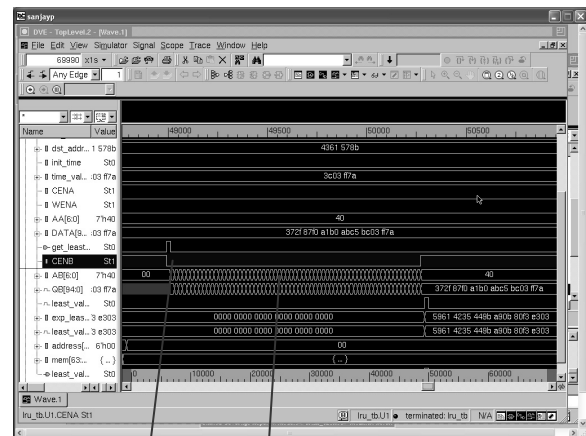
DA_valid

sa_valid, da_valid, init_time are activated to extract source address, destination address, time values. From Screen shot 2, upon active high on both the valid bits of the source address and destination address, respective addresses will be learn't in the address table which infers that the learning table is getting proper MAC addresses.



Screen shot 3

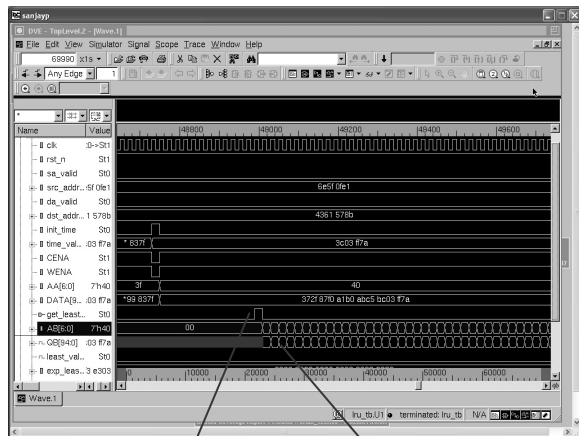
CENA and WENA Signals enabling memory write operation where the data is being written into the memory.



Screen shot 4

CENB enabling memory read operation as the learning table has been filled with all 64 entries.

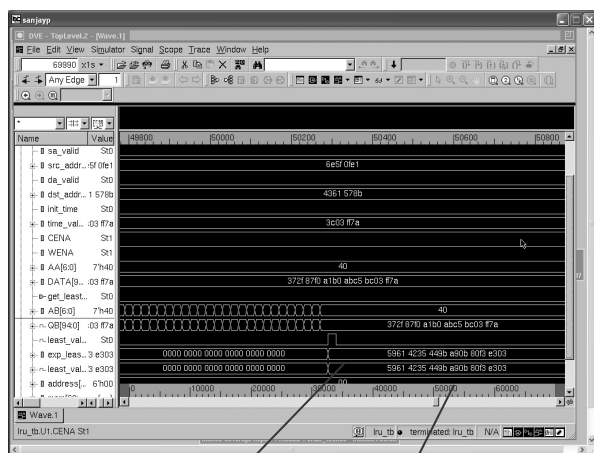
Address AB showing the data ready to be read at respective addresses.



Screen shot 5

Get_least_entry=1

Data QB is read



Screen shot 6

On Least entry valid =1, the least recently used entry is ready in the next clock finishing its search operation.

Least recently used entry found

A. Device Utilization Summary

Selected Device: 5v1x330ff1760-2

Table 1. Device Summary

Slice Logic Distribution	
Number of LUT Flip Flop pairs used	255
Number of fully used LUT-FF pairs	209 out of 255 81%
IO Utilization	
Number of IOs	197
Number of bonded IOBs	197 out of 1200 16%
Timing Summary	
Minimum period	3.510ns
Maximum Frequency	284.890MHz
Minimum input arrival time before clock	3.176ns
Maximum output required time after clock	2.830ns
Maximum combinational path delay	0.437ns

From the above synthesis report it is observed that the number of slices, LUTs, flip flops, shift register and IOs used is less and also the Timing Report.

XI. COVERAGE RESULTS

Date: Wed Jun 24 16:36:58 2012

User: sanjaycp

Version: Y-2006.06-SP2

Command line: urg -dir ./simv.cm

SCORE LINE COND TOGGLE BRANCH

81.43	94.74	26.19	66.93	82.61
-------	-------	-------	-------	-------

Source File(s):

/home/sanjaycp/synthesis/lru_pkt.v

A. Line coverage Report

Table 2. Line Coverage for Module: learn_table

	Line No.	Total	Covered	Percent
TOTAL		57	54	94.74
ALWAYS	59	11	11	100.00
ALWAYS	76	14	14	100.00
ALWAYS	97	32	29	90.62

B. Toggle Coverage Report

Table 3. Toggle Coverage for Module: learn_table

	Total	Covered	Percent
Totals	25	17	68.00
Total Bits	1388	929	66.93
Total Bits 0->1	694	507	73.05
Total Bits 1->0	694	422	60.81

Table 4. Nets Coverage Summary

Nets	11	10	90.91
Net Bits	582	581	99.83
Net Bits 0->1	291	291	100.00
Net Bits 1->0	291	290	99.66

Table 5. Register bits coverage summary

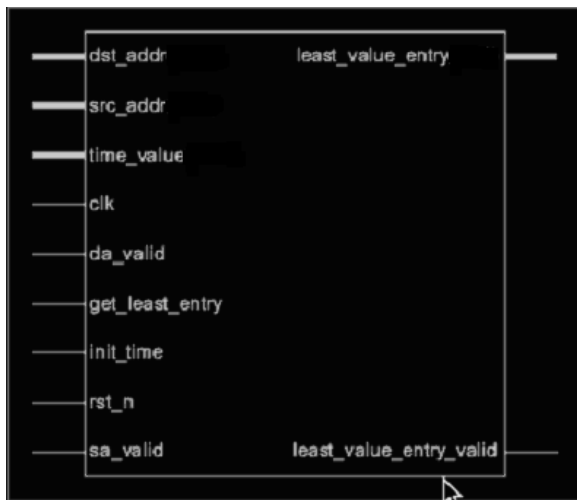
Regs	14	7	50.00
Reg Bits	806	348	43.18
Reg Bits 0->1	403	216	53.60

C. Branch Coverage Report

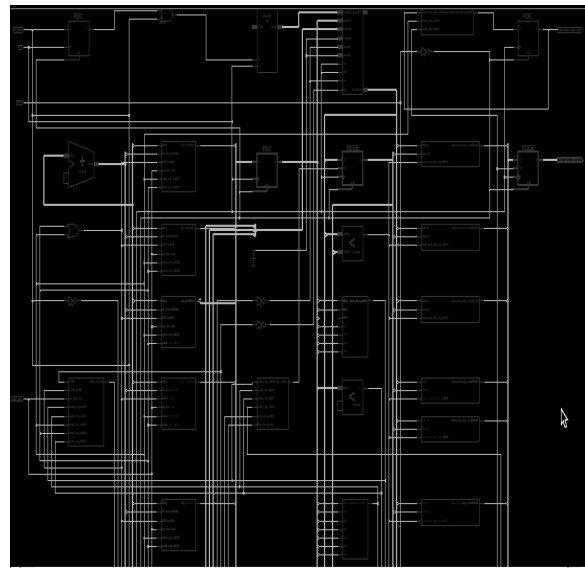
Table 6. Branch Coverage for Module: learn_table

	Line No.	Total	Covered	Percent
Branches		23	19	82.61
IF	59	7	7	100.00
IF	76	4	4	100.00
IF	97	12	8	66.67

XII. RTL Schematics



Top level model of the design



RTL blocks for the complete design

CONCLUSION

This paper presents the design verification methodology for the implementation of learning and aging process in Ethernet Bridge. The design undergoes complete verification steps involved in ASIC. Learning address table being the initial block in bridge plays a vital role in bridge performance.

REFERENCES

- [1] Alberto Leon-Garcia and Indra Widjaja, "Fundamental Concepts & Key Architecture"
- [2] Jim Edwards and Rich Seifert, "The All-New Switch Book - The Complete Guide to LAN Switching Technology" Second Edition
- [3] Samir Palnitkar, "A Guide to Digital Design and Synthesis" *Second Edition*
- [4] Zhi-xue Wang, Hai-rong Wang, Jun-hua Song, "The Application of RTL8019AS Network Card Chip in Ethernet Communication," *Modern Instruments*, vol.4, 2009, pp. 15-17.
- [5] Weng Fook Lee, "Verilog Coding for synthesis" A JOHN WILEY & SONS, INC., PUBLICATION
- [6] Icarus verilog. <http://icarus.com/eda/verilog/>.
- [7] Xilinx Inc. Spartan-3 Generation FPGA User Guide, 2008.
- [8] Xilinx Inc. Spartan-3E FPGA Family: Complete Data Sheet, 2008.

TEMPERATURE PREDICTION USING ARTIFICIAL NEURAL NETWORK

Deepak C. Patil

Computer Engineering

Fr.CRIT

Vashi, India

deepakpatil2312@gmail.com

Miron Mangalath

Computer Engineering

Fr.CRIT

Vashi, India

Miron.mangalath@gmail.com

Pritam P. Pachpute

Computer Engineering

Fr.CRIT

Vashi, India

pritam Pachpute137@gmail.com

Sushant Dhende

Computer Engineering

Fr.CRIT

Vashi, India

Sdhende1@gmail.com

ABSTRACT

Weather Data Mining is the data mining technique which deals with the study of huge metrological data and finding the hidden patterns inside it such as the rate of change of weather conditions. Same concept can be applied to the temperature forecasting to find interrelation between the meteorological data such as temperature, wind speed, humidity etc. using data mining technique to predict the temperature. This paper deals with utilization of 'Artificial neural network' for temperature prediction of Mumbai city, India. Our study is based on back propagation neural network which is trained and tested based on provided dataset. In formulating ANN based predictive model using back propagation three layered network is being constructed. This paper also discusses artificial neural network and back propagation algorithm used in the temperature prediction is explained in this paper.

Keywords- *Artificial Neural Network (ANN), Back Propagation Neural (BPN) Network, Meteorological data, Data Mining.*

I. INTRODUCTION

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time and a given location. [3] Weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to predict how the atmosphere will evolve. Weather forecasting is an ever-challenging area of investigation for scientists.

In this paper a temperature forecasting model using Neural Network is discussed. The weather parameters like maximum temperature, minimum temperature is

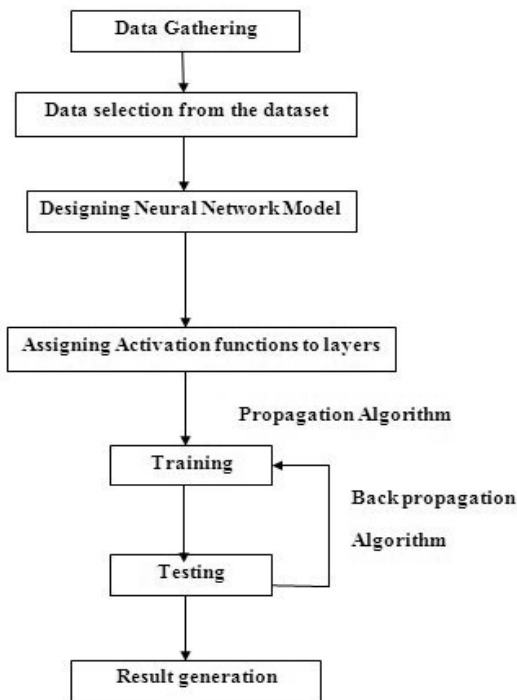
predicted using the features extracted over different periods as well as from the weather parameter time-series itself. The approach applied here uses feed forward "Artificial Neural Networks" (ANNs) with back propagation for supervised learning using the data recorded at a particular station (here Mumbai). The trained ANN was used to predict the one day ahead temperature conditions. The model can be suitably adapted for making forecasts over larger geographical areas.

Artificial Neural Network (ANN), a component of soft computing, is highly suitable for the situations where the underlying processes exhibit chaotic features. The concept of ANN is originated from the attempt to develop a mathematical model capable of recognizing complex patterns on the same line as biological neuron work [2]. In this we model the methods as that of biological neurons so that the Neural Network can identify meanings from imprecise or unstructured data. It can be identified as an "expert" in that field who can answer complex question given the circumstances that fall under its domain. ANN does not require any prior knowledge of the system under consideration and are well suited to model dynamical systems on a real-time basis. ANN uses a set of processing elements (or nodes) that we can compare as to that of neurons in the brain. These nodes are connected in such a way so that it can identify patterns in data once it is exposed to the data, i.e. the network learns from experience just as people do.

It is, therefore useful in real time analysis, e.g., weather forecasting, temperature predictions, fog prediction etc. Back propagation is the technique used along with the ANN so as to minimize errors in the predicted output over the provisioning of the learning data set. Back propagation is a learning technique that adjusts weights in the NN by propagating weight changes backward from the sink (output node) to the source node (input node).

II. MATERIAL AND METHODS

Temperature prediction system can be shown in the flow chart as follows:



A. Data Gathering

The data used for training the Neural network (NN) is taken from “wunderground.com”[6]. The data set that is used contains parameters such as Temperature, Dew point, Humidity, Visibility, Wind speed, Sea level pressure etc with their maximum, minimum and average values out of which the required attribute in temperature prediction were taken.(e.g. dataset of January 2012)

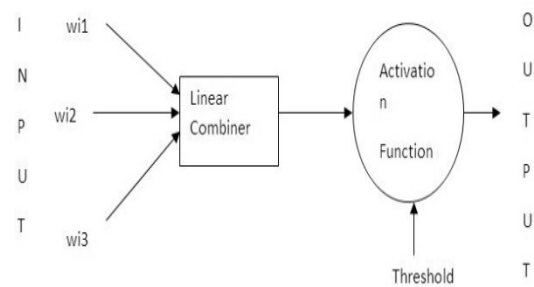
B. Data selection from the dataset

From the gathered meteorological data we have chosen maximum, minimum, average values of temperature & humidity from(eg. DEC 2011 to FEB 2012).

Values from Dataset			
Date	Maximum	Average	Minimum
2011-12-02	32	29	26
2011-12-03	34	29	24
2011-12-04	34	29	24
2011-12-05	31	28	25
2011-12-06	33	29	25
2011-12-07	33	28	24
2011-12-08	34	29	25
2011-12-09	34	29	24
2011-12-10	34	28	23

C. Designing Neural Network Model

The figure below shows the structure of a typical artificial neuron[1]:



It consists of nodes (processing elements) and arcs (interconnections) which is analogous to neurons in the brain. Each of these processing elements functions independently from the others and uses only local data (input and output to the node) to direct its processing. It can be viewed as a directed graph with source (input), sink (output), and internal (hidden) nodes. The input node is a part of the input layer, while the output nodes belong to output layer. The hidden nodes exist over one or more hidden layers [1] and each node is labelled with a function that indicates its effect on the data coming into that node. In this we insert a tuple as an input through the nodes and the output node determines what the prediction is. Although the structure of the graph does not change, the labelling of the edges may change (due to weight adjustment of edges of NN during back propagation).

Linear Combiner

The NN can be considered as structure with k input arcs coming from nodes $1, 2, \dots, k$. With weights of $w_{1i}, \dots, w_{ki}$ and input values of $x_{1i}, \dots, x_{ki}$ and f_i is activation function.

Input and weight on the arcs into node i are

$$[x_{1i}, \dots, x_{ki}], [w_{1i}, \dots, w_{ki}]^T$$

There is one output value from node i, y_i , which is propagated to all output arcs during the propagation process. The linear combiner performs summation to combine the inputs with the weights assigned to the edges, then, the output of a node is

$$Y_i = f_i\left(\sum_{j=1}^k w_{ji}x_{ji}\right) = ([w_{1i}, \dots, w_{ki}] [x_{1i}, \dots, x_{ki}]^T) \dots [1]$$

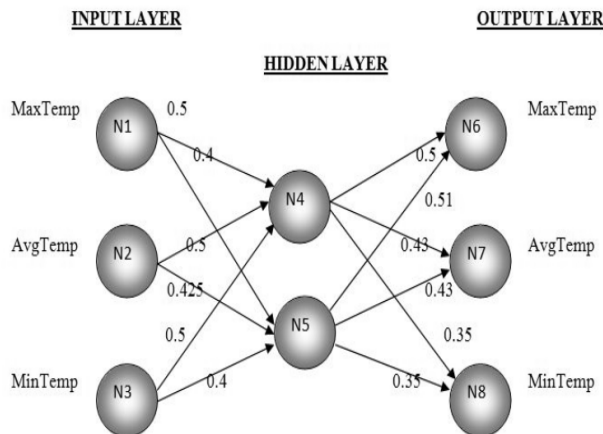
Activation Function

An activation function is sometimes called a processing element function or a sequencing function. The function is applied to the set of inputs coming in on the input links. There are various types of activation function including threshold, sigmoid, symmetric sigmoid and Gaussian [1] which are selected as requirement of the application.

Weights

The weight assigned to an arc indicates the relative weight between those two nodes. Initial weights are

usually assumed to be small positive numbers and are assigned randomly generally taken between 0 and 1 as per the application but it can be negative also.



Thus our implementation consists of NN model for temperature forecasting including parameters like maximum, minimum and average temperature as an input and output along with weights per link is given as below.

D. Assigning Activation functions to layers

Activation functions play a key role in the output generation. Thus an activation function is assigned to layers except the input layer which depends upon the type of application. In ANN the input should be above a certain level to classify it in correct output. If there are only 2 output nodes then output can be either 0 or 1 else we can use -1 and 1 as output. The activation function 'fi' is applied to input nodes 'xi' and their corresponding weights 'wi' and summed to classify the input tuple. The activation function can be applied in any input to hidden layer or any hidden to output layer.

As in our Temperature Prediction Project we have used 'Threshold' function for input to hidden layer nodes and 'Linear' function for hidden to output layer nodes.

AT Input Layer:

Input : $X = \{\text{MaxTemp}, \text{AvgTemp}, \text{MinTemp}\}$

AT Hidden Layer:

Input to each node:

$S_i = \text{Summation of } (W_{ij} * X_j)$ Where i: node from input layer, j: node from hidden layer

Output from node : (Using Threshold Activation Function)

$H_i = S_i - 10$ When $S_i > 38$

$H_i = S_i - 2$ When $S_i \leq 38$

AT Output Layer:

Input to each node:

$S_i = \text{Summation of } (W_{ij} * H_j)$ Where i: node from hidden layer, j: node from Output layer

Output from node : (Using Linear Activation Function)

Maximum temperature $Y_i = S_i + S_i / (10 * S_i + 1)$

E. Training

In this phase one record from the record set is given as an input to NN. The input is propagated through NN using above formulae (Summation of product of weight and input at each layer). The various activation functions are applied at each hidden layer (if more than 1 hidden layer) and accordingly the appropriate intermediate output is passed to next layer. This process is continued till the final output is generated at output layer.

F. Testing

Testing the data comes under supervised learning where NN is processed by adjusting the arc weights based on its performance with a tuple from the training set. The behaviour of the training data is known a priori and thus can be used to fine tune the network for the better behaviour in future in similar conditions.

The error thus introduced in the output and the desired output is used to train the neural network by adjusting the weights of the arcs (edges) of the neural network using "Back propagation algorithm" and formulae to modify weights with the help of difference.

During Back-propagation (Weight re-evaluation) for:

Difference $\text{diff} = d_i - Y_i$

Weights between the nodes coming to Y1 (Maxtemp): $W_{ij} = W_{ij} + (\text{diff} / 10) * 0.9$

Weights between the nodes coming to Y2 (Avgtemp): $W_{ij} = W_{ij} + (\text{diff} / 10) * 0.3$

Weights between the nodes coming to Y3 (Mintemp): $W_{ij} = W_{ij} + (\text{diff} / 10) * 0.15$

Temperature forecast result for next 3 day(s)		
Maximum	Average	Minimum
31.81105315045	26.98255445005	22.949386794179
31.81105315045	26.98255445005	22.949386794179
31.811053150448	26.982554450049	22.949386794179

G. Result Generation

After training the data weights are so adjusted so that the NN is capable of forecasting the temperature. So, depending on user's interest, forecasting max, min and

avg temperature for number of days(daily, weekly etc) is the required result.

The accuracy in the prediction of the temperature using NN model discussed above is

Accuracy of the prediction (%) :

(If value greater than 100% means value-100 is error)

Average Accuracy in Max Temp Prediction : 98.384700465307

Average Accuracy in Avg Temp Prediction : 98.716662622133

Average Accuracy in Min Temp Prediction : 99.779942583387

III. CONCLUSION

Back Propagation Algorithm and Advanced Neural Network can be used for predicting the atmospheric condition based on the training set provided to the neural network. In this approach of temperature prediction which combines both Back Propagation Network (BPN) and Advanced Neural Network Model effectively. Back Propagation Network and Advanced Neural Network based approach for weather forecasting is capable of yielding good results and can be considered as an alternative to traditional meteorological approaches.

So, this approach is capable of determining the non-linear relationship that exists between the historical data (temperature, wind speed, humidity, etc.,) supplied to the system during the training phase and on that basis, make a prediction of what the temperature would be in future.

IV. ACKNOWLEDGEMENT

We express our sincere gratitude to **Mrs. Kiruthika M.**, Associate Professor, Computer Department, FR.C. Rodrigues Institute of Technology, Vashi, Navi Mumabai for help and support. We would also like to thank her for providing us with her valuable inputs which enabled us to complete the project in a hassle free manner.

REFERENCES

- [1] MaragarethH.Dunham,"*Data Mining Introductory and Advanced Topics*",Pearson Education,6th Impression 2009.
- [2] Parag P. Kadu, KishorP.Wagh and Prashant N. Chatur, "*Temperature prediction system using back propagation neural network*", International journal of Computer Science and Communication Networks, Vol2(1),61-64.
- [3] MoheshHayati and ZarahMohebi,"*Application of Artificial neural networks for Temperature Forecasting*",World Academy of Science,Engineering and Technology 28 2007.
- [4] SoumadipGhosh and AmitavaNag,"*Weather data mining using artificial neural network*",IEEE 2011.
- [5] MoheshHayati and ZarahMohebi,"*Temperature Forecasting based on Neural Network Approach*",World Applied Science Journal 2(6):613-620,2007,ISSN 1818-4952.
- [6] www.wunderground.com/history

GENETIC APPROACH FOR FUZZY MINING USING MODIFIED K-MEANS CLUSTERING

Shaikh Nikhat Fatma

Department Of Computer , Mumbai University
Pillai's Institute Of Information Technology,
New Panvel, Mumbai , India
nikhats10@yahoo.com

Dr. J. W Bakal

Department Of Information Technology ,
Mumbai University
Mumbai , India
bakaljw@gmail.com

ABSTRACT

A fuzzy-genetic data-mining algorithm for extracting both association rules and membership functions from quantitative transactions is shown in this paper. It used a combination of large 1-itemsets and membership-function suitability to evaluate the fitness values of chromosomes. The calculation for large 1-itemsets could take a lot of time, especially when the database to be scanned could not totally fit into main memory. In this system, an enhanced approach, called the cluster-based fuzzy-genetic mining algorithm. It divides the chromosomes in a population into clusters by the k-means clustering approach and evaluates each individual according to both cluster and their own information.

Keywords— *Modified k-means Clustering, data mining, fuzzy set, genetic algorithm, Fuzzy Association Rules, Quantitative transactions*

I. INTRODUCTION

Data is very important for every supermarket. Data that was measured in gigabytes until recently, is now being measured in terabytes, and will soon approach the petabytes range. In order to achieve our goals, we need to fully exploit this data by extracting all the useful information from it. This information can be extracted using data mining. Data mining is the use of automated data analysis techniques to uncover previously undetected relationships among data items. Data mining often involves the analysis of data stored in large database. The typical business decisions that the management of a super market has to make is, what to put on sale in what quantity. Analysis of past transaction data is a commonly used approach in order to improve the quality of such decisions. Extraction of frequent item sets is essential towards mining interesting patterns from datasets. A typical usage scenario for searching frequent patterns is the so called "market basket analysis" that involves analyzing the transactional data of a super market in order to determine which products are purchased together and how often and also examine customer purchase preferences.

Mining association rules between sets of items in large databases was first stated by Agrawal, Imelinski and Swami in 1993 and it opened brand new family algorithms. Apriori algorithm is probably the most used algorithm in association rules mining. At present, more and more databases containing large quantities of data are available. These industrial, medical, financial and other databases make an invaluable resource of useful knowledge. The task of extraction of useful knowledge from databases is challenged by the techniques called data-mining techniques. One of the widely used data-mining techniques is association rules mining. Data Mining is commonly used in attempts to induce association rules from transaction data. Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes. Many types of knowledge and technology have been proposed for data mining. Among them, finding association rules from transaction data is most commonly seen. Most studies have shown how binary valued transaction data may be handled. For example, there may exist some implicitly useful knowledge in a large database containing millions of records of customers' purchase orders over the last five years. The knowledge can be found out using appropriate data-mining approaches. Data mining is most commonly used in attempts to induce association rules from transaction data. An association rule is an expression $X \rightarrow Y$, where X is a set of items and Y is a single item. It means in the set of transactions, if all the items in X exist in a transaction, then Y is also in the transaction with a high probability. For example, assume whenever customers in a supermarket buy bread and butter, they will also buy milk. From the transactions kept in the supermarkets, an association rule such as "*Bread and Butter $\rightarrow$ Milk*" will be mined out. Transactions with quantitative values are however commonly seen in real-world applications. Therefore a fuzzy mining algorithm by which each attribute used only the linguistic term with the maximum cardinality in the mining process. The number of items was thus the same as that of the original attributes, making the processing time reduced. The fuzzy association rules derived in this way are not complete. The algorithm can derive a more complete set of rules but with more computation time.[1][2]

Fuzzy set theory is being used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning. The theory has been applied in fields such as manufacturing, engineering, diagnosis, economics. Several fuzzy learning algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains. The mined rules are expressed in linguistic terms, which are more natural and understandable for human beings.

II. EXISTING SYSTEM

Data mining is commonly used in attempts to induce association rules from transaction data. Most previous studies focused on binary-valued transaction data. Transactions in real-world applications, however, usually consist of quantitative values. In the past, a fuzzy-genetic data-mining algorithm for extracting both association rules and membership functions from quantitative transactions. It used a combination of large 1-itemsets and membership-function suitability to evaluate the fitness values of chromosomes. The calculation for large 1-itemsets could take a lot of time, especially when the database to be scanned could not totally fit into main memory. Therefore, the k-means clustering approach was used to gather similar chromosomes into groups. Since the number for scanning a database will decrease, the evaluation cost can thus be reduced.

The existing system maintains a population of sets of membership functions, and use the genetic algorithm to automatically derive the resulting one. It will first transform each set of membership functions into a fixed-length string.

Each chromosome will represent a set of membership functions used in fuzzy mining. Then, it will use the K-means clustering approach to gather similar chromosomes into groups. All the chromosomes in a cluster will use the number of large 1-itemsets derived from the representative chromosome in the cluster and their own suitability of membership functions to calculate their fitness values. Since the number for scanning a database will decrease, the evaluation cost can thus be reduced. The evaluation results can be utilized to choose appropriate chromosomes for mating in the next generation. The offspring membership function sets will then undergo recursive evolution until a good set of membership functions has been obtained. Finally, the derived membership functions will be used to mine fuzzy association rules.

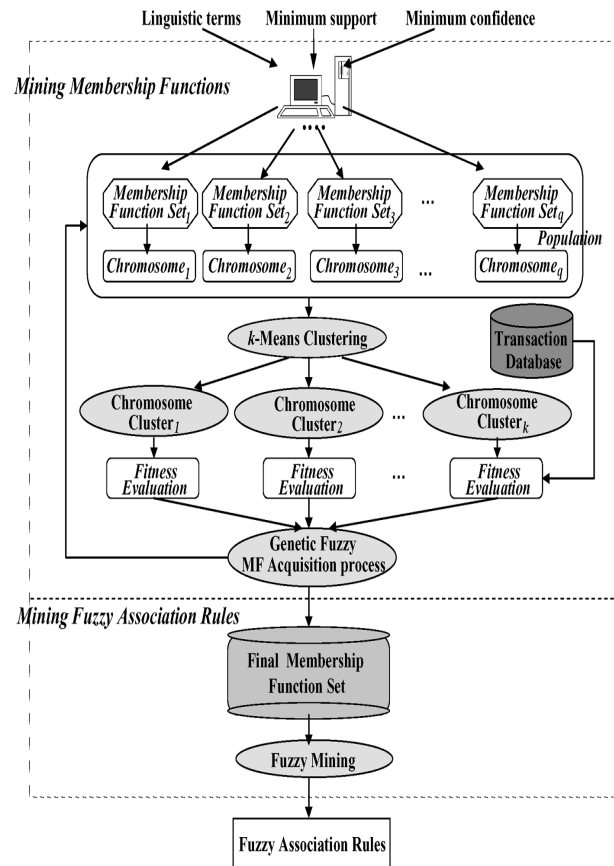


Fig 1: Framework for the existing cluster based fuzzy genetic data mining system

III. PROPOSED SYSTEM

A) Chromosome Representation:

Each set of membership functions is encoded. All pairs of (c.w)'s for a certain item are concatenated to represent its membership functions. Since and are both numeric values, a chromosome is thus encoded as a fixed-length real-number string rather than a bit string.[1][2]

B) Initial Population

A genetic algorithm requires a population of feasible solutions to be initialized and updated during the evolution process. As mentioned above, each individual within the population is a set of isosceles-triangular membership functions. Each membership function corresponds to a linguistic term in a certain item. The initial set of chromosomes is randomly generated with some constraints or forming feasible membership function.[1][2]

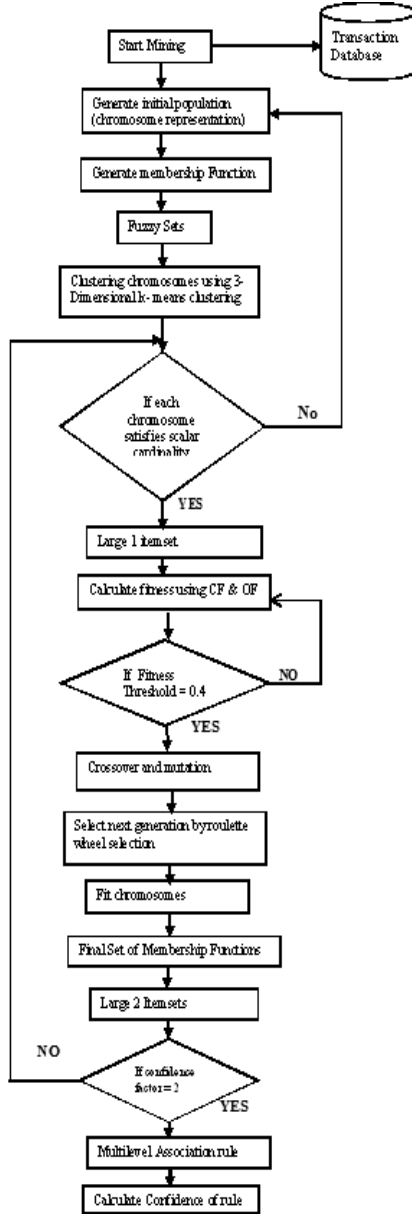


Fig 2: Flowchart for the proposed cluster based fuzzy genetic data mining system

C) Fitness and Selection

In order to develop a good set of membership functions from an initial population, the genetic algorithm selects *parent* membership function sets with its probability values for mating. An evaluation function is then used to qualify the derived membership function sets. It is shown as follows:

$$f(C_q) = \frac{|L_{1q}|}{\text{Suitability}(C_q)}$$

Where $|L_{1q}|$ is the number of large 1-itemsets obtained by using the set of membership functions in chromosome C_q and suitability of (C_q) represents the shape suitability of (C_q) . Suitability of (C_q) is defined as:

$$\sum_{i=1}^m [\text{overlap_factor}(C_{qi}) + \text{coverage_factor}(C_{qi})]$$

Where m is the number of item. Overlap_factor (C_{qi}) represents the overlap factor of the membership functions for an item I_j in the chromosome C_q and is defined as

$$\text{overlap_factor}(C_{qi}) = \sum_{k \neq i} \left[\max \left(\left(\frac{\text{overlap}(R_{jk}, R_{ji})}{\min(w_{jk}, w_{ji})} \right), 1 \right) - 1 \right]$$

Where $(R_{j1}, R_{j2}, \dots, R_{ji})$ is the coverage range of the membership functions I_j , is the number of membership functions for I_j and $\max(I_j)$ is the maximum quantity of in the transactions.

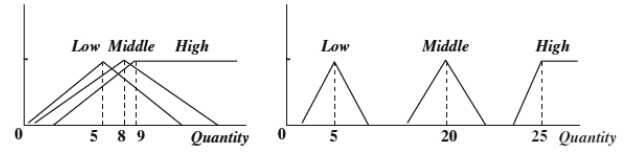


Fig 3: Two bad membership functions

D) Clustering Chromosomes

The coverage factors and overlap factors and the average of both of all the chromosomes are used to form appropriate clusters. The Modified k-means clustering approach is adopted here to cluster chromosomes. Since the chromosomes with similar coverage factors and overlap factors will form a cluster, they will have nearly the same shape of membership functions and induce about the same number of large 1-itemsets. For each cluster, the chromosome which is the nearest to the cluster center is thus chosen to derive its number of large 1-itemsets. All chromosomes in the same cluster then use the number of large 1-itemsets derived from the representative chromosome as their own. Finally, each chromosome is evaluated by this number of large 1-itemsets divided by its own suitability value.

E) Genetic Operators

Two genetic operators, *crossover* proposed in and the *mutation* are used. Assume there are two parent chromosomes

$$C_u^{rt} = (c_{11}, \dots, c_{1h}, \dots, c_z) \text{ and } C_w^{rt} = (c_{11}, \dots, c_{1h}, \dots, c_z)$$

The crossover operator will generate the following four candidate chromosomes from them:

- 1) $C_1^{t+1} = (c_{11}^{t+1}, \dots, c_{1h}^{t+1}, \dots, c_{1Z}^{t+1})$, where $c_{1h}^{t+1} = dc_h + (1-d)c'_h$;
- 2) $C_2^{t+1} = (c_{21}^{t+1}, \dots, c_{2h}^{t+1}, \dots, c_{2Z}^{t+1})$, where $c_{2h}^{t+1} = dc'_h + (1-d)c_h$;
- 3) $C_3^{t+1} = (c_{31}^{t+1}, \dots, c_{3h}^{t+1}, \dots, c_{3Z}^{t+1})$, where $c_{3h}^{t+1} = \min\{c_h, c'_h\}$;
- 4) $C_4^{t+1} = (c_{41}^{t+1}, \dots, c_{4h}^{t+1}, \dots, c_{4Z}^{t+1})$, where $c_{4h}^{t+1} = \max\{c_h, c'_h\}$.

Where the parameter d_i is either a constant or a variable whose value depends on the age of the population.

IV. ALGORITHM

INPUT:

A body of n quantitative transactions, a set of m items, each with a number of linguistic terms, a parameter k for Modified k -means clustering, a population size P , a crossover rate P_c , a mutation rate P_m , a support threshold α , and a confidence threshold λ and a fitness threshold.

OUTPUT:

A set of multilevel fuzzy association rules with its associated set of membership functions and its confidence.

Step 1: Randomly generate a population of P individuals, each individual is a set of membership functions for all the m items.

Step 2: Encode each set of membership functions into a string representation according to the schema explained above.

Step 3: Calculate the coverage_factor and the overlap_factor of each chromosome using the formulas defined above.

Step 4: Divide the chromosomes into K clusters by the Modified k -means clustering approach based on the three attributes (coverage factors and overlap factors and average of both) Find out the representative chromosome in each cluster, which is the nearest to the center.

Step 5: Calculate the number of large 1-itemsets for each representative chromosome.

Step 6: Calculate the fitness value of each chromosome using the number of large 1-itemsets of its representative chromosome and the suitability value of its own according to the formula defined above.

Step 7: Execute the crossover operation on the population.

Step 8: Execute the mutation operation on the population.

Step 9: Use the Roulette-wheel selection operation to choose appropriate individuals for the next generation.

Step 10: If the termination criterion is not satisfied, go to Step 6; otherwise, do the next step.

Step 11: Get the set of membership functions with the highest fitness value.

Step 12: This set of membership functions are then used to mine fuzzy association rules from the given database.

To mine fuzzy association rules from the above membership functions the algorithm works as follows:

Step 1: Transform the quantitative values of each transaction, for each attribute into a fuzzy set using above found membership functions.

Step 2: Calculate the count of each attribute region (linguistic term) R_{jk} in the transaction data.

Step 3: Collect each attribute region (linguistic term) to form the candidate set CI .

Step 4: Check whether $count_{jk}$ of each R_{jk} is larger than or equal to the predefined minimum support value α . If R_{jk} satisfies the above condition, put it in the set of large 1-itemsets ($L1$).

Step 5: IF $L1$ is not null, then do the next step; otherwise, exit the algorithm.

Step 6: Set $r=1$, where r is used to represent the number of items kept in the current large itemsets.

Step 7: Join the large itemsets Lr to generate the candidate set $Cr+1$ in a way similar to that in the apriori algorithm, except that two regions (linguistic terms) belonging to the same attribute cannot simultaneously exist in an itemset in $Cr+1$. Restated, the algorithm first joins Lr and Lr under the condition that $r-1$ items in the two itemsets are the same and the other one is different. It then keeps in $Cr+1$ the itemsets which have all their sub-itemsets of r items existing in Lr and do not have any two items R_{jp} and R_{jq} ($p \neq q$) of the same attribute R_j .

STEP 8: Do the following substeps for each newly formed candidate itemset.:

(a) Calculate the fuzzy membership value of each transaction datum. Here, the minimum operator is used for the intersection.

(b) Calculate the scalar cardinality (count) of each candidate 2-itemset in the transaction data.

(c) Check whether these counts are larger than or equal to the predefined minimum support value λ , put s in $Lr+1$.

Step 9: If $Lr+1$ is null, then do the next step; otherwise, set $r=r+1$ and repeat Steps 6 to 8.

Step 10: Collect the large itemsets together.

Step 11 : Construct association rules for each large q -itemset s with items($s_1, s_2, \dots, s_q$), $q \geq 2$, using the following substeps:

(a) Form each possible association rule as follows:

$s_1 \wedge \dots \wedge s_{k-1} \wedge s_{k+1} \wedge \dots \wedge s_q \rightarrow s_k$, $k = 1$ to q .

(b) Calculate the confidence factors for the above association rules.

Step 12: Output the association rules with confidence values larger than or equal to the predefined confidence threshold λ .

V. EXAMPLE

In this section, a simple example is given to illustrate the proposed Modified cluster-based fuzzy-genetic mining algorithm. Assume there are ten items in a transaction database: milk, bread, cookies, beverage, chocolate, icecream, coldrink, curd, fruit, butter. Since the item *milk* has three possible linguistic terms, *Low*, *Middle* and *High*, the membership functions for *milk* are thus encoded as (2, 11, 13, 11, 24, 11) for chromosome C2.

1. Randomly generate chromosomes

C0: 1 5 6 5 11 5, 1 10 11 10 21 10, 1 8 9 8 17 8, 0 6 6 6 12 6, 0 3 3 3 6 3, 0 3 3 3 6 3, 2 7 9 7 16 7, 2 3 5 3 8 3, 0 4 4 4 8 4, 1 3 4 3 7 3
C1: 1 9 10 9 19 9, 0 7 7 7 14 7, 2 11 13 11 24 11, 1 4 5 4 9 4, 2 3 5 3 8 3, 1 3 4 3 7 3, 2 4 6 4 10 4, 0 6 6 6 12 6, 1 3 4 3 7 3, 0 7 7 7 14 7
C2: 2 11 13 11 24 11, 2 14 16 14 30 14, 1 8 9 8 17 8, 1 5 6 5 11 5, 0 5 5 5 10 5, 0 2 2 2 4 2, 0 3 3 3 6 3, 1 4 5 4 9 4, 0 4 4 4 8 4, 0 7 7 7 14 7
C3: 2 11 13 11 24 11, 2 5 7 5 12 5, 1 9 10 9 19 9, 1 5 6 5 11 5, 0 3 3 3 6 3, 1 2 3 2 5 2, 0 5 5 5 10 5, 0 7 7 7 14 7, 1 4 5 4 9 4, 0 5 5 5 10 5
C4: 0 10 10 10 20 10, 2 12 14 12 26 12, 0 8 8 8 16 8, 1 7 8 7 15 7, 0 4 4 4 8 4, 0 3 3 3 6 3, 2 5 7 5 12 5, 0 4 4 4 8 4, 1 2 3 2 5 2, 1 6 7 6 13 6
C5: 1 7 8 7 15 7, 0 5 5 5 10 5, 2 4 6 4 10 4, 2 10 12 10 22 10, 0 4 4 4 8 4, 0 2 2 2 4 2, 1 7 8 7 15 7, 1 7 8 7 15 7, 1 4 5 4 9 4, 0 5 5 5 10 5
C6: 2 3 5 3 8 3, 4 13 17 13 30 13, 1 9 10 9 19 9, 0 6 6 6 12 6, 1 6 7 6 13 6, 1 3 4 3 7 3, 1 7 8 7 15 7, 0 3 3 3 6 3, 1 3 4 3 7 3, 0 7 7 7 14 7
C7: 2 7 9 7 16 7, 0 14 14 14 28 14, 0 5 5 5 10 5, 0 11 11 11 22 11, 2 3 5 3 8 3, 1 3 4 3 7 3, 1 3 4 3 7 3, 1 4 5 4 9 4, 0 3 3 3 6 3, 0 8 8 8 16 8
C8: 1 4 5 4 9 4, 0 12 12 12 24 12, 2 6 8 6 14 6, 1 9 10 9 19 9, 2 5 7 5 12 5, 0 3 3 3 6 3, 1 4 5 4 9 4, 2 4 6 4 10 4, 1 4 5 4 9 4, 0 7 7 7 14 7
C9: 0 5 5 5 10 5, 2 10 12 10 22 10, 2 11 13 11 24 11, 2 11 13 11 24 11, 1 5 6 5 11 5, 0 2 2 2 4 2, 1 7 8 7 15 7, 2 3 5 3 8 3, 1 3 4 3 7 3, 0 5 5 5 10 5

MILK

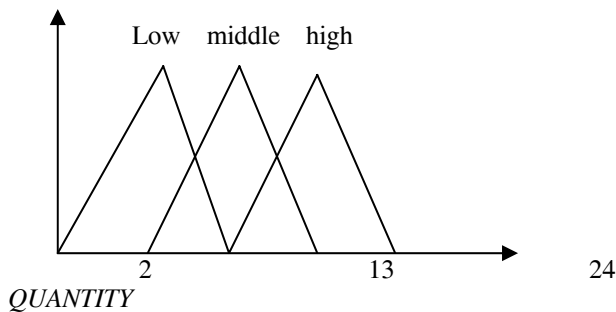


Fig No.6: The membership functions for milk in C2

2. Calculation of overlap factor , coverage factor and both

The overlap factor in suitable(Cq) is designed for avoiding the first bad case, and the coverage factor is for the second one.[2]

The dimensions used for clustering are overlap factor , coverage factor and average of both. Therefore, the calculation of minimum distance is as follows

$$\sqrt{(x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2}$$

Table No.1 : Calculation of CF & OF & Average

CHROMOSOME	OVERLAP FACTOR	COVERAGE FACTOR	AVERAGE OF OF & CF
C0	0.0	0.9478684	1.4369881
C1	0.9478684	0.0	0.48911962
C2	1.4369881	0.48911962	0.0
C3	0.7182525	0.22961593	0.7187355
C4	1.6091698	0.66130143	0.17218181
C5	0.66150403	0.28636444	0.775484
C6	0.69528997	0.25257847	0.7416981
C7	0.996007	0.048138615	0.440981
C8	1.2431476	0.29527926	0.19384035
C9	1.3238536	0.37598512	0.113134496

3. Clustering Chromosomes using 3-dimensional clustering

K=3

Chromosome no 0 In Cluster No. 0
 Chromosome no 1 In Cluster No. 1
 Chromosome no 2 In Cluster No. 2
 Chromosome no 3 In Cluster No. 1
 Chromosome no 4 In Cluster No. 2
 Chromosome no 5 In Cluster No. 1
 Chromosome no 6 In Cluster No. 1
 Chromosome no 7 In Cluster No. 1
 Chromosome no 8 In Cluster No. 2
 Chromosome no 9 In Cluster No. 2

4. Suitability of Chromosomes

The suitability factor used in the fitness function can reduce the occurrence of the two bad kinds of membership functions shown in Fig. 3, where the first one is too redundant, and the second one is too separate.

Table No.2 : Suitability of Chromosomes

Chromosome No.	Suitability
0	9.663095
1	8.815296
2	8.377814
3	9.020671
4	8.22381
5	9.071428
6	9.041209
7	8.77224
8	8.55119
9	8.479005

5. Fitness of Chromosomes

The fitness of each set of membership functions is evaluated by the number of large 1-itemsets generated by executing part of the previously proposed fuzzy mining algorithm. Using the number of large 1-itemsets can achieve a trade-off between execution time and rule interestingness. Usually, a larger number of 1-itemsets will result in a larger number of all itemsets with a higher probability, which will thus usually imply more interesting association rules. The evaluation by 1-itemsets is, however, faster than that by all itemsets or interesting association rules.[2]

Table No. 3: Fitness of Chromosomes

Chromosome No.	Fitness
0	2.897622
1	2.8359797
2	3.3421605
3	2.7714126
4	3.4047477
5	2.7559056
6	2.765117
7	2.8498993
8	3.2743979
9	3.3022742

6. Multilevel Association Rule and Confidence

IF MILKL AND BREADL THEN
COOKIESL 30.0 %
IF MILKL AND BREADL AND COOKIESL
THEN ChoclatesL 40.0 %

VI. COMPARISON OF THE EXISTING AND PROPOSED SYSTEM

For clustering the parameter k was taken as 3 and comparison was done between the k-means algorithm and the Modified k-means clustering and we got the following results.

Table No. 4: Comparison of clusters formed

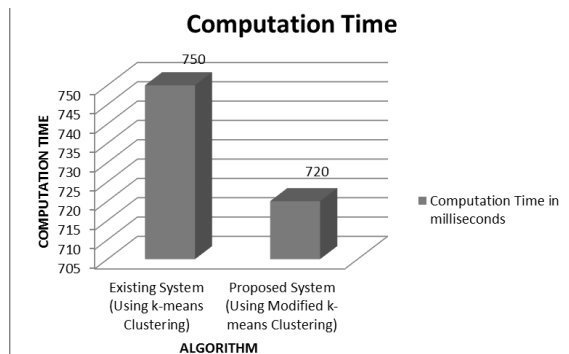
CLUSTERS FORMED BY 2-DIMENSIONAL K-MEANS ALGORITHM	CLUSTERS FORMED BY MODIFIED K-MEANS CLUSTERING
Chromosome no 0 In Cluster No. 0	Chromosome no 0 In Cluster No. 0
Chromosome no 1 In Cluster No. 1	Chromosome no 1 In Cluster No. 1
Chromosome no 2 In Cluster No. 2	Chromosome no 2 In Cluster No. 2
Chromosome no 3 In Cluster No. 1	Chromosome no 3 In Cluster No. 1
Chromosome no 4 In Cluster No. 2	Chromosome no 4 In Cluster No. 1
Chromosome no 5 In Cluster No. 1	Chromosome no 5 In Cluster No. 2
Chromosome no 6 In Cluster No. 1	Chromosome no 6 In Cluster No. 1
Chromosome no 7 In Cluster No. 1	Chromosome no 7 In Cluster No. 0
Chromosome no 8 In Cluster No. 2	Chromosome no 8 In Cluster No. 1
Chromosome no 9 In Cluster No. 2	Chromosome no 9 In Cluster No. 0

In order to check performance of the Association algorithms, we have applied the algorithm to item dataset. Comparison is done by considering the Computation Time.

Table No.5 : Comparison of Computation Time

SR NO.	ALGORITHM	COMPUTATION TIME
1	Existing system (using k-means clustering)	750 milliseconds
2	Proposed system (using modified k-means clustering)	720 milliseconds

Following graph depicts comparison of Computational time of existing system and proposed system with respect to Computational time for database of 10 items and 20 transactions. X axis in Graph denotes algorithm and y axis denotes Computational time. Computational time is measured in terms of milliseconds.



Graph 1 . Comparison of Algorithm using K-means and Modified K-means Clustering

CONCLUSION

This project presents an optimization method developed to deal with a data mining problem. Decision making in business sector is considered as one of the critical tasks. The objective is to provide a tool to help experts to find associations between the items bought by a customer in supermarket. An association rule may be more interesting if it reveals relationship among some useful concepts such as quantity of items bought by customer rather than which item is bought by the customer that is by using fuzzy mining. The favorite things of customers change all the time that is the membership functions should be adjusted dynamically, which can be done by genetic algorithm. To deal with all these aspects the fuzzy genetic association rule mining using modified k-means clustering is used. To achieve this, the algorithm has been developed in two parts. The first part finds out fit membership functions using genetic algorithm and modified k-means algorithm. The second part finds association rule (second level) using these fit membership functions and the confidence of that rule. This algorithm is tested on 10 items and 20 transactions.

REFERENCES

- [1] Chun-Hao Chen, Tzung-Pei Hong, "Cluster-Based Evaluation in Fuzzy-Genetic Data Mining", IEEE transactions on fuzzy systems, Vol. 16, No. 1, February 2008 249, pp. 249-262
- [2] T. P. Hong, C. H. Chen, Y. L. Wu, and Y. C. Lee, "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions," *Soft Computing*, vol. 10, no. 11, pp. 1091–1101, 2006.
- [3] Hung-Pin Chiu, Yi-Tsung Tang, "A Cluster-Based Mining Approach for Mining Fuzzy Association Rules in Two Databases", *Electronic Commerce Studies*, Vol. 4, No.1, Spring 2006, Page 57-74.
- [4] Tzung-Pei Hong, Chan-Sheng Kuo, Sheng-Chai Chi, "Trade-Off Between computation time and number of rules for fuzzy mining from quantitative data", *International Journal of Uncertainty, Fuzziness and Knowledge-Based systems*, Vol. 9, No. 5, 2001, page 587- 604.
- [5] M. Sulaiman Khan, Maybin Mueyba, Frans Coenen, "Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework", *The University of Liverpool, Department of Computer Science, Liverpool, UK*.
- [6] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems*, Vol. 13, No. 4, pp. 428-435, August 2005.
- [7] Miguel Delgado, Nicolás Marín, Daniel Sánchez, and María-Amparo Vila, "Fuzzy Association Rules: General Model and Applications", *IEEE transactions on fuzzy systems*, vol. 11, no. 2, April 2003
- [8] Tzung-Pei Hong, Li-Huei Tseng and Been-Chian Chien, "Learning Fuzzy Rules from Incomplete Quantitative Data by Rough Sets",
- [9] H. J. Zimmermann, "Fuzzy set theory and its applications", *Kluwer Academic Publisher*, Boston,

A SURVEY PAPER ON MANAGEMENT OF RISK FOR DIFFERENT MODELS OF SOFTWARE ENGINEERING

S. H. Bhosale

Department of Computer Engineering
A.C. Patil College of Engineering
Kharghar, Navi Mumbai, India
shbhosale@acpce.ac.in

M. M. Deshpande

Department of Information Technology
A.C. Patil College of Engineering
Kharghar, Navi Mumbai, India
mmdeshpande@acpce.ac.in

ABSTRACT

The challenges and realities in applying effective software risk management processes are difficult. Risk management is perhaps the most important part of any software development project. The intent of this paper is to get a glance of approaches to manage the risk for different models of software engineering. We have studied different techniques useful to handle the risks for various models, though some approaches are similar for a particular phase of different models. This study presents approaches to manage risks for every phase of the most commonly used models such as Waterfall model, Incremental model, Spiral model, Win-Win Spiral model, Agile model and Extreme Programming model.

Keywords-Risk Management, software project, model

I. INTRODUCTION

Risk plays a critical role in the development of a software project. The process of risk management is designed to reduce or eliminate the risk of certain kinds of event happening or having an impact on the software project. There are various models provided for the development of software projects. Each process has its own methods of developing the software.

While developing a software using Waterfall model, the risk factor is very high in case of large projects. This is because the requirement gathering is done only at the beginning of the phases and the customer is unable to communicate exact information for building the project. As a result the project is developed using vague ideas or inputs from the customer. Also, the customer can see the product only at the end. In case the customer is not satisfied with the product, it becomes very difficult to make those changes to the software. This leads to increase in the cost of the project as well as delay in its delivery [1].

Due to such problems (risk estimation is not performed on timely basis) 15-35% software projects are cancelled outright. Remaining suffer from schedule slippage, cost overruns or failure to meet the project goals [2]. Effective risk management is the most important

management tool that a project manager can employ to increase the likelihood of project success.

Spiral model and Incremental model are two risk driven models which help to identify the probable risk items that may occur at the beginning stage of software development. This helps to eliminate the factors that should be taken into consideration so as to avoid, if not entirely eliminate the possible risks as the project development progresses. A risk, if identified in timely manner will help to reduce the after effects it may cause once the project is delivered to the end user.

The objective of risk management is to identify, address and eliminate risk items before they become either threats to successful software operations or major sources of software rework. There is uncertainty in estimating the probability of losses associated with the unsatisfactory outcome, as assessments are often subjective and are product of surveying several domain experts.

Thus, risk exposure (RE) is a combination or product of the probability of unsatisfactory outcome (PO) and the loss to the parties affected if the outcome is unsatisfactory (UO) [4]. Unsatisfactory outcomes can be defined differently for the developers, users and the maintainers. For the customer and developers, unsatisfactory outcomes are budget overruns and schedule slippage while for the end users it is products with wrong functionality, wrong interface, performance shortfalls, reliability, etc. According to the maintainers, poor quality of software is an unsatisfactory outcome.

This paper is organized as follows: In section II we discuss about the risks involved with the various phases of waterfall model and the approaches to overcome the same. The phases of Incremental models along with its risks and techniques to reduce the risks are discussed in section III. Spiral model with the key risk factors and ways to eliminate them are described in section IV. In section V the Win- Win Spiral model is discussed with its risk and approaches to eliminate them. Section VI presents the risks involved with Extreme Programming while section VII describes Agile model with its risks and ways to overcome them.

II. WATERFALL MODEL

The waterfall model is a linear sequential model. It consists of the following distinct phases: Requirement

gathering (information engineering), analysis, design implementation and testing. Every stage must begin only after the former one is finished [3][4][5].

A. Risks for various stages of waterfall model

1. Requirement gathering and analysis

- Communication between humans.
- Misunderstanding or misinterpretation of information.
- Continuous requirement changes are difficult to accommodate due to the sequential nature of the model.

2. Design

- As it is the starting phase of the model, complex design complicates the next /latter phases. Due to this, the next phase will be harder to be done.

3. Implementation

- Different customs of different programmers makes it difficult to understand each other's code or logic.
- If a member leaves a team, they have to wait for another member to begin the work.
- Due to its sequential nature and relatively long stages, the working product can be seen late in the development phase.
- It becomes difficult for the developer to adapt to different tools and environment.

4. Testing

- It is the most dangerous part of the model. As there is no overlapping of stages, team members have to wait for the previous stages to be completed.
- This results in waste of time, cost and resources. It affects the schedule and deadline of the project.
- If the leader is unable to make the right decision, the whole group will be misled.
- Lack of quality assurance is conducted only during testing phase.

B. Approaches to overcome the risk

1. Requirement gathering and analysis

- Communicate extensively with the customer or the end user.
- Be more interactive
- Ask simple questions so that the customer is able to provide clear/well defined requirements in advance.
- Sign contract after the requirements are confirmed by the customer so that he will not be able to add extra functions to the software later on [3].

2. Design

- Keep the design as simple as possible.

- Give more emphasis on the details of the design [3].
- Design a well defined modular structure.
- Concentrate on available resources.

3. Implementation

- Keep the employee/programmer with similar skills/customs.
- Try to retain the existing members.
- Concentrate on available resources and tools.
- Have experienced members working on the project.

4. Testing

- Try to resolve risks at earlier stages.
- Test every corner in the program.
- Do not let the customer find the bug.

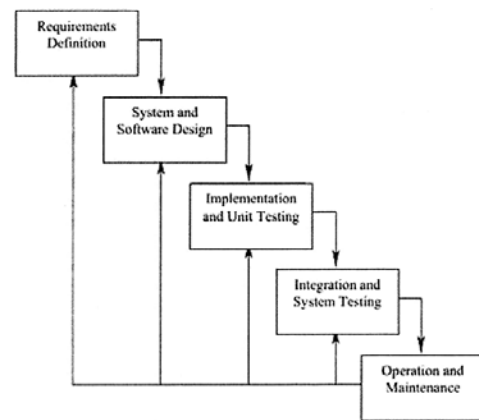


Figure 1: Waterfall Model [3]

III. INCREMENTAL MODEL

It consists of a series of waterfall lifecycles wherein the software development project is broken down into smaller segments called increments. It reduces the risk of gathering clearly defined requirements well in advance by grouping requirements, then implementing each group in an increment repeatedly until the system is complete and all the requirements are met [4,6].

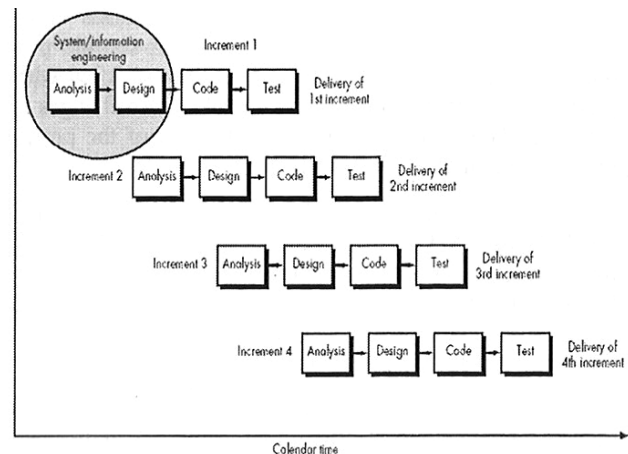


Figure 2: Incremental Model [13]

A. Risks for different phases of Incremental Model**1. Requirement gathering**

- Developers tend to postpone requirements so that they are included later on in the subsequent increments.
- This may take the system in a different or undesired direction.

2. Analysis

- Inadequate estimation of time, cost and resource requirements for each increment affects the overall project under development [3].
- It delays implementation of subsequent increments and thus results in unmet project deadlines.

3. Design

- Complex design complicates the next stages.

4. Coding

- Unique customs of individual programmer makes it difficult to understand other's logic.
- If a member leaves the team, they have to wait for another member to begin the work.
- It becomes difficult for the developer to adapt to various tools and environment.

5. Testing and maintenance

- Poor testing and maintenance conducted at the end of each increment allows any undiscovered error or bug in one increment to propagate through subsequent increments.

B. Approaches to manage the risks**1. Requirement gathering**

- Resolve all the identified requirements in the initial increments of the system and leave the rest (newly identified or changes in previously defined ones).

2. Analysis

- The developer will have an extra load to complete the project in order to meet the project's deadline.
- At times some of the requirements have to be ignored to deliver the software on time.

3. Design

- Keep the design simple.
- Give more attention to the details.
- Design modular structure for individual module.
- Concentrate on the available resources.

4. Coding

- Try to keep the developers with similar domain knowledge.
- Try to retain the existing developers.

5. Testing and maintenance

- Try to resolve risks at earlier stages only so that they do not affect the consequent phases.

IV. SPIRAL MODEL

The spiral model is aimed at determining the risks and trying to minimize them. It is an expansion of waterfall model because of the Risk Analysis phase. The addition of this phase makes it simple to identify risks and assess them thus reducing the chances of risk occurring at later stages [3][4][7].

A. Various risks approached during its Lifecycle**1. Planning**

- Requirements are gathered by communicating with the customer and risks are assessed.

2. Risk Analysis

- The project manager relies on the experience of the developers and their abilities in identifying and managing the risks.
- If the required skills are not available, it becomes difficult to rectify major risks in the project.

3. Design

- Complex design creates a problem to the consequent stages.

4. Engineering

- The software is produced along with testing at the end of this phase.
- Programmers with various skills find it difficult to understand other's logic.

5. Evaluation

- The customer evaluates the output of the software to date and gives his feedback.

B. Ways to overcome the risks of Spiral Model**1. Planning**

- The communication between the developer and the customers should be strong enough.
- The customer should be able to express his idea behind building the software in a precise manner.

2. Risk Analysis

- The analyst should be an expert in that particular domain.
- Analyst should be technical expert and have knowledge about the domain so that he can identify the risk items during analysis. This helps to eliminate risks in later stages. The success of the project is highly dependent on this phase.

3. Design

- Keep the design simple.
- Give more emphasis on the available resources.

4. Engineering

- Try to retain existing employee/programmer.
- Have members with similar skills to work on the same project.
- Concentrate on available resources and tools.

5. Evaluation

- The working model is provided to the customer at intervals so that he can run and validate the result. The customer is able to specify changes, if any so that they can be included in the next spiral.

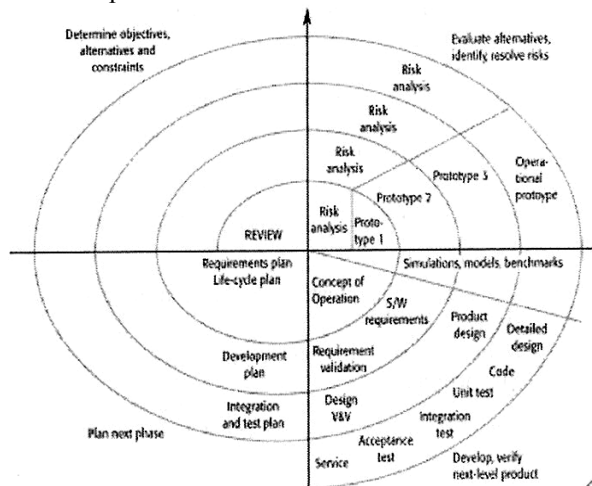


Figure 3 Spiral Model [4]

V. WIN-WIN SPIRAL MODEL

Win-Win Spiral model uses W Theory approach. Spiral model lacks well defined process guidance in determining the system's objectives, constraints and alternatives. Win-Win Spiral model identifies the system's stakeholders and their win conditions. It uses negotiation processes to determine mutually satisfactory set of objectives, constraints and alternatives for the stakeholders [4, 9].

Stakeholders are nothing but the project manager's subordinates and managers, customers, users, maintainers, significantly affected people such as developers or users of the product [8, 9, 10, and 11].

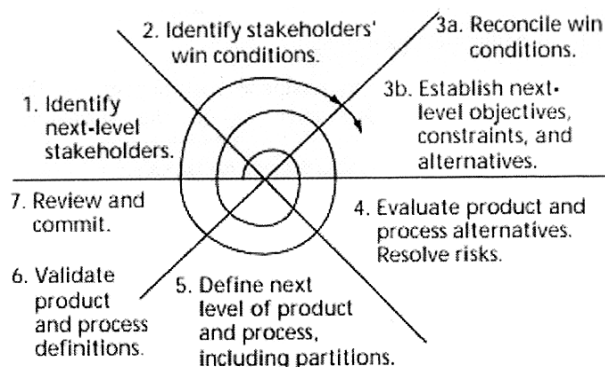


Figure 4 Win-Win Spiral Model [11]

A. Approaches to handle risks for this model

1. Establish a set of win-win preconditions [10]

- Understand how people want to win.
- Establish reasonable expectations.
- Match people's tasks to their win conditions.
- Provide a supportive environment.

2. Structure a win-win software process

- Establish a realistic process plan.
- Use the plan to control the project.
- Identify and manage your win-lose or lose-lose risks.
- Involve people during the entire life cycle of the model.

3. Structure a win-win software product

- Match products to user's, maintainer's win conditions.

VI. EXTREME PROGRAMMING

Extreme Programming (XP) is a developing method for projects where no specific demands are set up for the application. It is based on the values of simplicity, communication, feedback and courage. It works by bringing the whole team together in the presence of simple practices, with enough feedback to enable the team to see where they are and to tune to the practices to their unique station.

A. Risks encountered during development of the software using XP [4]

1. Team Size

- One of the risks that may occur when a project is developed with XP is that the project team becomes too large (more than 10 people).
- When this happens, the informal communication structure may evolve into miscommunication within the project.

2. Programming Experience

- Conceptually, though XP can be managed by different developers with various programming experience, the opposite is true.
- Many handlings and agreements are made informal during the development of the project.
- As there is not much control, it is necessary that the developers know what they are doing.

B. Steps to manage the above risks

1. Team Size

- By keeping the size of the team small (2-10 people), XP can offer a developing environment with informal communication and some project structure, wherever necessary.

2. Programming Experience

- Allow paired-programming for less experienced programmer.

- With paired-programming, the developer with less experience can be coupled to one with more experience.
- The experienced developer can be seen as a mentor that controls the project result of the developer with less experience allowing him to gain experience on XP.

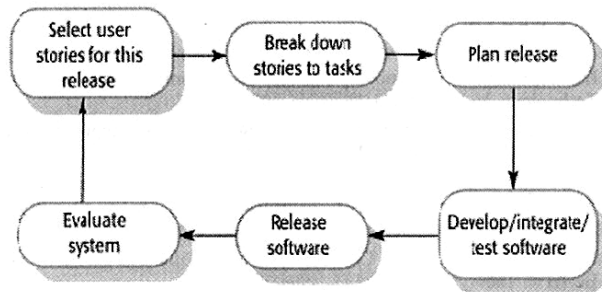


Figure 5: Extreme Programming [4]

VII. AGILE MODEL

Agile model is a lightweight methodology which begins the development of the software project without waiting for formal requirement and design specifications. It provides informal communication between different system stakeholders [4] [12].

It requires minimal planning. Hence it is optimal for those projects that lack structured planning [4].

A. Risks encountered during the development of the software

1. Size of software system

- In agile model, the developers do not wait for precise requirements from the customer. Rather they begin the development process with minimal planning.
- This results in relatively large number of increments thus increasing the span between the increments.
- For large sized projects, their cost increases as timely changes have to be done as per the customer's specifications or in case a bug is discovered.

2. Team Size

- One of the risks that may occur when a project is developed using agile model is that the project team becomes too large (more than 10 people).
- When this happens, the informal communication structure may evolve into miscommunication within the project.

3. Reliability on human

- It is extensively dependent on the experience of its developers along with their skills to interact with the customers.
- In case the developer lacks communication skills, it becomes a threat to the project.

4. Customer representative

- Lack of appropriate customer representative leads to the risk of project schedule as well as cost.

5. Distributed development environment

- This model is not suitable for distributed environment.
- It requires the developers to interact with the customers on timely basis to collect information for building the software.

6. Scope of the project

- As the development process does not wait for requirements to be acquired from the customer, developer are carried away from the main goal of the project.
- This leads to relatively large increments to satisfy the customer's requirements.
- The cost of the project increases.

B. Approaches to eliminate the above risks

1. Size of software system

- Projects which provide the software in small releases should be simple and small in size.
- Test cases can be easily performed on smaller releases thus implementing successful software.

2. Team Size

- By keeping the size of the team small (2-10 people), agile model can offer a developing environment with informal communication and some project structure, wherever necessary.

3. Reliability on human

- The developers working on the project should have expert knowledge about the domain.
- This leads to successful project development.

4. Customer representative

- The customer should be able to provide precise, well defined idea of the requirements to build the project.
- If the interaction between the developer and the representative is fruitful, it will lead to successful completion of the project.

5. Distributed development environment

- As it requires face to face interaction between the customer and developer, another mode of communication such as video conferencing should be held on daily basis.

6. Scope of the project

- During interaction between the customer and development team, they should be able to acquire clear requirements from the customer.
- This will help to keep the project small and simple.

CONCLUSION

The main contribution of this paper is to study different risk management techniques with respect to individual phase for each model of software engineering. After analysis of all the techniques, we found that each model is best suited for a particular domain or application area. Also, it depends on the size of the project.

Waterfall model is the most widely used methodology for development of a software project due to its sequential nature. It works well for small sized projects as the amount of risk involved is less.

Incremental model allows risks to be managed easily as risk items are identified and handled during intermediate iterations. It generates a working software quickly and early during the software lifecycle. It is less costly to change scope and requirement and hence more flexible.

The spiral model has an added advantage of a risk analyst who identifies the risk items at a very early stage in its lifecycle. This reduces the chance of any adverse effects on the software at later stages. The risk analyst has immense knowledge of the particular domain or application area on which the project is built. It is suitable for large, complex projects.

Win-Win Spiral model involves extensive collaboration of the stakeholders. Due to this the software is produced at a rapid speed. The cost of the software is less because of the rework. This also results in maintenance reduction.

Extreme Programming and Agile Model work well with minimum resources as well as less requirements. It involves team work and more suitable for small sized projects.

Thus, taking into consideration the approach of individual model to identify and assess risk in a project, we should make use of that model which best suits our domain or application.

REFERENCES

- [1] B. Boehm, "Software Risk Management: Principles and Practices", IEEE Software 8(1), pp. 32-41, January 1991.
- [2] Yu-Heng Wang, Jian Ji and Ying Qu, "The Earth-Moon Model on software project risk management", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010.
- [3] N. Munassar and A. Govardhan, "A Comparison between Five Models of Software Engineering," International Journal of Computer Science Issues (IJCSI), Vol. 7, No. 5, 2010, pp. 94-101.
- [4] H. Hijazi, T.Khdour and A. Alarabeyyat, "A Review of Risk Management in Different Software Development Methodologies", International Journal of Computer Applications (0975 – 8887) Volume 45– No.7, May 2012.
- [5] W. Royce, "Managing the development of large software systems," IEEE WESCON, 1970, pp. 1-9.
- [6] G. Tate and J. Verner, "Case Study of Risk Management, Incremental Development and Evolutionary Prototyping," Information and Software Technology, Vol. 32, No. 3, 1990, pp. 207-214.
- [7] B. Boehm, "A Spiral Model of Software Development and Enhancement," Computer, 1988, pp. 61-72.
- [8] B. Boehm, A. Egyed, D. Port, A. Shah, J. Kwan and R. Madachy, "A stakeholder win-win approach to software engineering education", Annals of Software Engineering 6 (1998) 295-321 295.
- [9] B. Boehm and P. Bose, "A Collaborative Spiral Software Process Model Based on Theory W", August 11, 1994
- [10] B..W. Boehm and R. Ross, "Theory W Software Project Management: Principles and Examples," IEEE Trans. Software Engr., Vol. 15, No. 7, July 1989.
- [11] B..W. Boehm and R. Madachy, "Using the Win Win Spiral Model: A Case Study", Computer IEEE, July 1998, pp. 33-44 .
- [12] V. Szalvay, "An Introduction to Agile Software Development," technical report, Danube Technology, 2004.
- [13] Roger S. Pressman, "Software Engineering :A Practitioner's Approach" , Fifth Edition, McGraw-Hill, ISBN 0-07-365578-3, 2004.

A BAYESIAN ITERATIVE RECONSTRUCTION WITH COMPUTED TOMOGRAPHY FOR MAXIMUM CONVERGENCE OF SCATTERED DATA

S. Asif Hussain
Associate Professor
Department of ECE
A.I.T.S, Rajampet ,A.P, INDIA
sah.ssk@gmail.com

Dr. M. N. Giri Prasad
Professor
Department of ECE
JNTUA ,A.P, INDIA
giriprasad@gmail.com

C. Venkatesh
Assistant Professor
Department of ECE
A.I.T.S, Rajampet ,A.P, INDIA
venky.cc@gmail.com

ABSTRACT

Medical Diagnostic devices play a vital role in imaging for easy diagnosis and therapy. Basically there are different types of image reconstruction such as iterative reconstruction and filtered back projection for working with higher radiation of CT images, but in terms of noise and flexibility iterative reconstruction are far better with fewer measurements. In this work iterative reconstruction is considered with expectation maximization and total variation methods with Poisson noise model for superior quality results. Expectation maximization is applied to observe visually in accordance with root mean square error between original and reconstructed images after rescaling to [0,255]. Similarly, Total variation method is extremely effective for recovering edges. The objective is to combine Expectation maximization and total variation (EM+TV) method with fewer projections to obtain good quality image which results in appropriate amount of CT radiation dose. Reconstructions and computer simulations with patient data shows that proposed work gives superior results for recovering the edges by smoothening the noise and sharp edges.

Keywords- *Iterative reconstruction; Poisson Noise; Smoothening; Filtered back projection.*

I. INTRODUCTION

Image reconstruction in CT is a mathematical process that generates images from X-ray projection data acquired at many different angles around the patient. Image reconstruction has a fundamental impact on image quality and therefore on radiation dose.

For a given radiation dose it is desirable to reconstruct images with the lowest possible noise without sacrificing image accuracy and spatial resolution. Reconstructions that improve image quality can be translated into a reduction of radiation dose because images of acceptable quality can be reconstructed at lower dose. Two major categories of methods exist, analytical reconstruction and iterative reconstruction.

Methods based on filtered back projection (FBP) are one type of analytical reconstruction that is currently widely used on clinical CT scanners because of their computational efficiency and numerical stability.

Many FBP-based methods have been developed for different generations of CT data-acquisition geometries, from axial parallel- and fan-beam CT in the 1970s and 1980s to current multi-slice helical CT and cone-beam CT with large area detectors. For a general introduction of the fundamental principles of CT image reconstruction.

Users of clinical CT scanners usually have very limited control over the inner workings of the reconstruction method and are confined principally to adjusting various parameters specific to different clinical applications. The reconstruction kernel, also referred to as “filter” or “algorithm” by some CT vendors, is one of the most important parameters that affect the image quality. Generally speaking, there is a tradeoffs between spatial resolution and noise for each kernel. A smooth kernel generates images with lower noise but with reduced spatial resolution. A sharp kernel generates images with higher spatial resolution, but increases the image noise.

Iterative reconstruction has recently received much attention in CT because it has many advantages compared with conventional FBP techniques. Important physical factors including focal spot and detector geometry, photon statistics, X-ray beam spectrum, and scattering can be more accurately incorporated into iterative reconstruction, yielding lower image noise and higher spatial resolution compared with FBP. In addition, iterative reconstruction can reduce image artifacts such as beam hardening, windmill, and metal artifacts. A recent clinical study on an early version of iterative reconstruction demonstrated a potential dose reduction of up to 65% (6) compared with FBP-based reconstruction algorithms.

Due to the intrinsic difference in data handling between FBP and iterative reconstruction, images from iterative reconstruction may have a different appearance (e.g., noise texture) from those using FBP reconstruction. Careful clinical evaluation and reconstruction parameter optimization will be required before iterative reconstruction can be accepted into mainstream clinical practice.

High computation load has always been the greatest challenge for iterative reconstruction and has impeded its use in clinical CT imaging. Software and hardware methods are being investigated to accelerate iterative reconstruction. With further advances in computational technology, iterative reconstruction may be incorporated into routine clinical practice in the future.

Types of Image Reconstruction:

Traditional algebraic methods are further categorized into Algebraic reconstruction technique (ART), Simultaneous iterative reconstruction technique (SIRT) and Iterative least-squares technique (ILST). Statistical reconstruction methods are also further categorized into Gradient and Conjugate gradient (CG), Maximum likelihood expectation maximization (MLEM), Ordered-subsets expectation maximization (OSEM), Maximum a posteriori (MAP). Iterative reconstruction has several merits over traditional FBP techniques and hence CT has attracted considerable attention in recent times. An attractive field normally in digital image processing techniques and particularly in biomedical imaging is 3D image reconstruction. The development of fast algorithms, like ordered-subset, expectation-maximization algorithm has partially solved the high computational cost of iterative 3D image reconstruction which is its major demerit.

A CT scan is carried out by using a special X-ray machine, which produces an image of a cross-section, or slice, of the body. The scanner consists of a 'doughnut-shaped' structure, or gantry, about two feet thick with a hole in its centre, through which you pass while lying on a couch.

A narrow fan-shaped beam of X-rays is produced from inside the gantry and rotates in a complete circle around you. The X-rays pass through your body and are detected by electronic sensors on the other side of the gantry. This information passes to a computer which produces a picture of the internal structure of the body. The pictures are displayed on a screen for examination by the radiologist.

To reconstruct the image from the X-ray measurements, each voxel must be viewed from multiple different directions. A complete data set requires many projections at rotational intervals of 1° or less around the cross section. Back-projection effectively reverses the attenuation process by adding the attenuation value

of each X-ray in each projection back through the reconstruction image. This requires a significant computer power to quickly generate the patient image. Because this process initially generates a blurred image, the data from each projection are mathematically altered (filtered) prior to back-projection to eliminate the blurring.

II. PROBLEM STATEMENT

To reconstruct a better image using fewer views in the computed tomography setting, thus reducing the overall dose of radiation. The numerical results in two and three dimensions show the efficiency of the proposed EM+TV method by comparison with those obtained by filtered back projection (FBP) or by EM only. Some of the terms and physical descriptions refer directly to X-ray computed tomography. Figure 1: Parallel beam geometry. Each projection is made up of the set of line integrals through the object.

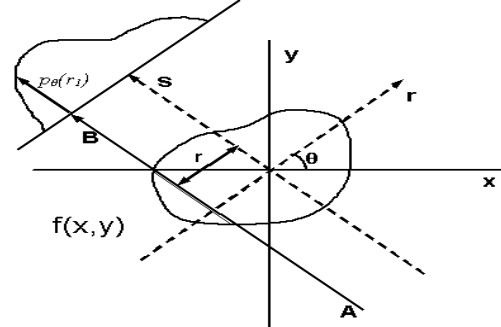


Fig2.1 shows the projections of Parallel geometry

The projection of an object at a given angle θ is made up of a set of line integrals. In X-ray CT, the line integral represents the total attenuation of the beam of x-rays as it travels in a straight line through the object. As mentioned above, the resulting image is a 2D (or 3D) model of the attenuation coefficient. That is, we wish to find the image $\mu(x, y)$. The simplest and easiest way to visualize method of scanning is the system of parallel projection, as used in the first scanners.

For this discussion we consider the data to be collected as a series of parallel rays, at position r , across a projection at angle θ . This is repeated for various angles. Attenuation occurs exponentially in tissue.

A. Existing system

Multi detector row computed tomography (MDCT) creates massive amounts of data, which can overload a picture archiving and communication system (PACS). To solve this problem, we designed a new data storage and image interpretation system in an existing PACS. Two MDCT image datasets, a thick- and a thin-section dataset, and a single-detector CT thick-section dataset were reconstructed.

Thin-section datasets and MPR images were used in 79% and 18% of cases, respectively. The radiologists' assessments of this system were useful, though the volume of images and times taken to archive, retrieve, and interpret thick-section images together with thin-section images were significantly greater than the times taken without thin-section images. The limitations were compensated for by the usefulness of thin-section images. This data storage and image interpretation system improves the storage and availability of the thin-section datasets of MDCT and can prevent overloading problems in an existing PACS for the moment.

Drawbacks

1. Only acquire data for one slice image at a time.
2. Expense, the high dose of radiation it delivers.
3. The laziness it promotes, and the havoc it can wreak when misinterpreted.

B. Proposed system

Proposed EM+TV Reconstruction Method:

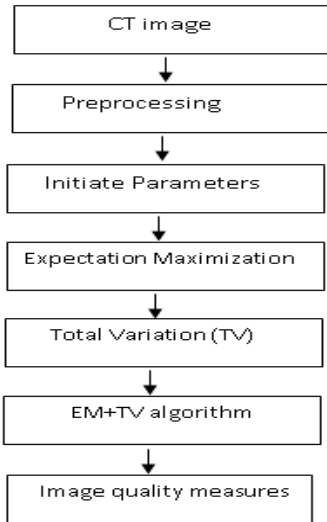


Figure 1. Flow chart of proposed work

The procedure for this method is presented in Fig.1. First of all; the gray color image is off set so that the gray level values are greater than 0. Then the image histogram is created. It is a chart that shows the distribution of intensities in a grayscale image. The input parameters are initiated.

1. Each point is given a probability (weight) to belong to each class.
The E step:
2. The probabilities of each point are updated according to the distances to the centers.
The M step:
3. Class centers are computed as a weighted average overall data points.
4. Final
E step: classify each point to the nearest (most probable) center.

III. EM+TV RECONSTRUCTION METHOD

The objective is to reconstruct an image with both minimal total-variation and maximal probability, given fewer noisy projections in the sinogram domain. So we can consider finding a Pareto optimal point by solving a scalarization of these two objective functions and the problem is to solve

$$\begin{cases} \underset{x}{\text{minimize}} & \int_{\Omega} |\nabla x| + \alpha \sum_{i=1}^M ((Ax)_i - b_i \log(Ax)_i), \\ \text{subject to} & x_j \geq 0, \quad j = 1, \dots, N, \end{cases} \quad (3)$$

with $\alpha > 0$ a tuning parameter. This is a convex constrained optimization problem and we can find the optimal solution by solving the Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} -\text{div}\left(\frac{\nabla x}{|\nabla x|}\right)_j + \alpha \sum_{i=1}^M \left(a_{ij} \left(1 - \frac{b_i}{(Ax)_i}\right)\right) - y_j &= 0, \quad j = 1, \dots, N, \\ y_j \geq 0, \quad x_j \geq 0, \quad j &= 1, \dots, N, \\ y^T x &= 0. \end{aligned}$$

By positivity of $\{x_j\}$, $\{y_j\}$ and the complementary slackness condition $y^T x = 0$, we have $x_j y_j = 0$ for every $j = 1 \dots N$. Thus, if we multiply the partial differential equation by x_j at pixel j , we obtain

$$-x_j \text{div}\left(\frac{\nabla x}{|\nabla x|}\right)_j + \alpha \sum_{i=1}^M \left(a_{ij} \left(1 - \frac{b_i}{(Ax)_i}\right)\right) x_j = 0, \quad j = 1, \dots, N,$$

or equivalently

To solve the above EM+TV minimization problem, we use an iterative semi-implicit finite-differences scheme. Each iteration is called a TV step. Given x EM, already computed from the EM step, we compute the new x discarding (5) (as a discrete minimize of (6)), by the following simple iterative scheme, presented below in two spatial dimensions and assuming now that x is represented as a 2D matrix: from an initial guess x_0 , we compute x_{n+1} with $n \geq 0$ from the following liberalized discrete equation,

Description

Given a statistical model consisting of a set X of observed data, a set of unobserved latent data or missing values Z , and a vector of unknown parameters θ , along with a likelihood function $L(\theta; X, Z) = p(X, Z | \theta)$, the maximum likelihood estimate (MLE) of the unknown parameters is determined by the marginal likelihood of the observed data

$$L(\theta; X) = p(X | \theta) = \sum_Z p(X, Z | \theta)$$

however, this quantity is often intractable.

The EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying the following two steps:

Expectation step (E step): Calculate the expected value of the log likelihood function, with respect to the conditional distribution of Z given X under the current estimate of the parameters $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = E_{Z|X, \theta^{(t)}} [\log L(\theta; X, Z)]$$

Maximization step (M step): Find the parameter that maximizes this quantity:

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$$

Note that in typical models to which EM is applied:

The observed data points X may be discrete (taking values in a finite or countably infinite set) or continuous (taking values in an uncountably infinite set). There may in fact be a vector of observations associated with each data point.

The missing values (aka latent variables) Z is discrete, drawn from a fixed number of values, and there is one latent variable per observed data point.

The parameters are continuous, and are of two kinds: Parameters that are associated with all data points, and parameters associated with a particular value of a latent variable (i.e. associated with all data points whose corresponding latent variable has a particular value).

A. About total variation regulated method EM algorithm

An EM algorithm for estimation :

To ensure convergence, parameter estimates are obtained via an EM algorithm alternating expectation and maximization. The penalized log-likelihood l is first expressed as a complete-data log-likelihood $l_C = l_{\xi} + l_{\eta_1} + l_{\eta_2}$, where

$$l_{\xi} = \sum_{i=1}^m \sum_{j=1}^{n_j} (z_{ij} \xi_{ij} - \log[1 + \exp(\xi_{ij})])$$

$$\xi_{ij} = \log(p / (1 - p)),$$

$$l_{\eta_1} = \sum_{i=1}^m \sum_{j=1}^{n_j} z_{ij} \log f_1(y_{ij}) - \frac{1}{2} \left(m \log(2\pi\sigma_{u_1}^2) + \frac{1}{\sigma_{u_1}^2} u_1^T u_1 \right),$$

$$l_{\eta_2} = \sum_{i=1}^m \sum_{j=1}^{n_j} (1 - z_{ij}) \log f_2(y_{ij}) - \frac{1}{2} \left(m \log(2\pi\sigma_{u_2}^2) + \frac{1}{\sigma_{u_2}^2} u_2^T u_2 \right).$$

Here, z_{ij} is an *unobserved* variable indicating whether y_{ij} comes from the first component ($z_{ij} = 1$) or the

second component ($z_{ij} = 0$). At each iteration, the EM algorithm proceeds with the E-step, in which z_{ij} is replaced by its conditional expectation $z_{ij}^{(r)}$ under the current parameter estimates $\{\hat{p}^{(r)}, \hat{\beta}_1^{(r)}, \hat{\beta}_2^{(r)}\}$ and $\{\hat{u}_1^{(r)}, \hat{u}_2^{(r)}\}$:

$$z_{ij}^{(r)} = \frac{\hat{p}^{(r)} \hat{f}_1(y_{ij})}{\hat{p}^{(r)} \hat{f}_1(y_{ij}) + (1 - \hat{p}^{(r)}) \hat{f}_2(y_{ij})}.$$

At the M-step, updated estimates $\hat{p}^{(r+1)}$, $\{\hat{\beta}_1^{(r+1)}, \hat{u}_1^{(r+1)}\}$ and $\{\hat{\beta}_2^{(r+1)}, \hat{u}_2^{(r+1)}\}$ can be obtained by maximizing l_{ξ} , l_{η_1} , l_{η_2} separately, in view of the orthogonal partition of l_C into three components. The M-step can be implemented by solving the following three sets of recursive equations based on the derivatives of l_{η_1} and l_{η_2} ,

$$\hat{p}^{(r+1)} = \sum_{i=1}^m \sum_{j=1}^{n_j} z_{ij}^{(r)} / n,$$

$$\begin{bmatrix} \hat{\beta}_k \\ \hat{u}_k \end{bmatrix} = \begin{bmatrix} \beta_{k,0} \\ u_{k,0} \end{bmatrix} + \mathfrak{S}_{\beta_k, u_k}^{-1} \begin{bmatrix} \frac{\partial l_{\eta_k}}{\partial \beta_k} \\ \frac{\partial l_{\eta_k}}{\partial u_k} \end{bmatrix}, \quad k = 1, 2.$$

Note that $\beta_{k,0}$ and $u_{k,0}$ are initial values of the corresponding quantities, which are then replaced by their updated estimates at each iteration, for given values of $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$. The matrix $\mathfrak{S}_{\beta_k, u_k}$ represents the negative second derivative of l_{η_k} with respect to $\{\beta_k, u_k\}$, $k = 1, 2$, formula of which is given in Appendix A.1. The shape parameter v_k is estimated by

maximizing $l_{\eta_k} = \sum_{i=1}^m \sum_{j=1}^{n_j} z_{ij} \log f_k(y_{ij})$ using updated estimates at the end of the M-step.

A step-by-step initialization procedure is adopted. Firstly, the two-component gamma mixture model without any covariate is fitted to the data, using a wide range of starting values. The estimates corresponding to the largest likelihood value are subsequently used as initial estimates for fitting the two-component gamma mixture regression model without any random effects. The results so obtained are then set as the initial estimates for the final mixture model with random effects.

At each cycle of the EM algorithm, it is assumed that the variance components $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ are given. In

practice, they are unknown and needed to be estimated. When convergence is attained, $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ are estimated by the most updated values $\hat{u}_1, \hat{u}_2$ and the corresponding elements of the information matrix. The process continues until all parameter estimates converge.

Convergence is determined when the absolute changes of estimates between two iterations are all less than 0.001. For the estimation of variance components, the REMQL approach is adopted to correct the bias due to maximum likelihood estimation. Suppose the observed information matrix $\mathfrak{I}_{p, \beta_1, \beta_2, u_1, u_2}$ is partitioned conformally to $p | \beta_1 | \beta_2 | u_1 | u_2$ as

$$\mathfrak{I}_{p, \beta_1, \beta_2, u_1, u_2}^{-1} = [A_{ij}], \quad i, j = 1, \dots, 5,$$

then estimators of variance components are given by:

$$\hat{\sigma}_{u_1}^2 = m^{-1} [\text{trace}(A_{44}) + \hat{u}_1^T \hat{u}_1],$$

$$\hat{\sigma}_{u_2}^2 = m^{-1} [\text{trace}(A_{55}) + \hat{u}_2^T \hat{u}_2].$$

Finally, asymptotic standard errors of the regression coefficients are obtained from the corresponding partition.

B. Back projection

Involves the reassigning of the convolved scan data to a 2D image matrix representing the section of the patient that is scanned. This is performed profile for profile for the entire image reconstruction process. The image matrix can be thought of as analogous to a chessboard, consisting of typically 512x512 or 1024x1024 picture elements, usually called "pixels".

Back projection results in an exact density being assigned to each of these pixels, which are then displayed as a lighter or darker shade of gray. The lighter the shade of gray the higher the density of the tissue within the pixel (e.g. bone).

An essential step in image reconstruction is back projection, which is the adjoint to forward Projection process that forms the projections of the object. Figure 3.3 shows the back projection along a fixed angle, $\emptyset$. Conceptually, back projection can be described as placing a value of $p(s, \emptyset)$ back into an image array along the appropriate LOR, but, since the knowledge of where the values came from was lost in the projection step, the best we can do is place a constant value into all elements along the LOR.

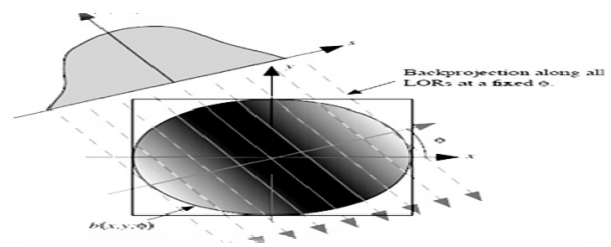


Figure 3.2: Back projection $b(x, y, \emptyset)$, into an image reconstruction array of all values $P(S, \emptyset)$ for a fixed value of $\emptyset$.

One might assume that straight back projection of all the collected projections will return the Image, but this is not the case due to the oversampling in the center of the Fourier transform. In Other words, each projection fills in one slice of the Fourier space resulting in oversampling in the center and less sampling at the edges. For example, if we perform back projections at only two angles, say $\emptyset_1$ and $\emptyset_2$ and examine the Fourier transform of the result we see that the contribution at the origin is doubled while there is only one contribution at the edges of the field of view.

Another way of understanding this oversampling in the space domain is with the forward projection of a single point source. If we simply back project the point source projections, the image would be heavily blurred since the projections are added back to the entire LOR from which they came. The oversampling needs to be re-weighted, or 'filtered', in order to have equal contributions throughout the field of view.

IV. REMOVAL OF ARTIFACTS & IMAGE INTENSITY MEASURES

Removal of artifacts and image intensity measurements: Technically-caused artifacts

After motion and physics, a third influential factor, the technology used for MR, is discussed. This group of artifacts can be explained by using technical limits, such as the size of the system or the limitation of the data volume generated.

Types of artifacts

We differentiate between the following artifacts related to technology:

- Truncation artifacts
- Wrap-around artifacts
- Distortion artifacts
- Artifacts caused by RF irradiation
- Truncation artifacts
- Truncation artifacts are shown as stripes or rings in the image.

Tuncation artifacts and data sampling :

Abrupt signal transitions in tissue may lead to artifacts caused by truncation. Period oscillations parallel to tissue interfaces are generated. Stripes or rings appear in the image with alternating high and low signal intensity (edge oscillations).

An object would be perfectly imaged, if an infinite data acquisition window were available. Through the limited time period available for the measurement, these are, however, interrupted at certain locations or not continued. Usually this does not have a negative affect on the MR image. Only highly contrast-enhanced tissue interfaces show artifacts caused by truncation.

Assessment of image resolution and noise quantization of the artifacts

A. Factors affecting the assessment of image resolution and noise quantization of the artifacts

Noise is measured. For each initialization method analyzed, the set of reconstructions obtained consisted of six images, each calculated from a different noise realization of the projection data. For each possible pair of reconstructions from such a set, the images forming the pair were subtracted. Then, the standard deviation in a region covering the liver was computed for each of the resulting difference images:

$$SD_i = \sqrt{\frac{\sum_k^N \tilde{\mu}_i^d(k)^2}{N-1}}$$

Where N is the total number of voxels in the region of interest in the i-th difference image. Finally, the standard deviations were averaged over all difference images to produce the final numerical estimate of the noise.

B. Image quality metrics

Image quality metrics are Figurers of merit used for the evaluation of imaging systems or processes. The image quality metrics can be broadly classified into two categories, subjective and objective. Subjective image quality is a method of evaluation of images by the viewers and it emphatically examines fidelity and at the same time considers image intelligibility. In objective measures of image quality metrics, some statistical indices are calculated to indicate the reconstructed image quality. The image quality metrics provide some measure of closeness between two digital images by exploiting the differences in the statistical distribution of pixel values. The most commonly used error metrics used for comparing compression are Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR).

The factors affecting the assessment of the image resolution are:

C. Computational Measures:

1) Mean Square Error (MSE):

Mean square error is the sum over all squared value differences divided by image size. It's a measure between the original image and the reconstructed image

$$MSR = \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N [(I(x,y) - I'(x,y))]^2$$

Where I(x,y) is the original image, I'(x,y) is the reconstructed image, and M,N are the dimensions of the images.

2) Peak signal to noise ratio (PSNR):

It is a measure of the peak error

$$PSNR = 10 \log \left[\frac{255}{\sqrt{MSE}} \right]^2$$

A lower value for MSE means lesser error, and as seen from the inverse relation between the MSE and PSNR, this translates to a high value of PSNR. Logically, a higher value of PSNR is good because it means that the ratio of Signal to Noise is higher. Here, the 'signal' is the original image, and the 'noise' is the error in reconstruction

3) Normalized Cross-Correlation (NCC):

Normalized Cross-Correlation is one of the methods used for template matching, a process used for finding incidences of a pattern or object within an image

NCC=

$$\sum_{j=1}^M \sum_{k=1}^N (I(X,Y) - I'(X,Y))^2 / \sum_{j=1}^M \sum_{k=1}^N (I(X,Y))^2$$

4) Structural Content (SC):

Structural content establishes the degree to which an image in the collection matches. It's the measure of image similarity based on small regions of the images containing significant low level structural information. The more the number of such regions common to both images, the more similar they are considered.

SC=

$$\frac{\sum_{j=1}^M \sum_{k=1}^N (I(X,Y))^2}{\sum_{j=1}^M \sum_{k=1}^N I'(X,Y)^2}$$

5) Maximum Difference (MD):

It's the variation of the method of paired comparisons.

$$MD = \text{Max} (|X_{j,K} - X_{j',K}| K)$$

6) Normalized Absolute Error (NAE):

It's the numerical difference between the original and reconstructed image.

NAE=

$$\sum_{j=1}^M \sum_{k=1}^N |X_{j,K} - X_{j',K}| / \sum_{j=1}^M \sum_{k=1}^N |X_{j,K}|$$

V. EXPERIMENTAL RESULTS



Figure: The phantom image

The above image describes the CT image of the phantom which has been captured by the CT machine and this image is given as the input to the next processes for the generation of the sinogram image and for the process of the back projection

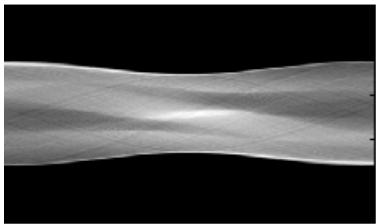


Figure: the sinogram image

This sinogram shows the data which is plotted between the sensor positions and the radon transform of the object, at various angles in the original CT machine

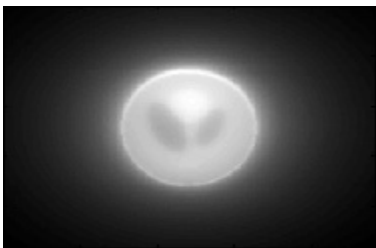


Figure 1. The image formed through the simple back projection

This figure shows the simple projection on the obtained sinogram image for the total back projection process to take place we need to acquire the total image, the simple projection is not sufficient

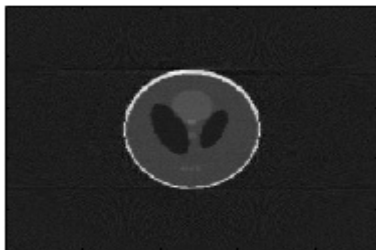


Figure: the reconstructed image formed through the expectation maximization and total variation algorithm

The above figure displays the total image which had been reconstructed by the process of the filtered back projection. The above results shows the reconstruction of the phantom image using the filtered back projection algorithm. As the real task of the project is to perform the operations on the CT images of living things. The

following are the results obtained when the operations are performed on the certain parts of the human body.

S No	parameters	values
1	Name	Medical phantom generated through EM &TV
2	Size	189 x 139
3	Pixels	26271
4	Minimum	0 @ 103 , 1
5	Maximum	254.7 @ 75 , 2
6	Back ground	55
7	average	64.804634
8	Sum	1702482.5342
9	Standard deviation	37.433054

CONCLUSION

The proposed method is combination of expectation maximization and total variation minimization for image reconstruction in computerized tomography, in the presence of Poisson noise. This method provides comparable results when using very few views, comparing to filtered back projection with many more views; also, the proposed method provided much improved results without artifacts, by comparison to those obtained by the proposed method needs much fewer measurements to obtain a good quality image, which results in the decrease of the radiation dose. The method has been easily extended to three dimensions. Future work includes faster implementation using the advantage of graphics processing units (GPUs), parallel computing, and applications to real data.

REFERENCES

- [1] F. Natterer, *The Mathematics of Computerized Tomography*. New York: Wiley, 1986.
- [2] D. C. Munson, J. D. O'Brien, and J. W. K., "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proceedings of the IEEE*, vol. 71, pp. 917–925, August 1983.
- [3] S. Basu and Y. Bresler, "O(N² log² N) filtered backprojection reconstruction algorithm for tomography," *IEEE Trans. on Image Processing*, vol. 9, pp. 1760–1772, October 2000.
- [4] G. T. Herman, *Image reconstruction from projections*. New York: Academic Press, 1980.
- [5] Kalender, W. A., "X-ray computed tomography", *Physics in Medicine and Biology*, vol. 51, no.13, pp. 29-43, 2006.

- [6] Kalender, W. A., Seissler, W., Klotz, E., Vock, P., "Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation", Radiology, vol. 176, no.1, pp.181-183, 1970.
- [7] Katsevich, A., "Theoretically Exact Filtered Backprojection-Type Inversion Algorithm For Spiral CT", SIAM Journal on Applied Mathematics, vol. 62, no. 6, pp. 2012-2026, 2002.
- [8] Katsevich, A., "An improved exact filtered backprojection algorithm for spiral computed tomography", Advances in Applied Mathematics, vol. 32, no. 4, pp. 681-697, 2004.
- [9] Danielsson, P. E., Edholm, P., Eriksson, J., Magnusson, M., "Towards exact reconstruction for helical cone-beam scanning of long objects. A new detector arrangement and a new completeness condition", In Proceedings of the Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, pp. 141-144, 1997.
- [10] Yang, J., Kong, Q., Zhou, T., Jiang, M., (2004), "Cone beam cover method: An approach to performing backprojection in Katsevich's exact algorithm for spiral cone beam CT", Journal of X-Ray Science and Technology, vol. 12, no. 4, pp. 199-214, 2004.
- [11] Yang, J., Guo, X., Kong, Q., Zhou, T., Jiang, M., "Parallel Implementation of Katsevich's FBP Algorithm. International Journal of Biomedical Imaging", vol. 2006, pp. 1-8, 2006.
- [12] Michael, G., "X-ray computed tomography", Physical Education, vol. 36, no. 6, pp. 442-451, 2001.
- [13] S. Asif Hussain and Dr. M.N. Giri Prasad, Dr. D. Satya Narayana "3d Medical image modeling with dynamic image reconstruction "in IJERT Vol 5, No.3 (2012) PP. 213-217
- [14] Michael, G., "X-ray computed tomography", Physical Education, vol. 36, no. 6, pp. 442-451, 2001.
- [15] Hiriyanaiyah, H. P., "X-ray computed Tomography for Medical Imaging", IEEE Signal Processing Magazine, vol. 14, no. 2, pp. 42-59, 1997.

COMPARATIVE ANALYSIS OF AUSCULTATION TECHNIQUE

Monali U More
E & TC Department
M.E-Electronics
V.P.C.O.E, Pune University
Baramati, India
monalimore12@gmail.com

Aparna R. Shastri
E & TC Department
M.E-Electronics
V.P.C.O.E, Pune University
Baramati, India.
aparnadjoshi@yahoo.com

Rajveer.K. Shastri
E & TC Department
M.E-Electronics
V.P.C.O.E, Pune University
Baramati, India.
rajveer_shastri@yahoo.com

ABSTRACT

Auscultation, listening to sounds coming from human organs, is a primary routine for screening and diagnosing many pathological conditions of the heart. Heart sounds-based methods of diagnosis are designed to discover and examine the relationships between heart sounds and heart disease using digital signal processing technologies. Most of the conventional methods for diagnosis using acoustic and vibration signals are primarily based on observing the amplitude differences in the time or frequency domain.

In this work, image patterns are obtained by using Short Time Fourier Transform (STFT), Wavelet Transform (WT), and Visual Dot Pattern (VDP) to identify the acoustic signals for heart sound analysis.

Experiments are carried out to evaluate the proposed system for heart sound analysis under various normal and abnormal conditions such as Aortic Stenosis, Mitral Stenosis, Aortic Regurgitation and Mitral Regurgitation. Principal Component Analysis (PCA) based on eigenspace system is used as classifier for image patterns obtained. Heart Sound signals patterns are successfully recognized as normal and abnormal heart sound with 73.33% accuracy of Short Time Fourier Transform, 80% accuracy of Wavelet pattern, and 83.33% accuracy of Visual Dot Pattern.

Keywords: Auscultation, Visual Dot Pattern (VDP), Short Time Fourier Transform (STFT), Wavelet Transform (WT), Principal Component Analysis (PCA).

I INTRODUCTION

The term Auscultation is simply the act of listening to the sounds made by internal organs, is a valuable medical diagnostic tool in diagnosis of heart sound. Auscultation methods provide the information about a vast variety of internal body sounds originated from the heart, lungs, bowel and vascular disorders. The ability to

sense heart sounds may be influenced by a number of factors, including the presence of ambient noise or other sensory stimuli. Therefore, only the skilled physician who is proficient in the skill of auscultation is likely to make accurate diagnoses upon cardiac auscultation. This may result in inaccurate or insufficient information due to the inability of the user [1]. It is, thus, desirable to enhance the diagnostic ability by processing the auscultation signals providing a visual display and automatic analysis to the physician for a better comparative study.

With the improvement of domestic living standard numbers of patients with heart disease are increasing rapidly. The heart disease is associated with living conditions, such as coronary heart disease (angina pectoris, myocardial infarction) and hypertension. It becomes more and more difficult to recognize and diagnose heart sound in traditional auscultation way, because of limitation of human hearing sensitivity and auscultator's clinical experience. How to find symptom and know the incidence status in advance, to prevent and diagnosis the heart disease [2].

Human heart generates four sounds S1, S2, S3 and S4 during its activity for one cardiac cycle which is not all audible. Fig 1 shows normal heart sounds S1, S2.

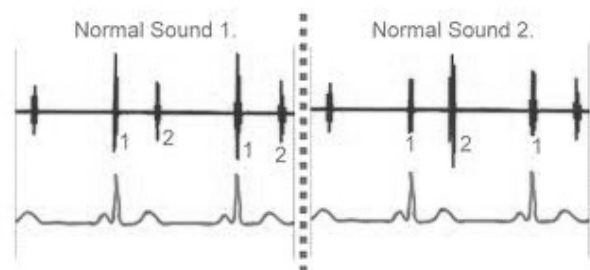


Fig 1: Normal heart sound S1, S2.

S1 is created by closure of the mitral and tricuspid valves. Contracting or pumping phase of chambers is called systole. The main feature of S1 is low tone and long time lasted. S2 created by the closing of the aortic and pulmonic valves as blood exits the heart to the body

and lungs. This is diastole. The resting or filling phase of chambers is diastole. In contrast with S1, it has characteristics of high tune and short time lasted. S1 and S2 contain important information of cardiac sounds auscultation [3]

In this work, image patterns are obtained by using Short Time Fourier Transform (STFT), Wavelet Transform (WT), and Visual Dot Pattern (VDP) to identify the acoustic signals for heart sound analysis. Experiments are carried out to evaluate the proposed system for heart sound analysis under various normal and abnormal conditions.

II HEART SOUND ANALYSIS METHODS

1) STFT: The STFT combines traditional time domain and frequency domain concepts into a single time-frequency framework. With the STFT one can characterize some time dependent frequency changes.

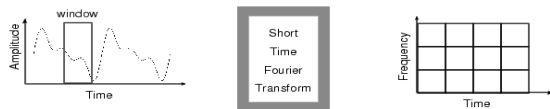


Fig 3 Short time Fourier transform

The STFT represents a sort of compromise between the time- and frequency-based views of a signal.

$$STFT_x(t, f) = \int [x(t) \cdot \omega * ((t - t'))] \cdot e^{-j2\pi f t} dt \quad (1)$$

It provides some information about both when and at what frequencies a signal event occurs [4]. However, you can only obtain this information with limited precision, and that precision is determined by the size of the window. While the STFT compromise between time and frequency information can be useful, the drawback is that once you choose a particular size for the time window, that window is the same for all frequencies. [5].

Fig 4 shows the simulated signal with sinusoidal frequencies $f_1=79.57\text{Hz}$ in duration 200-400 Sec and $f_2=7.95\text{Hz}$ in duration 600-800 Sec. Spectrogram of corresponding sine wave is as shown in Fig.5. in which colored part indicating amplitude of sinusoid with frequencies $f_1=79.57\text{Hz}$ Present in the duration 200-400 Sec and $f_2=7.95\text{Hz}$ in duration 600-800 Sec.

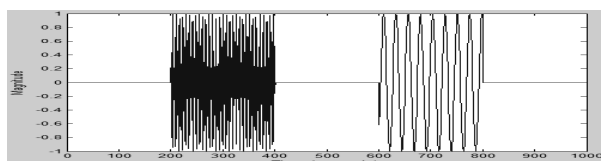


Fig 4 Sine wave of with frequencies $f_1=79.57\text{Hz}$ and $f_2=7.95\text{Hz}$

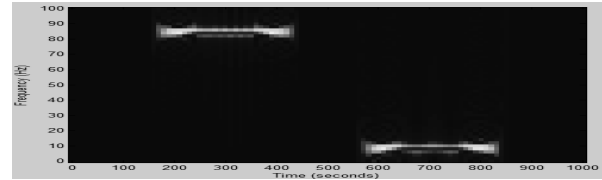


Fig 5 Spectrogram of sine wave

2) Wavelet Transform: A wavelet is a waveform of effectively limited duration that has an average value of zero. Compare wavelets with sine waves, which are the basis of Fourier analysis. Sinusoids do not have limited duration i.e. they extend from minus to plus infinity. And where sinusoids are smooth and predictable, wavelets tend to be irregular and asymmetric as shown in Fig 6. Fourier analysis consists of breaking up a signal into sine waves of various frequencies. Similarly, wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet[5].

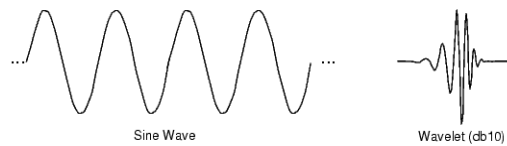
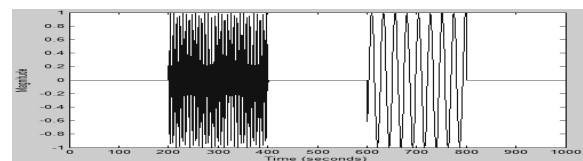


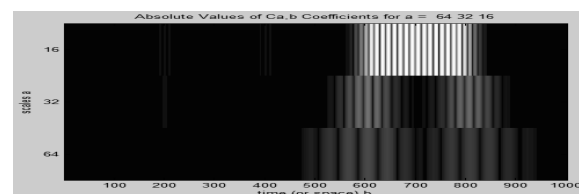
Fig 6 Sine wave and wavelet.

Wavelet analysis represents the next logical step: a windowing technique with variable-sized regions. Wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information.

Fig 7(a) shows the simulated signal with sinusoidal frequencies $f_1=79.57\text{Hz}$ in duration 200-400 Sec and $f_2=7.95\text{Hz}$ in duration 600-800 Sec.



(a)



(b)

Fig 7 (a) Sine wave of with frequencies $f_1=79.57\text{Hz}$ and $f_2=7.95\text{Hz}$, (b) Wavelet transform of sine wave.

III. VISUAL DOT PATTERN

Vibration signals generated by heart beat can be represented as superposition of sinusoidal signal. The

proposed visual dot pattern is methods for visually expressing the problem in easy to understand figure. It involves the transformation of sound pattern into a set of dots having mirror symmetry, in which the time waveform of sound signal is visualized as a snowflake-shaped pattern of six-fold symmetry [6].

VDP is a dot pattern obtained by transforming the correlation of value $X(t)$ of the time-axis signal at time t and its value $X(t+\tau)$ at time $t+\tau$ into polar coordinates $P(A(i), \theta(i), \phi(i))$ and drawing a plot. Fig 8 shows the technique for plotting VDP. One set of data ($X(t)$, $X(t+\tau)$) is plotted so as to be linearly symmetrical about the initial line of the polar coordinate graph. Then, a figure symmetrical about a point is created by repeatedly rotating the axis of linear symmetry and plotting. The transform from the discrete signal $x(i)$ at a fixed sampling time Δt to polar coordinates $P(i)$ is given by the following equations [7].

$$A(i) = \frac{(x(i) - x_{min})}{(x_{max} - x_{min})} \quad (2)$$

$$\Phi = \Theta + \frac{x(i+k) - x_{min}}{x_{max} - x_{min}} g \quad (3)$$

$$\Theta = \Theta - \frac{x(i+k) - x_{min}}{x_{max} - x_{min}} g \quad (4)$$

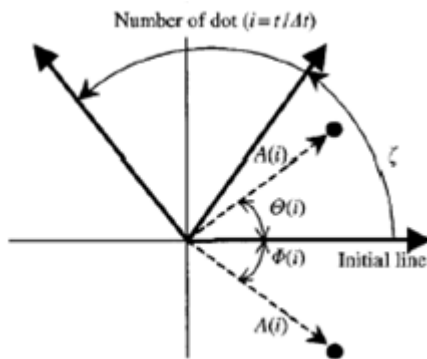


Fig 8 Principle of visual dot pattern

Where, i =Number of dots ($i = \frac{t}{\Delta t}$) with time ' t '

Δt =Sampling time

k =Time lag coefficient for plotting

N =Number of discrete signal

$A(i)$ =Radius of polar plot

$x(i) = t \cdot \Delta t$ sound signal

x_{max} = Maximum sound signal

x_{min} =Minimum sound signal

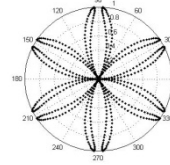
Θ = Rotation angle of initial line

g =Gain of plotting angle

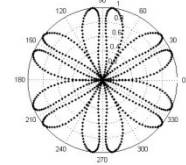
$\Phi(i)$ = Angle of dot in summarized dot pattern

$\Phi(i)$ =Negative angle of dot in summarized dot pattern

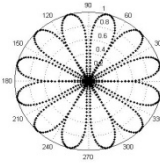
In plot coordinates $P(A(i), \Theta(i), \Phi(i))$, the time lag ' k ', and the gain ' g ' are taken as parameters, so, the discriminability of a fault with VDP depends on ' k ' and ' g '. Fig. 9 shows effect of frequency on VDP. Higher the frequency of signal more will be overlapping of dots making unable to discriminate between them.



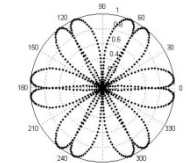
(a) $f=100\text{Hz}$



(b) $f=300\text{Hz}$



(c) $f=500\text{Hz}$



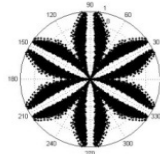
(d) $f=700\text{Hz}$

Fig 9: VDP pattern of sine wave with different frequency.

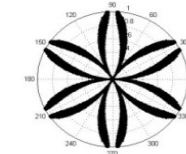
VDP of high discriminability having a wide contour is always obtained by setting k to $k=5$ and gain value $g = 20^\circ - 40^\circ$ so as to visualize the pattern clearly without overlapping. If gain is less than 25 or more than 29 VDP pattern becomes unclear.

The algorithm maps normalized time waveform into the visual dot space creating a scatter plot of neighbor amplitudes on polar graph. The point in time waveform is mapped into a radial component and adjacent point is mapped to angular component [7].

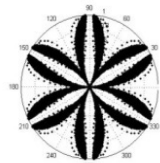
Fig.11 shows VDP patterns of normal heart sound with different beat rate



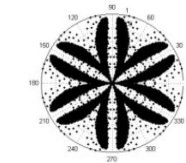
(a) Innocent



(b) Normal sinus rhythm
60bpm



(c) Normal sinus rhythm
90bpm



(d) Normal sinus rhythm
130bpm

Fig.11. VDP patterns of Normal heart sound with different beat rate.

Fig.12. shows the VDP patterns of abnormal HS namely AR (Aortic Regurgitation), AS (Aortic Stenosis), MR (Mitral Regurgitation), MS (Mitral Stenosis).

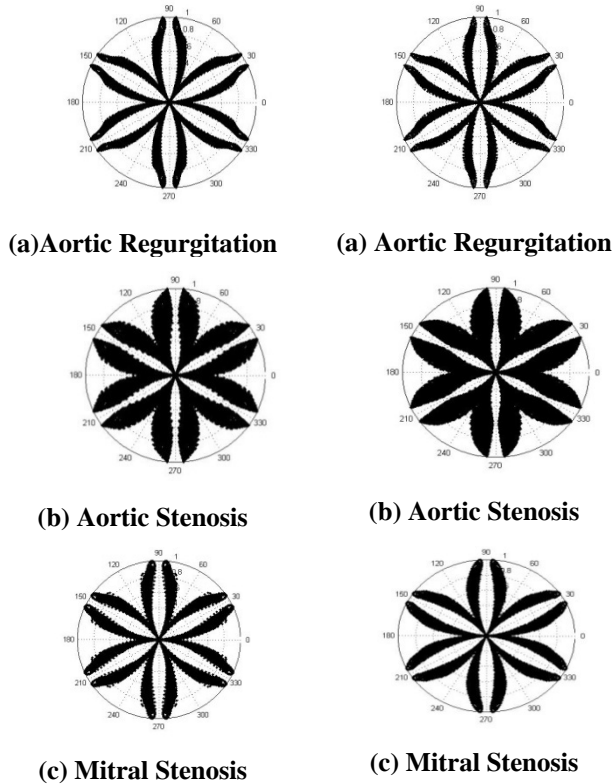


Fig 12 VDP patterns of abnormal HS (a) AR (b) AS (c) MS

Fig 11 & 12 shows VDP of normal and abnormal heart sound by which we can easily visualize the difference between them.

The input signal $X(t)$ is first normalized to the range through finding the high (x_{max}) and low (x_{min}) values for 'i' points of data in the window, so the overall amplitude is not a factor in the characterization. VDP functions somewhat like autocorrelator and is also sensitive to general frequency content and waveform variability [8]. These VDP are computed for sample of about two second duration.

IV. PCA CLASSIFIER

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal

components are guaranteed to be independent only if the data set is jointly normally distributed. Principal component analysis is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen-Loève transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD) [10].

The goal of Principal Component Analysis (PCA) is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original dataset. PCA is a technique that effectively and efficiently represents pictures of faces & recognizes the face from large dimension [10]

V. PCA ALGORITHM

This procedure consists of two phases: 1) training phase and 2) recognition phase. In training phase, the samples of data, on which the system needs to recognize, are used to create an Eigen Matrix which transforms the samples in the image space into the points in Eigenspace. The image samples are taken as grayscale images and are transformed from 2D matrix to 1D column vector of size $N^2 \times 1$ by placing the image matrix columns consecutively. These column vectors of n images are placed column wise to form the data matrix (image set) X of dimension $N^2 \times n$. And m be the mean vector of data vectors in matrix X given by

$$m = \frac{1}{n} \sum_{i=1}^n x^i \quad (5)$$

The vector of data matrix X are centered by Subtracting the mean vector m from all the column vectors of X to get the covariance matrix Q of the column vectors and is given by

$$Q = XX' \quad (6)$$

The eigenvalues and corresponding eigenvectors are computed for the covariance matrix

$$QV = \lambda V \quad (7)$$

Where V is the set of eigenvectors associated with the Eigenvalues λ . Set the order of the eigenvectors $v_i \in V$ according to their corresponding eigenvalues from high to low. This matrix of eigenvectors is eigenspace V . The data matrix X is projected onto the eigenspace to get P consisting n columns, where

$$P = V^T X \quad (8)$$

In the Recognition phase, the image I to be recognized is converted to 1D vector and form J as mentioned above which is projected on to same eigenspace to get Z .

$$Z = X^T I \quad (9)$$

And the Euclidean distance d between Z and all the projected samples in P is measured using L_2 norm (Euclidean distance) of an image A and image B is

$$L_2 = \sqrt{\sum (A_i - B_i)^2} \quad (10)$$

The projected test image is compared to every projected training image and the training image that is found to be closest to the test image is used to identify the training image [11].

VI. EXPERIMENTS AND RESULTS

This section presents the results of the experiments conducted to compare the performance of PCA and proposed method. The methods are implemented in MATLAB on grayscale images of fifteen different Heart Sound patterns obtained from STFT, WT and VDP. The image patterns obtained from STFT/WT/VDP for five different classes namely Aortic regurgitation, Aortic stenosis, Mitral Stenosis, Mitral regurgitation and Normal Heart Sound. Each class contains three image patterns. Table 1 indicates how the classification is done.

Table 1 Type of Classes

No of Classes	Name of Class
I	Aortic regurgitation
II	Aortic Stenosis
III	Mitral Stenosis
IV	Mitral regurgitation
V	Normal Heart Sound

This Fig 13 shows five classes, each consisting of three different samples. All image cycles taken as gray scale images and transformed 2D matrix to 1D column vector.

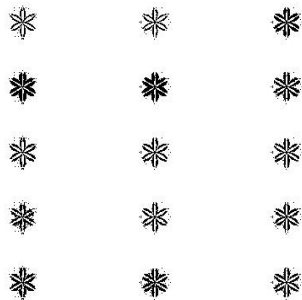


Fig 13 Database of PCA

Eigen values and corresponding Eigen vector are computed for the datamatrix. Then Eigen vectors are set according to their corresponding eigenvalues from high to low to get Eigen space. Fig 14 shows the plot of eigen values set in descending order.

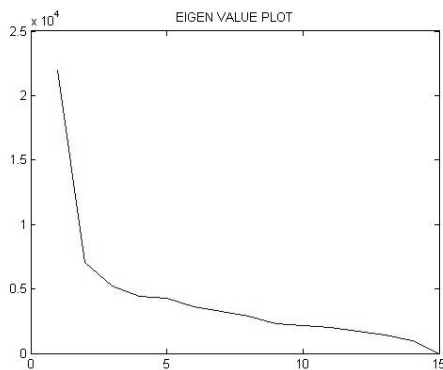


Fig 14 Eigen Value Plot

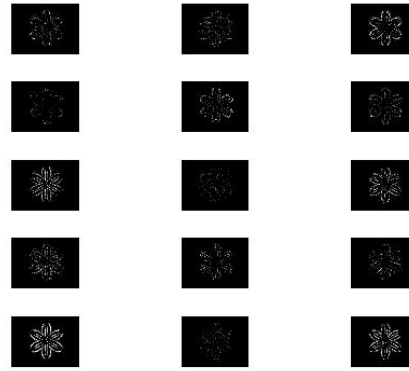


Fig 15 Eigen images

This matrix of eigenvectors is nothing but eigenspace. The data matrix is projected onto the eigenspace. Fig 15 shows eigen images indicating principal component.

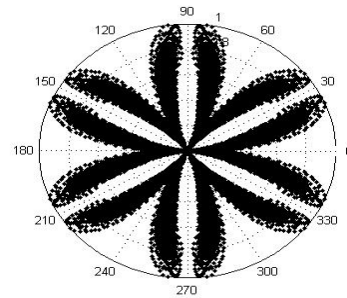


Fig 16 Visual Dot Pattern of Aortic stenosis

In recognition phase, any random image pattern from the dataset is considered as test image as shown in Fig 16 AS2_3 i.e. third sample from class two name as Aortic Stenosis. Euclidean distance of between this test image and all projected samples are calculated. Table 2 shows Euclidean distance between this test image and all projected images.

Table 2 Euclidian distance

CLASS	Stored Images	Euclidian distance
1	AR1_1	12813.671
	AR1_2	12569.5368
	AR1_3	4678.328527
2	AS2_1	8429.208762
	AS2_2	9834.658764
	AS2_3	9.42E-11
3	MS3_1	8381.621126
	MS3_2	8307.267163
	MS3_3	8827.616575
4	MR4_1	8468.854151
	MR4_2	8938.147179
	MR4_3	7265.517391
5	NR5_1	5302.510014
	NR5_2	6472.362705
	NR5_3	7707.893251

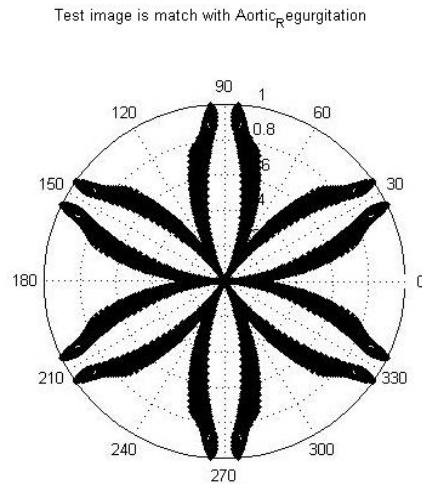


Fig 17 Output of PCA Classifier

Table 2 itself indicate that test image i.e. AS2_3 has minimum Euclidean distance. So result indicates test image belong to class-2 i.e. Aortic Stenosis as shown in Fig 17.

CONCLUSION

In work novel method called Visual Dot Pattern has been explained. This system uses fifteen heart sound divided into the five classes, each class contains three heart sounds. STFT, WT and VDP has been applied to all heart sounds to get corresponding image patterns at two sampling frequencies $F_s=44$ KHz and $F_s=8$ KHz. With these image patterns applied to PCA Classifier gives 73.33% accuracy with STFT, 80% accuracy with WT and 83.33% accuracy with VDP distinguishing heart sound as normal/abnormal Heart Sound as Aortic regurgitation, Aortic stenosis, Mitral Stenosis, Mitral regurgitation and Normal Heart Sound. Also, this method gives 100% accuracy for two diseases namely aortic stenosis and mitral stenosis. VDP requires less computational time compared to STFT and WT. Table-4 shows comparative result.

Table 4 Comparative Result

PCA Classifier	Fs=44Khz		Fs=8Khz	
	Accuracy	Time	Accuracy	Time
STFT	73.33%	4.91Sec	73.33%	4.45 Sec
WT	80%	3.65 Sec	93.33%	9.59 Sec
VDP	83.33%	2.65 Sec	93.33%	3.09 Sec

VDP is simply the graphical representation of signal. It is effective to analyze heart sound through visualization

of sound signals. As VDP is simply the picture of polar plot, pictures becomes rapidly identified, tabled, remembered & integrated into coherent whole. VDP can make differences obvious even to in experimental persons.

REFERENCES

- [1] "Ascultation of heart the Heart", *review of clinical signs. September 2002*, pp 39-43.
- [2] Wang Haibin, Hu Yuliang, Liu Lihan, Wang Yan, "Heart Sound Analysis based on Autoregressive Power Spectral Density Heart Sound Analysis based on Autoregressive Power Spectral Density" *ICSPS, 978-1-4244-6893-5/2010 IEEE*
- [3] <http://www.learntheheart.com/PDF2-heartsounds.pdf>
- [4] P. S.Vikhe, N.S.Nehe.Heart Sound Abnormality Detection Using Short Time Fourier Transform and Continuous Wavelet Transform. *ICETET, 978-0-7695-3884-6/09© 2009 IEEE*
- [5] MATLAB Documentation and Toolbox.
- [6] Clifford A. Pickover "Computers, Patterns, Chaos and beauty graphics from an unseen world"
- [7] Katsuhiko Shibata, Atsushi Takahashi, Takuya Shirai. Fault Diagnosis of Rotating Machinery through Visualization Of Sound Signals. *Mechanical Systems and Signal Processing (2000) 14(2), 229-241.*
- [8] Jian-Da Wu, Chao-Qin Chuang. "Fault diagnosis of internal combustion engines using visual dot patterns of acoustic and vibration signals. *NDT&E International 38 (2005) 605-614, Elseveir Ltd.*
- [9] Dr A Samson, Shabeer Hameed. "Diagnosis of Multiple Faults of Ball Bearings Using Visual Dot Pattern" *NCTT09, 6-7 Nov 2009*
- [10] Lindsay I Smith, A Tutorial on Principal Component Analysis, February 26, 2002.
- [11] Neeraja, Ekta, "Face Recognition Using Improved Fast PCA Algorithm" *CISP, 978-0-7695-3119-9/08 © 2008 IEEE.*

DIFFERENTIATED INSTRUCTION AND ITS IMPLICATION IN MOTIVATING L2 WRITING

Dr. Malavika Sharma
Department of Humanities, PIIT
New Panvel, Mumbai, India
doctor.malavika@yahoo.com

ABSTRACT

Motivating students to write effectively and develop their cognitive ability is indeed a challenge for language teachers. It is observed that there is deterioration in students' writing abilities and also their ability to think independently. In order to find solution to this problem, a model of teaching for the undergraduate students has been developed based on the concept of 'Differentiated Instruction' (Tomlinson, 2004). Results indicate that learners benefit in terms of noticing their shortcomings in the area of language, coherence of thought, and writing analytical paragraphs. With the help of this method the students were able to practice and try out new language forms and structures, involve themselves in the tasks, and develop their creative and critical thinking that led them to enhance their self-esteem. This article introduces the concept of differentiated instruction and highlights the research evidence for its effectiveness. It also proposes a method of application of the differentiated instruction in the classroom and the need to introduce differentiated assessment as a strategy to encourage every student to meet the same standards in different ways.

I INTRODUCTION

Motivating students to write effectively and to think on their own is a challenge for language teachers. The observation is that there is deterioration in students writing abilities and also their ability to think independently. The present scenario could be the result of two reasons.

The first one is that in the teacher –centered classroom there is hardly any involvement of the students. As such the needs of the learners are not taken into account. In such a situation the activities assigned to the learners appear as part of their routine work that needs to be fulfilled.

Secondly teachers have not paid interest to the fact that the increase in the number of students means the possibility of varied learning needs in addition to their diverse cultural and linguistic background.

This predominant method of instruction used by the teacher can be effective only with the students who are comfortable in learning with that method. The other students who do not learn well through that method will not be able to master the required skill. This is a problem faced not only by language teachers, but also by technical teachers.

II THE NEED FOR AN ALTERNATIVE METHOD

An alternative approach in teaching is therefore necessary in order to focus on the needs of the learners. According to Hobson (2004) teachers must deliver developmentally appropriate instruction while providing lessons that are stimulating, challenging and are differentiated to meet the varied learning needs of students. Researchers have reiterated the need to respond to the diversity existing in the classrooms (Fischer and Rose, 2001; Flem et al., 2000; McCoy and Ketterlin-Geller, 2004; Mulroy and Eddinger, 2003; Sizer, 1999; Tomlinson, 2001b, 2004a).

Moreover taking into consideration the differences and interests of the students will help to enhance their motivation to learn. Ignoring these fundamental differences may result in some students falling behind, losing motivation and failing to succeed. Carol Ann Tomlinson (2004) also suggests that the tasks designed by the teachers should be moderately challenging and interesting enough to arouse their curiosity. It should neither be too simple nor too complex.

In response to the realities of such a diverse student population, many educational researchers, and teacher training programs have made it a priority to equip teachers with a knowledge base for understanding and working with the differences that students bring to the classroom.

III THE CONCEPT OF THE DIFFERENTIATED INSTRUCTION MODEL

Differentiated Instruction is an approach that recognizes the needs and preferences of the students. It can be defined as the method where the teacher tries to address students particular readiness needs, their particular interests, and their preferred ways of learning. This concept is grounded in the work of Lev Vygotsky (1978). His work highlights the need of the individual to work in association with social activity which consequently leads to cognitive development. Cognition in this respect is to be seen as a social product which can be achieved through interaction (Zone of proximal development, Vygotsky, 1978.) Research conducted in this area by Fisher et al., (1980), strongly supports the ZPD concept.

The idea of differentiation is not confined to a single content area or a specific grade level. This method to classroom instruction is one that can be applied to any content area at any grade level. Tomlinson identifies three elements of the curriculum that can be differentiated: Content, Process, and Product.

A Content

In content the focus is on the concepts, principles and skills that students are expected to learn. The teacher adjusts his/her instruction to suit the diverse abilities of the learners keeping in mind that the same concepts are being addressed.

B Process

Here the learners are expected to interact and work together as they develop knowledge of new content. The teacher helps by imparting information about the content as part of the introductory session to the whole class. This is followed by dividing the class into small groups or pairs and assigning the tasks accordingly. Care is taken by the teacher to maintain flexibility in the formation of the groups.

C Product

In product the learners are given the opportunity to apply and extend their understanding of the content. Continuous assessment of the learners output informs the teacher about the different approaches, choices, interests and abilities and gives him/her an idea of scaffolding to meet the varying needs of the learners.

According to the proponents of differentiation many improvements have been noticed in the performances of the learners in terms of the attainment of accuracy, and development of their confidence level. McAdams (2001) reported significant improvement in the test scores of low-scoring students in the Rockwood School District (Missouri), following the use of differentiated instruction. Teachers in this study also indicated that their students were more motivated and enthusiastic about learning. Tony Erben, Ruth Ban and Martha Castaneda (2009) have in their book **Teaching English Language Learners through Technology** highlighted the importance of Differentiated Instruction and how it has helped teachers meet the learning needs of a heterogeneous group of students. This has been exemplified in their work where they have successfully demonstrated by using one computer in a class of 45 students and the positive outcome of such an effort. The one computer can be used in the classroom by teachers with the help of designing activities which are best known as tiered activities (Erben, Ban and Castaneda, 2009.)

Tiered activities or assignments are creating activities by making slight adjustments within the same lesson in order to enable students who are at different cognitive levels to work at their own pace. All students learn the same fundamental skills and concepts, but through varying modes and activities. The tier according to researchers challenges students at their ability levels. The teacher's challenge is to make sure all tasks,

regardless of the tier level, are interesting, engaging and challenging.

Apart from creating activities that meet the learning needs of students, author Carol Ann Tomlinson (2004) also gives importance to the learning environment as a key factor in the implementation of the differentiation model. She suggests that the learning environment is the "weather" of a classroom and includes the classroom's operation and tone. In her research she has found how class rules, furniture arrangement, lighting procedures, and processes all affect the classroom's mood. Learners' motivation is largely affected due to these factors.

IV DIFFERENTIATING ASSESSMENT AND ITS SIGNIFICANCE

Assessment is an integral part in the teaching and learning process. In a differentiated instruction classroom assessment helps the teacher maintain record to monitor the progress of students who are struggling and who may need more intensive support. This intense monitoring and documentation facilitates differentiation, because it enables teachers to keep track of students' progress and design differentiated lessons more accurately. Further, it also helps teachers from spending time on tasks that are boring, inappropriate, or excessively challenging.

The assessment policy implemented by a teacher in a differentiated classroom takes into consideration the basic principles of differentiation that is, varying the assessments in terms of learning difficulty and learning style. This enables the learner to consider assessment as a learning tool where they can measure their progress.

An effective and productive method of assessment in a differentiated classroom can be made possible through ongoing and continuous assessment as well as implementing different ways of assessment. In other words, students with producing difficulty can record their understanding of the content either by writing or speaking. In this way the teacher can achieve the required product by changing the mode of response for each student.

The teacher can also use formative assessment that is preparing portfolios for each individual student that are related to their individual needs, skills and progress. This can be implemented in addition to the summative tests and examinations that are conducted by the University. Formative assessment is useful for teachers as it takes place during instruction. So the teacher is actively involved in gathering and acting on information about students' learning in order to improve their work while it is actually taking place. Assessment thus plays a very important role in the differentiated classroom as the learner is able to construct an understanding for himself/herself. The difference in assessment procedure lies in giving attention to how quantitative and qualitative evidence of understanding can amplify learning before, during and after instruction.

V APPLICATION OF THE METHOD IN THE CLASSROOM

The present study is primarily concerned with how to improve the writing abilities of the undergraduate Engineering students. In order to devise suitable teaching and learning techniques for developing writing ability for use in regular classrooms, an experiment was conducted to teach students to write appropriate analytical paragraphs and monitor their progress over the course as well as to ascertain whether instruction in these helped them to progress better on writing tests. Their learning of writing analytical paragraphs was judged by

- i) their performance on the tests as well as
- ii) their retrospections on the skills they have acquired.

The performance of the students has been studied on questions on passages which are meant to elicit the following cognitive skills (or sub-skills), namely,

- i) Knowledge
- ii) Comprehension
- iii) Application
- iv) Analysis

The above taxonomy is based on Bloom's taxonomy of cognitive skills. For the purpose of this study the higher order cognitive skills namely Synthesis and Evaluation which is included in Bloom's taxonomy has not been considered.

Student performance is studied prior to and as a result of training in writing coherent paragraphs. Their differential performance on the four cognitive skills isolated here and elicited through the questions will also be studied. A Control Group was also set up, in which writing paragraphs was taught by the regular method used in classrooms, namely, giving students practice in writing.

Hypotheses

1. Experimental Group students, trained in analytical skills will show a significant improvement in overall scores in tests, while the Control Group students, who have not been given this treatment in their course, will not show a significant improvement in scores.
2. The Experimental Group students will show greater improvement in coherence of thought in writing compositions in comparison with the Control Group.
3. The rate of progress in writing grammatically correct sentences in paragraph writing will be higher in the Experimental Group than the Control Group.

Subjects

Out of the total population of undergraduate Engineering students in different disciplines, students of Computer and Information Technology, were selected as the sample, because of easy availability, as the experimenter taught in that college. It was felt that these students would represent the wider population of undergraduate Engineering students as such.

Students of Pillai's Institute of Information Technology of the first year comprised the sample. Two matched groups were needed for the study, namely, the Experimental and the Control. The Experimental Group of 25 students was randomly drawn from first year Computer Engineering, and the Control Group of 25 students was randomly drawn from first year Information Technology.

A test was conducted to measure the level of comparability of the two groups. The *t* test performed on these groups indicated that the value was not significant ($t=0.23$). Therefore, the two groups could be considered to be matched.

Experimenter

The present researcher, who was the regular teacher of these students, conducted the experiment.

Materials and Procedures

An Entry Stage test (the same as the test of comparability) was conducted at the beginning of the course for both Experimental and Control Groups. A similar, though not entirely identical, Exit Stage test was conducted at the end of the course for both groups.

The Experimental Group students were divided into groups, formed on the basis of their levels of understanding. Based on this, tiered activities and assignments were designed and given to the Experimental Group. The course lasted for three months, with one training session per week being conducted in the regular class.

The tests were graded according to the levels of the groups in linguistic and cognitive terms. The Entry Stage test was considered to be Test 1, followed by one test termed as the "Early Developmental Stage", and by another test termed the "Later Developmental Stage". The last test, that is the fourth, was termed the Exit Stage test.

The passages selected for the tests and the questions asked at the end of each were devised by the experimenter. The experimenter assessed the tests and conducted the students' interviews immediately after the tests, in order to tap their recollection of the background information used and the skills used to answer the questions. On the basis of the skills that students claimed to use they were instructed to join the next level of group and perform the tasks. The experimenter did not try to isolate the skills as such but made it a point to enable learners to acquire the knowledge of the particular skill from peers. The experimenter performed the role of the facilitator and extended his/her support to enable learning.

The tasks set comprised questions based on Knowledge, Comprehension, Application and Analysis. Both groups attempted these exercises after suitable explanation and discussion. The similarity of treatment extended to the two groups, ends at this point. Students of the Control Group were treated as in a regular class, which meant that after each test the answers would be corrected and

discussed. The students of the Experimental Group were given an added input that is, training in writing coherent paragraphs, in addition to the explanation and discussion of the answers. An example of tiered lesson is given below:

Tiered Activities

To alter the depth of a lesson

Subject area: Communication Skills

Grade level: Undergraduate Engineering

Topic: Elements of the Communication Cycle

Table 1.0

	Basic	Intermediate	Advanced
Outcome/performance indicators	Students will write a dialogue between a customer who wants to buy a cellphone and a salesman.	Students will be asked to comment on a dialogue that is displayed on the computer screen.	A crossword puzzle based on the components of communication cycle is given
Assessment	Assessment using the computer-asynchronous interaction.	Assessment using the computer-asynchronous interaction.	Assessment using the computer-asynchronous interaction.
Instruction/learning activity	Teacher will provide examples of effective and coherent dialogues through asynchronous interaction. Students will read the dialogues and learn the use of effective communication and acquire the skill of identifying the components of communication.	Teacher will explain the essential elements of communication through the example of dialogue writing. Students will identify one element of the communication cycle, list 4-5 uses of the element and explain each of the uses.	Students will learn the different components/elements of the communication cycle. They will be in a position to evaluate the appropriate medium or channel necessary in a particular situation
Resources	Students will evaluate their own writing with the help of examples of dialogue writing and exchange their writing with the Intermediate students.	List of communication situations based on the area of interest of students that have different medium or channel used.	Online communication situations, giving them opportunities to communicate with Intermediate level and the Basic level students and also encouraging them to assess their work.
Homework	Students will collect good and bad examples of dialogue writing from magazines, younger students' work, and the media.	Students will collect good and bad examples of persuasive communication situations from newspapers, magazines and periodicals.	Students will search the web for multiple examples of communication situations and communicate the same with the class.

It was felt that designing activities on the basis of the levels of understanding of the students was required to develop appropriate writing skills and enable them to grasp the content. It was expected that practice in writing activities based on the learners needs and their learning styles would lead to their spontaneous use of the skills acquired by the learner at the end of the course. The idea behind the entire approach was to help the learner understand the content without getting bogged down at the level of syntax or vocabulary.

The responses of the learners were analysed by two means, namely the quantitative method and the qualitative method. The quantitative method, because of its claimed objectivity, is often regarded as the sole means for determining the progress or failure of the student, as it leaves no room for bias or prejudice on the part of the experimenter. However, qualitative data can also provide insights into the nature of mental processing which the former cannot. The qualitative data helped the experimenter gain information about the needs of the learners, their area of interest, their preferred learning style and their progress from the Entry to Exit Stages.

Results

The Entry Stage data from the Experimental and Control Groups is given below:

Table 1.1
Total scores of the Experimental and the Control Groups at Entry Stage

Experimental Group				Control Group			
Students	Total	Mean	Std	Students	Total	Mean	Std.
(N=25)	Scores Max=14	Scores	Dev.	(N=25)	Scores Max=14	Scores	Dev.
1	6.0			1	7.0		
2	9.0			2	9.0		
3	5.5			3	7.0		
4	5.0			4	8.5		
5	7.0			5	6.0		
6	9.0			6	7.5		
7	5.5			7	7.0		
8	7.5			8	7.0		
9	6.5			9	8.5		
10	7.5			10	6.0		
11	7.0			11	6.0		
12	7.0			12	7.0		
13	10			13	7.5		
14	7.0			14	4.5		
15	7.0			15	7.5		
16	5.5			16	8.0		
17	6.5			17	11.5		
18	6.5			18	3.0		
19	6.5			19	7.5		
20	7.0			20	6.0		
21	6.5			21	10.0		
22	6.5			22	11.0		
23	8.5			23	7.5		
24	8.0			24	7.5		
25	7.5			25	7.5		
Totals	175.5	7.02	1.55	Totals	185.5	7.42	2.26

Table 1.1 shows the total scores of the Experimental and Control Groups at the Entry level. It also gives the mean scores and the standard deviation of each of the groups.

The total and mean scores indicate that the Experimental Group and the Control Group are comparable.

The following table shows the overall performance of the Experimental Group in all sub-skills combined from the Entry to the Exit stage.

Table 2.0

Comparison of Entry and Exit Stages of the Experimental Group for all sub-skills combined Entry Stage Exit Stage

Students (N=25)	Totals All Skills Max = 14	Mean scores	Std. Dev.	Totals All Skills Max = 14	Mean scores	Std. Dev.	Diff. in Means	t values
1	6.0			9.0				
2	9.0			11.5				
3	5.5			9.0				
4	5.0			8.0				
5	7.0			10				
6	9.0			10.5				
7	5.5			9.0				
8	7.5			9.5				
9	6.5			9.0				
10	7.5			9.5				
11	7.0			8.5				
12	7.0			10				
13	10			11.5				
14	7.0			9.5				
15	7.0			9.5				
16	5.5			9.0				
17	6.5			9.0				
18	6.5			9.0				
19	6.5			9.5				
20	7.0			10.5				
21	6.5			9.5				
22	6.5			9.0				
23	8.5			11				
24	8.0			10.5				
25	7.5			10				
Totals	175.5	7.02	1.55	241	9.64	0.847	1.25	2.35

$p > .02$

Table 2 shows the difference in the mean scores of the students in the Experimental Group from Entry to Exit Stage in proficiency in all the four skills using tiered activities and flexible grouping in the class. The t test administered in order to find out the difference between performance at the Entry and the Exit Stages shows that there is a significant difference at the .02 level. This indicates that students have improved as a result of the training given to them in writing skill. The mean scores also point to this fact. At the Entry Stage, the mean score was 7.02, whereas at the Exit Stage the mean score moved up to 9.64, thereby indicating the overall improvement of the students.

On the other hand, the standard deviation has decreased from 1.55 at the Entry Stage to 1.25 at the Exit Stage. This indicates that the range of dispersal of scores has decreased, thereby pointing to the fact that all the students are improving.

Next, the performance of the Experimental Group on the individual sub-skills of Knowledge, Comprehension, Application and Analysis is considered in Table 2.1, 2.2, 2.3 and 2.4

Table 2.1
Comparison of Entry and Exit Stages of the Experimental Group in Knowledge

Students (N=25)	Total Scores (Max=2)	Mean Scores	Std. Dev.	Total Scores (Max=2)	Mean Scores	Std. Dev.	Diff. in means	t values
1	1.0			1.5				
2	1.5			1.5				
3	1.0			1.5				
4	1.0			1.5				
5	1.0			2.0				
6	2.0			1.5				
7	1.0			1.5				
8	1.5			2.0				
9	1.0			1.5				
10	1.0			1.5				
11	1.0			1.5				
12	1.0			2.0				
13	2.0			2.0				
14	1.0			1.5				
15	1.5			1.5				
16	1.0			1.5				
17	1.0			1.5				
18	1.0			1.5				
19	1.0			1.5				
20	1.5			2.0				
21	1.0			1.5				
22	1.5			1.5				
23	2.0			2.0				
24	1.5			2.0				
25	1.5			2.0				
Totals	31.5	1.26	0.35	41.5	1.66	0.23	0.3	1.49

$p > .05$

Table 2.2
Comparison of Entry and Exit Stages of the Experimental Group in Comprehension

Entry Stage				Exit Stage				Diff. in Means	t values
Students (N=25)	Total Scores Max=3	Mean Scores	Std. Dev.	Total Scores Max=3	Mean Scores	Std. Dev.			
1	1.0			1.5					
2	2.5			3.0					
3	1.0			1.5					
4	1.0			1.5					
5	1.5			2.0					
6	2.0			2.5					
7	1.0			1.5					
8	1.5			2.0					
9	1.0			1.5					
10	1.5			2.0					
11	1.5			1.5					
12	1.5			2.0					
13	2.5			2.5					
14	1.5			2.0					
15	2.0			2.5					
16	1.0			2.0					
17	1.5			2.0					
18	1.5			2.0					
19	1.0			1.5					
20	1.5			2.5					
21	1.5			2.0					
22	1.5			2.0					
23	2.0			2.5					
24	1.5			2.0					
25	1.5			2.0					
Totals	37.5	1.5	0.43	50.0	2.0	0.40	2.14	1.21	

$p > .05$

Table 2.3
Comparison of Entry and Exit Stages of the
Experimental Group in Application

ENTRY STAGE				EXIT STAGE				Diff. in Means	t values
Students (N=25)	Total Scores Max=4	Mean Scores	Std. Dev.	Total Scores Max=4	Mean Scores	Std. Dev.			
1	2.0			3.0					
2	3.0			4.0					
3	1.5			2.5					
4	1.5			2.5					
5	2.0			3.0					
6	2.0			3.0					
7	1.5			2.5					
8	2.0			2.5					
9	2.0			3.0					
10	2.0			2.5					
11	2.0			2.5					
12	2.0			3.0					
13	2.0			3.0					
14	2.0			2.5					
15	1.5			2.5					
16	1.5			2.0					
17	1.5			2.0					
18	2.0			2.5					
19	2.0			3.0					
20	2.0			3.0					
21	2.0			2.5					
22	1.5			2.0					
23	2.5			3.5					
24	2.5			3.5					
25	2.0			2.5					
Totals	48.5	1.94	1.54	68.5	2.74	0.46	1.14	0.78	

p>.05

Table 2.4
Comparison of Entry and Exit Stages of the
Experimental Group in Analysis

ENTRY STAGE				EXIT STAGE				Diff. in Means	t values
Students (N=25)	Total Scores Max=5	Mean Scores	Std. Dev.	Total Scores Max=5	Mean Scores	Std. Dev.			
1	2.0			3.0					
2	2.0			3.0					
3	2.0			3.5					
4	1.5			2.5					
5	2.5			3.0					
6	3.0			3.5					
7	2.0			3.5					
8	2.5			3.0					
9	2.5			3.0					
10	3.0			3.5					
11	2.5			3.0					
12	2.5			3.0					
13	3.5			4.0					
14	2.5			3.5					
15	2.0			3.0					
16	2.0			3.5					
17	2.5			3.5					
18	2.0			3.0					
19	2.5			3.5					
20	2.0			3.0					
21	2.0			3.5					
22	2.0			3.5					
23	2.0			3.0					
24	2.5			3.0					
25	2.5			3.5					
Totals	58.0	2.32	0.42	81.0	3.24	0.31	0.37	2.78	

p>.01

Tables 2.1, 2.2, 2.3 and 2.4 show the difference in the mean scores and *t* values of the students in the Experimental Group from the Entry to the Exit Stages in Knowledge (0.3, *t*=1.49), Comprehension (2.14, *t*=1.21), Application (1.14, *t*=0.78), and Analysis (0.37, *t*=2.78). The *t* tests performed on each of these sub-skills show a significant difference from the Entry to the Exit Stages. This points to an improvement in the students' ability to tackle questions as a result of extensive training in the four skills that is speaking, reading, writing and listening. In Knowledge, Comprehension and Application the improvement level is at the .05 level, whereas in Analysis the level of significance is at .01 level thereby indicating an improvement of all the students from the Entry to the Exit Stages.

The standard deviation scores in Knowledge, Comprehension, Application and Analysis are seen to have dropped from the Entry Stage to the Exit Stage, indicating that the range of dispersal of scores has been reduced, that is, the difference between individual students has been reduced. Thus the group as a whole has shown improvement as a result of training in tasks designed on the basis of tiered activities and flexible grouping.

Having analyzed the improvement of the Experimental Group we now move on to the analysis of the Control Group.

Table 3.0
Comparison of Entry and Exit Stages of the Control
Group in all skills combined

ENTRY STAGE				EXIT STAGE				Diff. in means	t values
Students (N=25)	Totals: All Skills Max=14	Mean Scores	Std. Dev.	Totals: All Skills Max=14	Mean Scores	Std. Dev.			
1	7.0			6.5					
2	9.0			9.5					
3	7.0			8.0					
4	8.5			9.0					
5	6.0			7.0					
6	7.5			8.5					
7	7.0			8.5					
8	7.0			9.0					
9	8.5			9.5					
10	6.0			8.5					
11	6.0			7.0					
12	7.0			8.0					
13	7.5			9.0					
14	4.5			6.5					
15	7.5			8.5					
16	8.0			9.5					
17	11.5			10.5					
18	3.0			5.5					
19	7.5			8.0					
20	6.0			6.5					
21	10.0			10.5					
22	11.0			11.5					
23	7.5			8.5					
24	7.5			7.5					
25	7.5			7.0					
Totals	185.5	7.42	2.26	208	8.32	1.43	1.89	0.53	

p>.20

Table 3 shows the overall performance of the Control Group in proficiency in writing skills from the Entry to the Exit Stages.

Next the performance of the Control Group on the individual sub-skills of Knowledge, Comprehension, Application and Analysis is considered in Tables 3.1, 3.2, 3.3 and 3.4.

Table 3.1
Comparison of the Entry and Exit stages of Control Group in the level of Knowledge

ENTRY STAGE				EXIT STAGE				
Students (N=25)	Total Scores Max=2	Mean Scores	Std. Dev.	Total Scores Max=2	Mean Scores	Std. Dev.	Diff. in means	t values
1	1.5			1.0				
2	2.0			1.5				
3	1.5			1.5				
4	1.5			1.5				
5	1.0			1.5				
6	2.0			2.0				
7	1.5			2.0				
8	1.5			2.0				
9	1.0			1.5				
10	0			1.5				
11	1.0			1.5				
12	1.0			1.5				
13	1.5			2.0				
14	1.0			1.5				
15	1.5			2.0				
16	1.5			2.0				
17	2.0			1.5				
18	1.0			1.0				
19	1.5			1.5				
20	1.0			1.0				
21	2.0			2.0				
22	2.0			2.0				
23	1.0			1.5				
24	1.0			1.0				
25	1.0			1.0				
Totals	33.5	1.34	0.46	39	1.56	1.38	2.51	0.96

p>.20

Table 3.2
Comparison of the Entry and Exit stages of Control Group in Comprehension

ENTRY STAGE				EXIT STAGE				
Students (N=25)	Total Scores Max=3	Mean Scores	Std. Dev.	Total Scores Max=3	Mean Scores	Std. Dev.	Diff. in Means	t values
1	1.5			1.5				
2	2.0			3.0				
3	1.5			2.0				
4	1.5			2.0				
5	1.5			2.0				
6	1.0			2.0				
7	1.0			2.0				
8	1.5			2.5				
9	1.5			2.0				
10	1.5			2.0				
11	1.0			1.5				
12	1.5			2.0				
13	1.5			2.0				
14	1.5			1.5				
15	2.0			2.5				
16	1.5			2.5				
17	2.5			2.0				
18	1.0			2.0				
19	2.0			2.5				
20	1.0			1.5				
21	2.0			2.5				
22	2.5			3.0				
23	1.5			2.0				
24	2.0			2.0				
25	2.5			2.0				
Totals	40.5	1.62	0.45	52.5	2.1	0.40	2.25	1.29

p>.20

Table 3.3
Comparison of the Entry and Exit Stages of Control Group in Application

ENTRY STAGE				EXIT STAGE				
Students (N=25)	Total Scores Max=4	Mean Scores	Std. Dev.	Total Scores Max=4	Mean Scores	Std. Dev.	Diff. in Means	t values
1	2.0			2.0				
2	3.0			3.0				
3	2.0			2.5				
4	2.5			2.5				
5	1.5			1.5				
6	2.0			2.0				
7	2.5			2.5				
8	1.5			2.0				
9	2.5			2.5				
10	2.5			2.5				
11	1.5			1.5				
12	2.0			2.0				
13	2.5			2.5				
14	0			1.5				
15	2.0			2.0				
16	2.0			2.0				
17	3.0			3.0				
18	1.0			1.5				
19	2.0			2.0				
20	2.0			2.0				
21	3.0			3.0				
22	3.0			3.0				
23	2.0			2.0				
24	2.0			2.0				
25	2.0			2.0				
Totals	52.0	2.08	1.83	55.0	2.2	0.46	1.34	0.100

p>.20

Table 3.4
Comparison of the Entry and Exit Stages of Control Group in Analysis

ENTRY STAGE				EXIT STAGE				
Students (N=25)	Total Scores Max=5	Mean Scores	Std. Dev.	Total Scores Max=5	Mean Scores	Std. Dev.	Diff. in Means	t values
1	2.0			2.0				
2	2.0			2.0				
3	2.0			2.0				
4	3.0			3.0				
5	2.0			2.0				
6	2.5			2.5				
7	2.0			2.0				
8	2.5			2.5				
9	3.5			3.5				
10	2.0			2.5				
11	2.5			2.5				
12	2.5			2.5				
13	2.0			2.5				
14	2.0			2.0				
15	2.0			2.0				
16	3.0			3.0				
17	4.0			4.0				
18	0			1.0				
19	2.0			2.0				
20	2.0			2.0				
21	3.0			3.0				
22	3.5			3.5				
23	3.0			3.0				
24	2.5			2.5				
25	2.0			2.0				
Totals	59.5	2.38	0.72	61.5	2.46	0.93	0.84	0.108

p>.20

Tables 3.1, 3.2, 3.3 and 3.4 show the difference in the mean scores of the students in the Control Group from the Entry to the Exit stages in Knowledge (2.51), Comprehension (2.25), Application (1.34) and Analysis (0.84). The t -test performed on all the sub skills show an improvement at the .20 level. The level of improvement is less in comparison to the Experimental Group where the level of improvement in the sub skill of Knowledge, Comprehension and Application is at .05 levels. This indicates a trend towards significance for the students in the Experimental Group. In the sub skill of Analysis there is a significant improvement at .01 level of the Experimental Group. This point to the fact that students in the Control Group require training in order to improve their skills. They will not be able to perform with ease in the sub skills with practice.

Table 4.0

Comparison of t - values in the sub-skills, and the rate of progress at Entry and Exit Stages of the Experimental and Control Groups

Experimental Group		Entry Exit stages		Control Group	
Sub- skills	t - values of diff. in means	The sig. level	Rate of Progress	t - values of diff. in means	The sig. level
Knowledge	1.49	.05	20%	0.96	.20
Comprehension	1.21	.05	16.6%	1.29	.20
Application	0.78	.05	20%	0.100	.20
Analysis	2.78	.01	18.4%	0.108	.20
Total of all sub- skills	6.26			2.45	

Table 4 shows the t -values and the rate of progress of the Experimental and the Control Group from the Entry to the Exit Stages. The t values show that there is a significant difference at .05 level from the Entry to the Exit Stages in the sub-skill of Knowledge, Comprehension and Application for the Experimental Group. This is not the case with the Control group where the significant level in all the sub skills is at .20 level. This means that students do not improve significantly with practice. The students show greater improvement as a result of training in tasks designed on the basis of tiered activities and using flexible grouping which is indicated in the t values of the Experimental group.

When we compare the rate of progress of the Experimental and the Control groups we find that students in the Experimental group have improved at a higher rate. This would suggest that students in the Experimental group have benefited more than the Control Group from the tasks that the subject teacher designed keeping in mind the levels of the students.

VI DISCUSSION

The results indicate that the students of the Experimental Group have improved significantly in their overall performance. This supports Hypothesis 1. The scores of the Control Group (in its overall performance) suggest that training using the concept of the differentiated instruction and using the technological devices is a necessity.

An analysis of the individual sub skills shows that students of Experimental Group have made a significant improvement in scores on questions involving Knowledge, Comprehension Application and Analysis. This is not so with the Control Group. This supports Hypothesis 2. We can therefore, conclude that training in appropriate skills is absolutely necessary.

In the case of Analysis students in the Experimental Group show a significant improvement at .01 levels from the Entry to the Exit stages. This also supports Hypothesis 2.

With regard to the rate of progress, the Experimental Group showed rapid progress than the Control Group and this can be attributed to group activity, learning from peers, and tiered lessons using technology.

VII QUALITATIVE ANALYSIS

After analyzing the recording of the statements of the Experimental Group it was found that students felt comfortable and relaxed when tasks were given on the basis of their level and their area of interest. It was also found that improving their writing abilities is more successful when they are allowed to discuss and interact with their peers. There was a significant improvement in their motivation to write grammatically correct sentences and develop their coherence of thought in expression. This seems to support Hypothesis 3.

Moreover the freedom of using cellphones, books, dictionaries, newspapers and the internet stimulated them to engage in the activities. Students were empowered to control their learning. A sense of responsibility and curiosity was also reflected in their handling of the task. The independence and the involvement that was observed in the classroom by the teacher while performing the task helped the teacher conclude that using differentiated instruction with technology can promote reading and writing competence. Here the learning process itself was enjoyable and taken as fun. The tasks designed to suit the needs of the students also enabled them to develop their critical and creative thinking.

The teaching method used in these classes emphasizes the need to acquire knowledge about the digital media, and develop networks that supports strategic and

affective learning. With strategic learning the teacher could focus on factors such as offering the students flexible opportunities to demonstrate their skills as well as providing ongoing and relevant feedback. In addition the teacher could provide multiple examples and models of skilled performance and extend their support to students to practice with these tools. The idea behind this recommendation is to select goals, methods, assessment and materials in a way that will minimize barriers and maximize flexibility. (Rose & Meyer, 2002). The teacher can comfortably handle all these responsibilities with the help of the digital media. This can further be useful in extracting the defining features of a pattern which can practically provide constructive information to the teacher to differentiate it from similar patterns simply by viewing the multiple examples. Presentation of multiple examples in a classroom by the means of printed material can be difficult for the teacher. So the use of digital materials can be used in storing large collection of information which can easily be modified to suit the needs of the learners.

Secondly a congenial atmosphere is created in the classroom that can help the teacher in the formation of teams. The aim is to encourage and engage the students in meaningful activities.

This method also accounts for a third responsibility of the teacher in terms of encouraging students in presenting their output in the class. Flexible means of expressions needs to be appreciated and acknowledged by the teachers in order to boost their morale.

Immediate assessment of student readiness is another area that needs to be considered to ensure the acquisition of the skills. Successful pre-assessment and ongoing assessment helped the teacher to provide varying needs, interests and abilities. Teacher can have liberty in using formal and informal assessments. It can be in the form of surveys, procedures, performances or interviews. Research shows that involvement of the students in the process of learning is more profound if they are met with achievable challenges. Sometimes the output of the students can be handy in terms of enabling the teacher to work out on varied or alternative expressions and procedures.

The introduction of the differentiated instruction as a model did provide the teacher with a lot of opportunity to tap the cognitive skills of the student. But it is felt that this method should be implemented slowly and gradually in the classroom. This can help the teachers and educators to have a thorough knowledge of the content and prepare a wide variety of tasks. Moreover the effectiveness of the tasks lies in sharing the creative load between teachers in the same area. This step will facilitate teachers and educators develop ideas and approaches that can enhance differentiation.

CONCLUSION

To conclude differentiating the needs, pace the learning styles of the learners and designing tasks to meet the real life needs of the learners is an attempt towards motivating students to overcome their writing difficulties. Engaging students in meaningful tasks can enhance cognitive development as well as develop their curiosity in learning. This can be done only when teachers realize that all learners are not the same. The area of interest, their goals, their dreams, their aspirations and moreover their background differs. So the need is to consider these issues and design activities that encourage the involvement and participation of the student. Further this method of instruction can provide teachers with opportunities to appropriately challenge the scope of students in classrooms today.

REFERENCES

- [1] Hall. Tracey, Strangman. Nicole, Meyer. Anne, "Differentiated instruction and implications for UDL implementation," CAST (Centre for Applied Special Technology) Wakefield, 2003
- [2] Erben. Tony, Ban Ruth, Castaneda Martha, "Teaching English language learners through technology", 49-55, 56-64, Routledge, New York, 2009.
- [3] Nunan. D, "The learner-centred curriculum", Cambridge University Press, Cambridge, 1988.
- [4] Vygotsky, L.S., "Mind in Society: the development of higher psychological processes", Cambridge, MA: Harvard University Press, 1978.
- [5] Hall, T., "Differentiated instruction, effective classroom practices report", National Center on Assessing the General Curriculum, CAST, U.S., Office of Special Education Programs, 2002.
- [6] Lawrence, Brown, D., "Differentiated instruction: inclusive strategies for standard based learning that benefit the whole class", American Secondary Education, 32(3), 34-62, 2004.
- [7] McAdams, S., "Teacher tailor their instruction to meet a variety of student needs", Journal of Staff Development, 22(2), 1-5, 2001.
- [8] Stronge, J., "Teacher effectiveness and student achievement: what do good teachers do?" Paper presented at the American Association of School Administrators Annual Conference and Exposition, San Francisco, California, 2004.
- [9] Tomlinson, C. A., "Deciding to differentiate instruction in the middle school: one's school's journey", Gifted Child Quarterly, 39 (2), 77-114, 1995.
- [10] Chapelle, C.A., "Computer applications in second language acquisition" Cambridge: Cambridge University Press, 2001.
- [11] Willis, Jane, "A framework for task-based learning", Pearson Education Limited, Addison Wesley Longman Limited, 4-10, 14-15, 23-26, 116, England, 1996.
- [12] Prabhu, N.S. "Second language pedagogy: a perspective", Oxford University Press, 1987.
- [13] Sinclair, J.M., "Corpus, concordance and collocation", Oxford University Press, 1991.
- [14] Tomlinson, C.A., "Differentiation of Instruction in the elementary grades", ERIC Digest, ERIC Clearing House on Elementary and Early Childhood Education, 2000.

ADAPTABILITY - A COMPETENT SKILL TO HARNESS THE POWER OF EFFECTIVE COMMUNICATION

Anita Yeshwant Padhye

Assistant Professor, Basic Science & Humanities Dept.,
K J Somaiya Institute of Engineering and Information Technology,
Sion, Mumbai-22

It is not the strongest of the species that survives, nor the most intelligent that survives,
it is the one that is the most adaptable to change

– *Charles Darwin*

ABSTRACT

In the era of specialization of knowledge and information technology, we ought to become adaptive to continuously evolving needs of socio-economic and cultural structures of age, gender and race especially during the process of communication. This paper highlights the potentiality of adaptability skills which when harnessed lends plasticity to the emotional and ideational experience of an individual resulting in personal communication. The outcome of it is felt when the individual overcomes the intrinsic resistance to function productively, grows personally and expands professionally.

Keywords: - *adaptability, skills, sender, receiver.*

INTRODUCTION

Communication is a skill that combines behavioural skills with that of practical wisdom. Words, actions and techniques measure the outcomes. Effectively communication, otherwise, is the self-assessment of the communicator to vouch for one's own credibility. According to Mahatma Gandhi, "Adaptability is the problem the resistance and assimilation".

THE PROBLEM :

Primarily, initiatives on the grounds of our education system produce thinkers with lesser creativity or skilled creators with limited outputs. The result is that it propels attitudinal inertia. Even training programmers fade out from the memory of trainees, as balancing both the art and the science of Communication takes its toll.

Our society is overloaded with information. Easy accessibility leaves an individual overwhelmed as he/she unconsciously is a participant of a message competition. The number of choices available forces us to make the choices of what can be and what cannot be ignored. Some choices empowers us where as some paralyses us.

(1)

Sensitivity to and an understanding of how things work and differ in their contexts; adaptation is the way, groups relate to their environment to grow, survive and to branch out in new directions. It empowers one to become resilient in a crisis situation.

We get calcified with the thought of 'satisfactory underperformance' due to information overload and monotony and fail to enjoy the outcome of experiential learning.

THE COMMUNICATION ANGLE :

Communication process is based on the thinking models of the parties involved. There is always a difference in the desired state. The difference is identified on the basis of the analysis by the sender (in terms of communication) for the receiver (in terms of communication) on demographic, disposition and knowledge planes. Aristotle calls it "finding a common ground".

The demographic analysis is concerned with the age, range, socio-economic status, religions and/political affiliations of a homogeneous or a heterogeneous 'audience' (in terms of composition).

The disposition analysis is concerned with stresses, problems and needs.

The knowledge analysis is about the topic specification or the information which could be simple or complex. Similarly, it includes new information which may benefit the receivers.

Such analysis becomes the crux of the sustainability of communication as it refers to connectivity between the sender and the receiver. In spite of all such care, the communication may suffer from monotony because of the limited 'knowledge adaptability' of the sender. The liberal and creative uses of general and technical forms of communication break such monotony. Hence,

content, style and approach of presentation show cases the subject competency, linguistic competency, and the organizational competency of the individual.

Any content is the extension of the responses to already existing content. There is always an element of 'psychological preparedness' in the mind of the sender to communicate, thereby, adapting message for audience for clarity to consider for their time and energy. (P8)

Personal communication helps us to make and sustain the relationships between the sender and the receiver. Repeated reassessment of communication skills make it resourceful as it is all about a sight and thought of a bigger picture.

We perceive truth and reality in a different manner. Perhaps, this trait of communication makes it more dynamic, interactive, and symbolic and recursive due to the transmission of the meanings between the encoder and the decoder. Irrespective of an individual's background, adaptability becomes an indirect test to scale individual's

- i) Openness to new ideas
- ii) Adaptations to situations
- iii) Handling of unexpected demands
- iv) Changing strategy

Communication has a degree of its multiplicity, ambiguity, contradiction, paradox and general confusion. * So, every act of communication is an algorithm which requires great understanding of human nature. It is an analysis of differences in communicator's background, perceptions, and point of view, skills and talents of people and their thoughts and feelings. Every individual perceives truth and reality in a different manner which attributes to the formation of his behaviour. Information from the surroundings is perceived through sensory organs to form psychological maps; acknowledging such maps leads to enhancing communication. For example, adapting communication style for establishing rapport with the people who are impressed by the use of visuals of diagram flow-charts, and other graphics, particularly for complex presentation may give prompt feedback. The digitals (people who are very logical) get satisfied by defining terms and providing order of sequence to information. The auditorys (people whose learning style is by listening) move around by varying voice, speed and delivery. Kinesthetic are the people who learn by doing the things by sitting down occasionally and talking to them in a relaxed position at their eye level make them feel very comfortable. Communication becomes easier when 'preferred representational system predicates' are paced particularly during summarizing, advocating and asking for assistance. (2)

Hence, acquiring proficiency in such kind of behavioural adaptability improves ability, affectivity to

build up a strong foundation and rapport with that of the audience. The outcome of this understanding that supports the communication, when structured for clarity also calls upon simplicity, to acquire a change; for example, when we write. The sender knows his subject better than the receiver. According to the readability research of experts – Gunning, Dale, Chale & Flesch, "Writing slightly below the reader's level of understanding communicates in the best manner". (3)

Candid information testifies for both reliability and information credibility of the encoder only when it is expressed in a concrete and comprehensive manner. In modern times, new processes of communication are the collaborative nets and the hybrid systems to combine communication for mass and for personal communication and the horizontal communication. A search for equivalents heightens the complexities and limitation of communication. (4)

Learning and teaching acts have become the judgment of organizational / institutional capabilities to relate the matter to bring about connectivity in communication. There is a word of caution that an easier and simple knowledge management is often misunderstood by those who are the vendors to peddle information management. The biggest limitation to communication is the presence of variables. It is very difficult to judge 'how' and 'where' in the process of communication technology.

"Digital technologies have become a powerful tool for quality pedagogy by designing innovative learning 'environments'. These environments are created by making learning activities more participatory, by engaging active real-word connected, self-directed and relevant features. Technological literacy is a key to 21st century skill for all of us to be confident and competent to do research and synthesize ideas through a range of online mediums. This may increase learners' ownership of their learning to connect it to the real world rather than conceiving it as a passive body of learning. Adding new professional learning opportunities makes one feel comfortable to construct new ideas to be integrated by the technology which has its own philosophy. (5)

As quoted by President & CEO of IBM "Whatever business you're in its going to commodities over time so you have to keep moving it to a higher value & change". (6)

So what is expected outcome from the stakeholders? The answer is innovation sustainable models inclusiveness, diversity and holistic approach sensitizing learning and opportunity richness.

THE MANAGEMENT / CORPORATE ANGLE:

Adaptability scouts for all these general necessities as it helps one to move beyond percentiles to holistic personality, from physical infrastructure to mind structure, from faculty influence / guide to one's actions of decisions. Hence, all the requisites of globalization 'triple bottom line', 'more for less for more', 'disruptive innovation sustainability', 'technological revolution', 'social responsibility', 'ethics', 'governance' etc. are just limited to the business sector, rather in the world of borderless technology these terms provide a sort of mentoring to the learner to adapt according to the societal changes.

THE INDIVIDUAL ANGLE:

Individually, adaptability has its own advantages -

-Develops individual's knowledge, skills and attitudes towards social experience, composure and social acknowledgement by analyzing options and their implications in an objective and focused manner.

-Updates information enthusiastically and the articulation skills. In addition to this, an adaptive performer anticipate changes in work demand by undergoing training to prepare self to overcome work deficiencies.

-Continues feedback for goal attainment, vital for enriched transfer of training and becoming more resilient for constructive solutions. The individual may effectively adjust plans, goals, actions or priorities to deal with changing situations and not getting paralyzed by uncertainty and ambiguity.

THE LIMITATIONS:

Even though we come across great many benefits of adaptability but the 'psychological stringency' in the form of filtering, perceptions, faulty memory, poor listening skills and emotional interference, affect communication. The feedback is obviously the gap between the intended and the desired and the interpreted meanings of the messages.

Being adaptable makes learning and development multipronged to the extent where appeal is to growth needs rather than the deficiency needs of an individual.

CONCLUSION

In the age of seamless integration of teaching methodologies to create a multi-sensory learning experience for the students, adapting visually rich and more cognitive learning experiences may prove a benchmark. In the corporate sector, where

'innoinformation' (innovation & information) and 'intrapreneurship' (tapping the talent of internal employees) are the new words, isn't it the adaptability skill of an individual, handy, to become competitive in the borderless world of knowledge. In order to harness the power of effective communication, it strictly should follow and adapt to create and to innovate! So, adaptability is of you to your message rather than adapting message for the audience. The futuristic value of adaptability is that it capacitates the adaptor to increase the depth of his / her knowledge domain and width of overall perspective and sensitivity of the organization and the community.

REFERENCES

- [1] Barry Schwartz - The paradox of Choice.
- [2] Management Information Systems and Corporate Communication - 11, P.536.
- [3] <http://www.micropowerandlight.com/>
- [4] Business communication – Making connections in a Digital world – Raymond V Lesikar, Marie E Flatley, Kathryn Rentz, Neerja Pande, Eleventh edition. Tata MC graw Hill Education Pvt. Ltd., New Delhi
- [5] Ibid.
- [6] Innovative learning: Interview of Simon Breakspear, leading thinker and educationist, Monday- February 27, 2012 TOI Ascent.

PROMOTIONAL MIX POLICIES AND PRACTICES IN SMALL SCALE INDUSTRIES IN RAJASTHAN

Dr. Poonam Somani
Asst. Prof., MBA Department
Sri Balaji College of Engg.
& Tech., Jaipur, India
drpoonamsomani@yahoo.in

Surya Prakash Takhar
Asst. Prof., MBA Department
Sri Balaji College of Engg.
& Tech., Jaipur, India
stakhar@rediffmail.com

Dr. Hari Singh
Prof., MBA Department
Sri Balaji College of Engg.
& Tech., Jaipur, India
Harisingh_2000@yahoo.com

ABSTRACT

Promotion is the function of informing, persuading and influencing the consumers purchase behaviors. The dynamic nature of promotion makes it one of the most difficult and interesting area of marketing policies.¹ Promotion planning is a systematic decision making relating to all aspects of the development and management of a firm's promotional effort.² Promotion includes personal selling, advertising and other sales promotion techniques. Therefore, the managements of small enterprises should realize that successful marketing requires careful attention to promotional efforts.

Keywords- advertising; pricing; promotion; practices; product; SSI; policies

I. INTRODUCTION

The following are three major decision areas the field of promotion on which the entrepreneurs should focus their attention:

1. Personal selling
2. Advertising
3. Sales promotion

Here we shall bring out the promotional practices followed by Small scale industries in respect of these elements of promotion.

To begin with we shall bring out the relative importance of each of three elements in the promotion mix according to various product items under agro based industries. Table 1.1 gives the distribution of units arranged according to product items to bring out the fact which element is most important. 90% Cotton units followed by 80% Vegetable oil units and 70% Rice units use advertising technique of promotion. Thus, advertising is the most important with (60%) in Agro based industries and followed by personal selling 30%. Sales promotion is the least popular element of promotion mix for units belonging to agro industries.

Table 1.1

DISTRIBUTION OF SAMPLE UNITS OF VARIOUS PRODUCT ITEMS UNDER AGRO BASED INDUSTRIES ACC. TO RELATIVE IMPORTANCE OF PROMOTIONAL MIX COMPONENTS IN RAJASTHAN

Elements	Product Items					
	Dal	Vegetable Oil	Rice	Cotton Pressing & Ginning	Guar Gum	Total
Advertising	4 (40)	8 (20)	7 (70)	9 (90)	2 (20)	30 (60)
Sales Promotion	2 (20)	1(10)	2 (20)	-	-	05 (10)
Personal Selling	4 (40)	1(10)	1(10)	1(10)	8(80)	15(30)
Total	10	10	10	10	10	50(100)

Figures in parentheses are percentages of the total.

The computed value of chi-square notation is $\chi^2_{0.5} = 22.33$ which is greater than the tabulated value 15.507 at 5% level of significance with degree of freedom $V=8$. This shows that the Hypothesis of association between various product items under Agro based industries and relative importance of promotion mix is accepted.

Table 1.2

Distribution of Sample Units of Various Product Groups acc. to Relative Importance of Promotional Mix Components in Rajasthan.

Elements	Product Groups					
	POP	Agro Based	Chemicals	Agriculture Implements	Bone Grinding	Total
Advertising	3(30)	3(30)	2(20)	-	4(40)	15 (30)
Sales Promotion	2(20)	1(10)	2(20)	-	1(10)	06 (12)
	5(50)	6(60)	6(60)	-	2(20)	19 (38)
Personal Selling	-	-	-	10(100)	-	10 (20)
NO Technique						
Total	10	10	10	10	10	50(100)

Figures in parentheses are percentages of the total.

The Table 1.2 gives the distribution of sample units according various different industries, to see whether or not they are engaged in advertising, sales promotion and personal selling. The advertising is resorted to by 30% units. The most preferred choice is personal selling 38%. Sales promotion is again least preferred the computed value of chi-square notation is $\chi^2_{0.5} = 58.57$ which is greater than the tabulated value 21.026 at 5% level of significance with degree of freedom $V=12$. This shows that the Hypothesis of association between product groups and relative importance of promotion mix is accepted.

Agriculture implements industries do not use any promotional weapon because they say there is nothing new in their product. Our product is a necessity for farmer and they people are illiterate so promotional tools are useless for them. Main problem which SSI units face is limited financial resources as a result of this the cheapest way of advertising is followed. Agro based industries mainly depend on advertisement, they feel it is cheapest way of promotion, increases goodwill and supplements salesman, moreover, customer can make direct contact.

II. ADVERTISING

The importance of advertising has been summed up in the pithy statement, "Advertisement is telling and selling".⁴

The term advertising originates from latin "adverto" which means to any object or purpose. In the marketing context, advertising has been defined as any paid form of non personal presentation and promotion of ideas, goods or services by an identified sponsor.⁵

"Advertising consists of all the activities involved in presenting to an audience a non personal, sponsor-identified, paid for manage about a product or organization."⁶

In developing a advertising program, managers must always start by identifying the target market and buyer motives. Then they can make the five major decisions in developing an advertising program, known as five Ms. Mission: What are the advertising objectives? Money: How much can be spent? Message: What message should be sent? Media: What media should be used? Measurement: How should the results be evaluated? The advertising objectives must flow from prior decisions on target market, market positing & marketing mix. The objectives can be classified according to whether their aim is to inform, persuade, or remind.

III DECIDING THE ADVERTISING BUDGET

How much should a company spend on advertising? While this is undoubtedly a major decision all area no standard formula or norm can be applied. In fact, there is no standard norm at all. The practices vary from industry to industry from company to company within the square industry. And same firm's expenditure differs our time. So we can only indicate the general approaches in this regard. The important ones among the practices followed for deciding the advertising budget are :

- 1) Capacity to afford
- 2) A fixed percentage of turnover
- 3) Promotional expenses by competitors

Table 1.3

Distribution of Sample Units of Various Product Items Under Agro Based Industries according to Practices Followed for Deciding the Advertising Budget in Rajasthan.

Practices Followed	Product Items					
	Dal	Vegetable Oil	Rice	Cotton Pressing & Ginning	Guar Gum	Total
Capacity to Afford	1(10)	7(70)	3(30)	7(70)	-	18 (60)
Percentage of Sales	2(20)	1(10)	3(30)	2(20)	2(20)	10(33)
Promotional Expenses by Competitor	1(10)	-	1(10)	-	-	2(6.67)
Total	4	8	7	9	2	30(100)

The Table 1.3 gives the distribution of sample units of various product items according to methods used for deciding the advertising budget. It shows that the majority of the units in the Agro based industries use capacity to afford method, followed by those using the percentage of sales and the promotional expenses by competitor. The computed value of chi-square notation is $\chi^2_{0.5} = 16$, which is greater than the tabulated value 15.507 at 5% level of significance with degree of freedom $V= 8$. This shows that the Hypothesis of association between various product items under Agro based industries and advertising budget is accepted.

The distribution of sample units according to different product groups and methods used for deciding budget is given in table 1.4. It shows that the affordable method is mostly used in bone grinding and chemical units. Agriculture implements unit do not make advertisement budget. The computed value of chi-square notation is $\chi^2_{0.5} = 7$ which is lesser than the tabulated value 12.59 at 5% level of significance with degree of freedom $V=6$. This shows that the Hypothesis of association between various product groups & method of preparing budget is rejected.

Table 1.4

Distribution of Sample Units of Various Product Groups acc. to Practices Followed for Deciding the Advertising Budget in Rajasthan

Practices Followed	Product Groups				
	POP	Agro Based	Chemicals	Bone Grinding	Total
Capacity to Afford	2(66.67)	2(66.67)	2(100)	4(57.14)	10(66.67)
Percentage of Sales	1(33.33)	1(33.33)	-	3(40.86)	5(33.33)
Promotional Expenses by Competitor	-	-	-	-	-
Total	3	3	2	7	15

IV. DECIDING THE MEDIA

We have so far covered the advertising objective, deciding the advertising budget. We will now discuss the next decision area - deciding the media. Media as the very implies is nothing but a medium or channel for carrying the intended advertising message to the target audience.⁷ A manufacturer can make use of the following advertising media to spread his message to the people (1) press advertising (2) outdoor advertising (3) film advertising (4) radio advertising (5) television advertising (6) outdoor displays (7) catalogues (8) directories (9) direct mails (10) circulars etc.. Before assessing and evaluating the various media available, firm has to study in detail its target audience. It is essential for firm to understand the characteristics by the target customers, their braying habits, their lifestyles etc and after that reach, frequency and impact on target buyers of media selected.

Table 1.5

Distribution of Sample Units of Various Product Items Under Agro Based Industries acc. to Deciding the Advertising Media in Rajasthan.

Media Used	Product Items					
	Dal	Vegetable Oil	Rice	Cotton Pressing & Ginning	Guar Gum	Total
Radio	-	1	1	-	-	2(6.67)
Wall Writing	2	4	3	2(22.22)	-	1(3.33)
Outdoor Hoarding / Poster / Hand bills	2	1	-	3(33.33)	1(50)	12(40)
Newspaper / Magazine	-	1	1	3(33.33)	-	5(16.67)
Direct Mail	-	-	-	-	-	-
Directories, Cinema Slide	-	1	1	-	-	2(6.67)

Table 1.5 shows that the total number of sample units belonging to various product items under Agro based industries is using advertising as medium of promotion, overall 60%, units are engaged in advertising, 40% use wall writing / posters and 23.33% through outdoor hoarding, only 16.67% are concentrating on newspapers / magazines.

The computed value of chi square notation is $C^2_{0.5}=26.6$ which is lesser than the tabulated value 41.33 at 5% level of significance with degree of freedom $V=28$. This shows that the Hypothesis of association between various product items under Agro based industries & advertising media used is rejected.

Table 1.6 gives the media used position according to different product groups. It shows that of the total number of sample units, 46.67% use wall writing, the method is used by 66.67%. POP units followed by 50% chemical unit 33.33% of agro and bone grinding also follow this method.

The computed value of chi square notation is $C^2_{0.5}=10.24$ which is lesser than the tabulated value 32.67 at 5% level of significance with degree of freedom $V=21$. This shows that Hypothesis of association between various product groups & media used is rejected.

Table 1.6

Distribution of Sample Units of Various Product Groups acc. to Deciding the Advertising Media in Rajasthan

Media Used	Product Groups				
	POP	Agro Based	Chemicals	Bone Grinding	Total
Radio	-	-	-	-	-
Wall Writing	-	-	1(50)	-	1 ((6.67)
Outdoor Hoarding /Poster / Hand bills	2(66.67)	1(33.33)	1(50)	3(42.86)	7(46.67)
Newspaper /Magazine	-	1(33.33)	-	2(28.57)	4(26.67)
Direct Mail	-	-	-	1(14.28)	1(6.67)
Directories, Cinema slide	-	-	-	-	-
Any other	3	3	2	7	15(100)

V. PREPARATION OF ADVERTISING MESSAGE

The most creative part of advertising is planning an advertising campaign. Advertising campaign is a combination of messages communicated in a particular way to audiences to achieve advertising objectives. Organizations handle their advertising in different ways. In small companies it is handled by someone in sales or marketing department, who works with an advertising agency. A large company will often set up its own advertising department, whose manager reports to the vice president of marketing.

To bring out the relative use of two choices in preparing the advertising message, the sample responses have been presented according to product items under Agro based industries and different product groups.

Table 1.7

Distribution of Sample Units of Various Product Items Under Agro Based Industries acc. to Preparation of Advertising Message in Rajasthan.

Advertising Message	Product Items					
	Dal	Vegetable Oil	Rice	Cotton	Guar Gum	Total
Self or Own staff	4(100)	4(75)	5(71.43)	8(88.89)	2(100)	25(88.33)
Advertising Agency/ Professional Artists	-	2(25)	2(28.57)	1(11.11)	-	5(16.67)
Total	4	8	7	9	2	30(100)

Table 1.7 shows that 83.3% of units prepare their advertising message themselves or have it done by their staff & the remaining 16.7% use the services of advertising agencies. The computed value of chi-square notation is $C^2_{0.5} = 8$ which is lesser than the tabulated value 9.49 at 5% level of significance with degree of freedom $V=4$. This shows that the Hypothesis of association between various product items under Agro based industries and advertising message is rejected.

Table 1.8

Distribution of Sample Units of Various Product Groups acc. to Preparation of Advertising Message in Rajasthan.

Advertising Message	Product Groups				
	POP	Agro Based	Chemicals	Bone Grinding	Total
Self or Own staff	3(100)	2(66.67)	1(50)	5(71.4)	11(73.33)
Advertising Agency / Professional Artists	-	1(33.33)	1(50)	2(28.51)	4 (26.67)
Total	3	3	2	7	15(100)

Table 1.8 brings out that the proportion of units which prepare this advertising message themselves is highest in POP units followed by Bone grinding & Agro based. The computed value of chi-square notation is $c^2_{0.5} = 5.18$ which is lesser than tabulated value 9.49 at 5% level of significance with degree of freedom $V= 4$. This shows that the Hypothesis of association between various product groups and advertising message is rejected.

V1. ADVERTISING COURAGE

To maximize overall awareness, the maximum number of the target audience should be reached by the advertising. This is known as coverage. It is often defined in terms of geographical boundaries of the market identified as local, regional, national or international. When advertising leverage is so viewed, following results have been obtained.

Table 1.9

Distribution of Sample Units of Various Product Items Under Agro Based Industries acc. to Advertising Coverage in Rajasthan.

Coverage	Product Items					
	Dal	Vegetable Oil	Rice	Cotton Pressing & Ginning	Guar Gum	Total
Local	4(100)	6(75)	5(71.43)	8(88.89)	1(50)	24(80)
Regional	-	2(25)	2(28.5)	1(11.11)	1(50)	06(20)
National	-	-	-	-	-	-
International	-	-	-	-	-	-
Total	4	8	7	9	2	30(100)

The Table 1.9 gives the classification of sample units belonging to various product items under agro based industries according to their advertising coverage either being local, regional, national or international. It shows that 80% have local coverage and 20% have regional coverage. In case of units belonging to Dal 100% cover local areas.

The computed value of chi-square notation is $C^2_{0.5} = 8.05$ which is lesser than the tabulated value 21 at 5% level of significance with degree of freedom $V=12$. This shows that the Hypothesis of association between various product items under Agro based industries and advertising coverage is rejected.

VII. PERSONAL SELLING

Personal selling is an age old art of selling goods. It is the process of assisting and persuading a prospective buyer to buy a product in a face to face situation. "Salesmanship is the process whereby the seller ascertains and activates the needs or wants of the buyer and satisfies the needs or wants to mutual continuous advantage of both the buyer and the sellers".¹⁰

In most of the small scale businesses, personal selling plays a key role in the marketing effort. It is common with many of small entrepreneurs that they make this own affords through this process of marketing. In personal selling it is possible for the salesman to carry the qualified target consumer through a logical and persuasive reasoning process. Further the salesmen turn out to be a very good market researcher. He gathers information which often helps him in making timely decisions.

CONCLUSIONS

Promotion planning is a systematic decision making relating to all aspects of the development and management of a firm's promotional effect. It is a form of communication used by a firm to inform, persuade, or remind people about its products, services, image, ideas, community involvement, or impact on society. The promotion may be company sponsored or controlled by independent media. A firm's promotional plan usually stresses individual products and services with the objective of moving consumers from awareness to purchase.

Therefore, the management of small enterprises should realize that successful marketing requires careful attention to promotional efforts. The following are three

major decision areas in the field of promotion on which the entrepreneurs should focus their attention.

1. Personal selling, 2. Advertising 3. Sales promotion.

Personal selling is face to face contact between a seller's representative and those people whom the seller wants to communicate "Personal selling consists in individual personal communication in contrast to mass, relatively impersonal communication of advertising, sales promotion and other promotional tools."

Advertising is an important element of promotion Mix by which the markets provides information regarding products and services to mass public. "Advertising is any paid form of non-personal presentation & promotion of ideas, goods or services through mass media by an identified sponsor."

Third element sales promotions are specific efforts that are designed to have an immediate impact on sales. It refers to much kind of incentives and techniques directed toward consumers and traders with the intention to produce immediate sales effects.

To begin with we have brought out the relative importance of each of the three elements in various product groups.

After that importance of advertising is discussed. It is only through this form of promotion we can attract the attention of large audience, stimulate demand, develops brand preferences and finally contribute to the increase in the sales volume. The advertising expenditures are large, but the average cost of advertising in a firm is typically 1 to 3 percent of sales.

Most advertising money is spent on news papers, television and direct mail. Other frequently used advertising media are radio, magazine, yellow pages and outdoor displays. How much is to be spent on advertising is one of the crucial decisions to be taken by the advertising managers. There are three methods deciding budgets:-

- 1- Percentage of sales.
- 2- Capacity to afford.
- 3- Promotional expenses by competitor.

Depending upon their form, benefit and coverage, a firm can engage itself in advertising by using media, such as magazines & newspapers, radio and television, outdoor displays, direct mail, catalogues circulars etc.. Each of these media differs in terms of extent of its reach frequency and impact. The effectiveness of advertising as a tool of communication depends largely on the message which is intended to be conveyed. It is therefore necessary that adequate care is observed in developing the advertising message so that it meets the requirement of being desirable, acceptable and effective, large firms spend a huge amount of money on engaging the services of specialized agencies to get their advertising message prepared.

But, this practice is not common in Small scale industries. The reason being that these units are neither adequately aware of the role that advertising message plays in sales promotion, nor do they have the necessary channel to make use of specialized services. Therefore, the only alternative available to them is doing job themselves. And lastly, advertising coverage is analyzed. Advertising coverage refers to geographical area which a manufactures intends to reach through

advertising. It may be local, regional, national or international.

To sum up, it can be concluded that though advertising has many obvious advantages, only few units are in a position to use this method as a promotional technique.

Sales promotion is the catch wall of all various promotional tools that are not formally classifiable such as samples, coupons, trade fairs. Due to professionalism in recent years, sales promotion is increasingly viewed as an important tool in its own right. It plays a critical role in introductory and maturity stage of product life cycle and also appears to be especially effective during periods of rapid inflation. But compared to other two tools it has been less favoured by selected small units.

Thus, hereby we want to suggest that there is a vast scope for the units to be benefited by advertisement. It would be also in the fitness of things if they are able to pay attention on some of the sales promotion methods like samples coupons, contests and demonstrations.

LIMITATIONS

Despite our best effort the present study suffers from various limitations as detailed below:-

1. Rajasthan being the largest state in terms of area, so the study has not covered the entire population of SSI due to limitation of both resources and time;
2. It was observed that the owners – managers of some industrial units have shown reluctance and have refused to cooperate in furnishing the necessary data for study;
3. In some cases the information provided by them was so incomplete that it was difficult to use it for the analysis of results;
4. In the absence of sufficient studies on the marketing of Small scale industries sizeable secondary data for references on subject was also not available;
5. A number of sample units were not available as per address recorded in the Directorate of Industries;
6. Hardly any set of written policies are prepared by entrepreneurs;

REFERENCES

- [1] Boone and Kurtz, *Contemporary Marketing*, The Dryden press, Illinois, 1980, p.284.
- [2] R.Joel, Evans and Barry Bermon, *Essentials of marketing*, Mac Millan Publishing Co., New York, p.286.
- [3] Hans Scholl Hammer and Arthur H. Kurileff, *Entrepreneurship and Small Business Management*, John Wiley and Sons, New York, 1979, pp.184-185.
- [4] Philip Kotler, *Marketing management: Analysis, Planning, Implementation and Control*, Prentice- Hall, New Delhi, , 2000, p.578.
- [5] V.S. Ramaswamy, *Marketing Management*, Mac Millan India Ltd., New delhi 2006, p.535.
- [6] William J. Stanton, *Fundamentals of Marketing*, Mc Graw Hill Co., 1994, p.502
- [7] V.S. Ramaswamy, op. cit., p.551.
- [8] Philip Kotler, op. cit..
- [9] Czinkota Kolabe, *Marketing Management*, 2007, p.387.
- [10] Pederson and Wright, *Salesmanship*, p.27.
- [11] Peter Drucker, *Managing for results*, Allied publishers (Pvt.) Ltd., Bombay, 1970, pp. 21-24.
- [12] W. J. Stanton, op. cit., p.254
- [13] P.K. Agarwal, *Marketing Management*, Pragati Prakashan ,Meerut,2007,p.354.
- [14] P.O. Converse, *The Other Half of Marketing*, Proceedings of the Boston Conference on Distribution, 1954, pp. 22-25

UNDERSTANDING THE PERCEPTION OF TELEVISION ADVERTISEMENTS AMONG CHILDREN IN MUMBAI

Suvrashis Sarkar

Researcher for Ph.D. with Mumbai University
at Jamnalal Bajaj Institute of Management Studies.
Brand mgt. & mkt. communication professional.
Mumbai, India.
suvrashis@gmail.com

ABSTRACT

This study has been done with the main intention to examine the perception and understanding of television advertising among children of Mumbai. There is a growing use of children or children based creative theme in TV commercials recently, it was hence felt by the author to analyze what children themselves as target audience have a perception for modern day TV advertisements and what way has television advertisements impacted the lives of children in Mumbai.

Need for research :

Marketers see children as a rich and influential market segment because children have pester power and can influence the purchasing habits of other family members. Therefore the children's market has a huge opportunity for marketers and in India television happens to be one of the main media used for reaching children as target audience. However, concerns about the ethics and appropriateness of advertising to children have always been voiced, including academic study on this topic too. Despite such study no definite research on children's understanding of television advertising in urban India has been done, hence this study has been attempted in Mumbai.

Objective of the study:

1. *To study the basic understanding among children of Mumbai regarding the purpose of television advertisements*
2. *To investigate the variables of liking television advertisements among children in Mumbai*
3. *To investigate the effect of television advertisements in the life of children in Mumbai*

Research design and methodology:

A sample of 250 children including 130 boys and 120 girls studying in school grades 1 – 5 covering the approximate age group of 5 years to 11 years were personally interviewed in presence of their parents on August and September 2008 and again in July 2012 from various places of Mumbai and Thane. The selection of children was purely on random sampling for

primary data, most of the contact interviews happened in shopping malls and housing societies comprising of people from SEC B, B+ and A. Secondary data has been collected referring to published material on the work done in similar context by psychologists internationally.

Key words:

Children, urban India, modern consumers, television, advertisements

I. INTRODUCTION

Rapid economic growth has given rise to fundamental shift in the consumer spending patterns which in-turn has impacted manufacturers, marketers and retailers of consumer products and services. To compete successfully in this changing environment it has become extremely necessary to understand the rapid change in consumer behaviour and consumption patterns. In recent years, India has seen a remarkable transformation of consumerism. While family structure has modified and shrunk from joint to nuclear, income levels have gone up resulting in more disposable money finally resulting in a constantly changing aspiration of consumers. According to Euromonitor International, a new middle class is the emerging big group in the market with varied level of interests, dynamic behavior pattern and increasing disposable money at hand. Not surprisingly, there is an almost general agreement of the shift in dominance of global consumer markets from the advanced nations to emerging economies like India. According to Deloitte, this increasing middle-class is pushing a significant growth in consumer credit. The middle-class with heightened spending power has shown investments or spends in areas like healthcare, education, financial products, etc. Employment in cities has also resulted in the urban population growing exponentially. According to the National Institute of Urban Affairs report of 2011, the urban Indian population will be close to 600 million by 2031. It states that number of metro cities with population on 1 million plus has increased from 35 million in 2001 to 50 million in 2011 and is expected to increase to 87 million by 2031. These facts become more important for a marketer to understand since communicating with such a dynamic urban consumer in India becomes a challenge to connect

and sell your products. Accordingly so the evolution of modern day advertising plays a significant role to take away the mind share of this dynamic consumer base in urban India.

Among this dynamic consumer base the most important section is of children who have undoubtedly evolved in the last decade as the real “game changers” in the modern market.

Today advertising penetrates into the life of every person, including children. The children’s market is important to advertisers because of the enormous purchasing power of the children and their parents. A survey done two decades back including 2400 children aged 7–12 years in six countries, including China, Japan, France, UK, Germany and USA, indicated that their estimated annual spending power ranged from US\$1.7 billion in Germany to US\$11.3 billion in the USA. The survey results also indicated active participation of children in making family purchase decisions in a number of product categories. The newly emerging middle classes, high populations and one-child policies in some Asian countries mean that these ‘little kings’ are obtaining more and more influence over where the family money is spent. Research has suggested that children in Urban cities have been quick to make demands and exercise pester power. They are often the first to learn of new products from the mass media (Roedder John, D. 1999), Children have always considered television to be the most important information sources for new products. In India children are ideally considered in the age group of 0-13 years and this section comprise of approximately 34% of the country’s population. This again makes children as one of the vital part of the consumer market to understand and serve. Today how a marketer can sell his goods or services as got a major hurdle to cross through a child’s understanding, buy-in and influence in the family’s purchase decisions. Hence this study was conducted to understand from a marketer’s point of view how urban children perceive advertisements today. Media like print, outdoor, internet do not influence children much since they are less exposed to these however television remains the sole medium which has attracted children the most in everyday walk of life and definitely plays a significant role in influencing child consumers. Let’s take a look at a typical Indian kid who comes home tired after 7-8 grueling hours at school, and all he looks for on television is something which can make him laugh and feel fresh. A child spending an average of 4 hours per day watching television may be exposed to 15000 commercials every year. What do children in Mumbai know about television advertising and what do they learn from it? This study attempts to examine children’s response to television advertising for aspects relating to the communication process. This study will be of major interest to both marketers and public policy makers since marketers are always keen to know if their

advertising is effective while policy makers are concerned with protecting the interests of the child.

II. RISE OF TELEVISION ADVERTISING IN INDIA

It all started in July 1984 when India’s first soap opera “Hum Log” was aired and Nestle India sponsored it to promote its product

“Maggi Noodles”. Nestle paid the telecast fee and production cost of Hum Log and got 5 minutes of commercial time in exchange to advertise Maggi. That was the grand beginning in Indian television which was followed by the series of Nukkad, Buniyad, Katha Sagar, Khandan, Vikram Betal, Dada Dadi ki kahaniyan, etc which every child of that time including me has grown up watching. These serials were the only family soaps and hence children and adults together watched them, hence advertisements that these serials had in the breaks in between had equal audience of children and adults.

The time grew with 2-3 dedicated programmes for children which even adults watched equally along, hence there was no much difference in advertisements for children as a specific target audience in those days. After the Asian Games of 1982 and Los Angeles Olympics of 1984, Colour TV was the new age gadget that took off in India. This revolutionized further from DD to Satellite TV with cable connections as the modern option which has now metamorphosed to DTH having HD viewing experience.

Sustained economic growth in India in the last three decades has resulted in consumption moving decidedly beyond the basic survival needs of people. Beyond the broad estimates of spending by Indian consumers, it is very interesting to see how the spending priorities have changed over the years and may further change in the next few years.

figures: million people	2006	2011	2016 (estimate)
Access to internet	32	101	300
Smart phone users	2	48	178
House with a four-wheeler	5	12	19
Access to digital TV	9	82	364
Laptop usage	2	12	61
Engineers and MBAs graduating	0.5	1	2
Indians traveling over seas	8	12	16

Source: Technopak Analysis

It’s evident from this report that television will be the topmost priority for an Indian in the years to come. The overall television industry has grown from Rs.257 billion in 2009 to Rs. 297 billion in 2010, which is

expected to reach Rs.630 billion at a CAGR of 17% by 2015, according to FICCI & KPMG research. This growth drive in television industry has been largely credited to advertising.

The top 10 advertising sectors on Indian TV are:

Sectors	% share
Food and beverages	14
Personal care	13
Hair care	5
Telecom	4
Automobile	4
Fashion	4
Banking and insurance	3
Healthcare	3
Household appliances	3

Source: TAM AdEx (2010-11)

It is very important to note that among these top 10 advertising sectors on TV, most of them have used children in their communications and have attempted to create the child's appeal in the family of the target audience. But did children actually notice them and how much have they understood this particular exercise of marketers is definitely a concern for study, hence this research was attempted in an urban market of Mumbai, representing urban India.

III. WHAT CHILDREN FEEL ABOUT TELEVISION ADVERTISING

Many parents and educators considering children as naïve and vulnerable believe that advertisers exploit children's unsuspecting nature and lack of experience, and that advertising-induced demands lead to family tensions and conflict.

Much research in this area assumes that children lack the maturity needed to cope with advertising - a view of child as innocent, and advertiser as exploiter. This focus on children in terms of what they lack, and as more or less incompetent compared to adults, is heightened by the application of Jean Piaget's theory of child development. Furthermore, the prevalence of effects and linear-sequential models of influence has led to positive, quantitative research presenting children as passive receivers of advertising. Experimental methods and laboratory studies are the norm, with little attention paid to children's perspectives or their social and cultural backgrounds. The shortcomings of prior research were highlighted by Lawlor and Prothero (2002), who made a powerful case for adopting a child's perspective and a "non-commercial and meaning-based" approach when researching children's understanding of advertising intent. They noted for example that while researchers examined children's appreciation of advertisers' intent,

children themselves may not view advertising simply as a product purchasing catalyst. Methodological differences have led to much debate about the development of children's understanding with respect to advertising (Martin 1997, Gunter and Furnham 1998). Summarizing research in this area, Young (1998) highlighted that pre-school kids can make crude distinctions between TV advertising and programmes. The age at which children appear capable of relating the basic content of advertising to shopping ranges from two to five. Around the age of six, children realize that ads exist to inform as well as to amuse, and eight year olds tend to know that advertisers provide information in a rhetorical way. By middle to late childhood, there is usually an understanding of advertising's advocacy, informative, and rhetorical functions.

Jean Piaget's theory of cognitive development has had a great impact on research related to the communication process of advertising to children. The theory proposes that a child's ability to think and reason progresses through a series of distinct stages. Each stage has its own unique characterization, yet could vary considerably depending on intelligence, cultural background and socio-economic factors. Piaget (1970) proposed four stages of cognitive development that can be summarized as below:

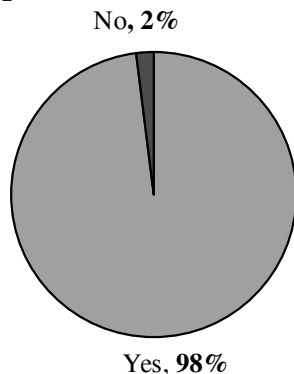
1. The sensorimotor stage (0–2 years): behaviour is primarily motor and the child does not manifest conceptual thinking.
2. The pre-operational stage (2–7 years): the child's thinking about objects and ideas is poorly organized and only dominant features are used for making judgments.
3. The concrete operations stage (7–11 years): the child can think conceptually and organize ideas in a coherent and stable manner.
4. The formal operations stage (11+ years): the child can think in abstract terms.

Many studies on children's communication processing of television advertising have been based on Piaget's cognitive development theory. Other approaches have also emerged to provide detailed explanations of Piaget's theory using concepts about information processing of short- and long-term memory of commercial information (Roedder, 1981). Studies have generally indicated that child's ability to distinguish programmes from commercials increased with age (Robertson and Rossiter, 1974; Meringoff and Lesser, 1980). Children's comprehension of television advertising and its persuasive intent also increased with age.

In this research, children were prompted about their awareness and knowledge of television advertising. They were always approached in presence of their parents and were asked open-end questions to enable them to open up and express their personal views. They were asked what advertisements are, what their purpose are and what they like in a TV ad.

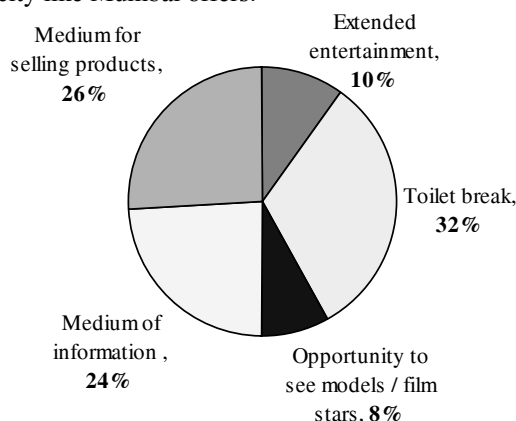
The actual findings of this research have been detailed as below;

On asking **do you like TV ads**, most of the children agreed of liking TV ads.



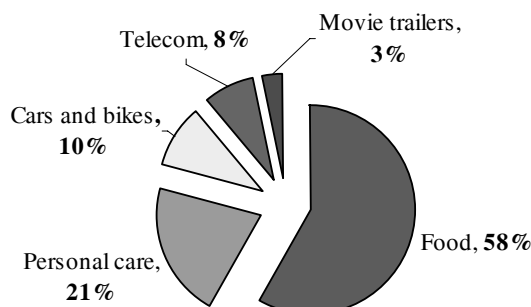
What does a TV ad mean to you?

This question was answered by most kids as a toilet interval in between their TV serial but the second most answered statement was that advertisement is a mode to sell company's products. This shows the sound understanding and maturity among modern urban children. This could be a result of education, family upbringing and also the modern lifestyle that an urban city like Mumbai offers.



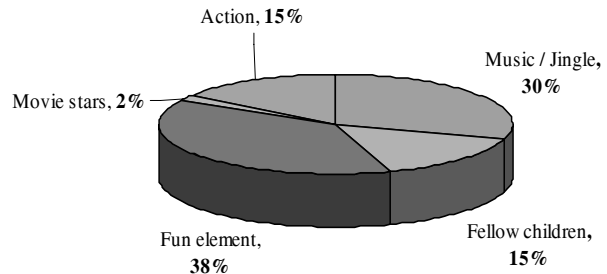
What type of TV ads do you like?

Food was the top favorite category of TV ads which the children responded. This category has been summarized which children replied as chocolates, snack foods, corn flakes, cold drinks, fast-food restaurants, etc



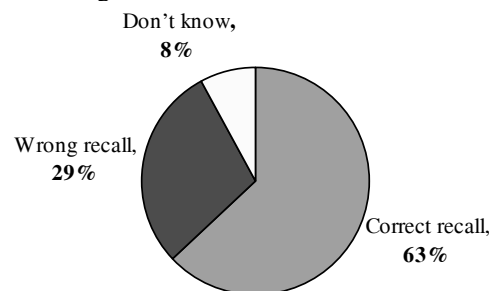
What are the variables that you enjoy to see in a TV ad?

Most of the children answered fun-element in a TV ad which attracts them the most, a particular hit music or jingle also is definite success factor among children followed by action elements like stunts, etc



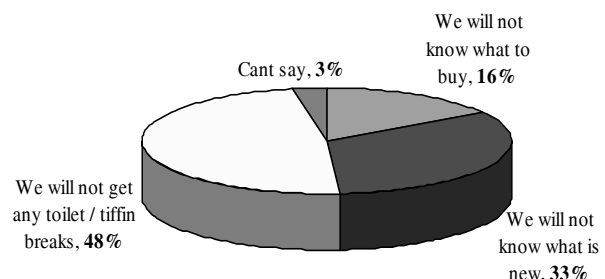
Brand recall?

Children were influenced to remember an ad through event description, jingle, etc and were tested for brand recall. Chocolates, fast-food joints, shampoos, car & bikes, telecom, movie trailers were the categories tested upon. To this most of the kids had shown positive brand recall proving the success of TV advertising among children as target audience.



To test the importance of TV advertisements in the life of urban children, they were asked to answer **what will happen if there are no TV ads from tomorrow?**

The importance of TV ads as a break or interval was strongly highlighted by most children followed by the strong fact that awareness of new products in the market will cease to exist without TV ads came out very well.



Since ads in between is seen as a disturbance in viewing the complete serial or programme comes up very often, children were asked about their views on **What will happen if ads come only in start and at the end, but not in middle of a programme ?**

The responses have been grouped as below:

We will enjoy un-interrupted programme	23%
We will not get any toilet / Tiffin break	54%
Programme will get boring to watch	14%
Confusion, don't know what to say	9%

These responses have definitely underlined that TV ads are strongly perceived as an interval from the monotony of continuous TV viewing and is definitely needed. This also proves that Children can not imagine TV without advertisements; also advertisements are a necessity in life today.

Since all the children were contacted and interviewed in presence of their parents, it was important to understand the parents views on what they felt **how has TV ad viewing affected their children?**

Children are better informed of what is new in market	15%
Children recognise the latest cars / bikes much better than us adults	23%
Children know which product is better	21%
Children decide which movie to watch and where to dine over the weekend	15%
Children are not only influencers but are also becoming decision makers in family purchase decisions	26%

CONCLUSION

It is evident from these results that television advertisements have been responsible for social maturity of children to a substantial extent and also urban children show sound understanding and positive perception towards advertisements. This could be due to education, family upbringing and also the modern lifestyle that an urban city like Mumbai offers.

The results also indicated that children are beginning to understand what advertising is and are aware of the persuasive intention of television advertising. Children from grade 4 onwards understood that television channels carried advertising for commercial value. The main reason for liking and disliking commercials depended on their entertainment element. An understanding of television advertising, recall of brands from slogans and comprehension of advertising content were consistently related to the cognitive development of children. Brand recognition from liked and disliked commercials was strong.

The results on children's cognitive responses to television advertising obtained in this study were similar to previous studies. A remarkable increase in

understanding what advertising is and the purpose of television advertising at the age of 7–8 years was also consistent with Jean Piaget's proposed stage on concrete operations. The children entering this stage of cognitive development were able to think conceptually and organized the selling idea in a more consistent manner. This indicates that Piaget's theory of cognitive development finds empirical support in urban Indian environment. According to Piaget's classification system, children at the age of 9 -10 years begin to enter the formal operations stage where they can think in more abstract terms. The findings have public policy implications that younger children need greater protection against persuasive marketing communication in the mass media. The findings that children liked food and beverage commercials were similar to those obtained by Yavas and Abdul-Gader (1993). This is probably because children are more familiar and experienced with the product category. Fun and enjoyment was the main reason for liking a commercial. Brand recognition for liked and disliked commercials was strong among children. Commercials enjoyed and disliked by children are more likely to be remembered. However, the influence of the brand information on purchase intention and purchase behavior needs further investigation. Advertisers also need to recognize the purpose for using a celebrity in children centric product. Advertisers should consider other strategies to enhance brand recognition such as incorporating brand names in slogans and package recognition. In this study children demonstrated a fairly good understanding of pro-social messages. This finding should be encouraging for government officials and academia. The government should produce more campaigns to cultivate positive social behavior and to protect children's interests, in a children-friendly way. It was found that the majority of children in Mumbai knew what advertising is and had developed an appreciation of television commercials. Children are also aware of the selling intention as well as the institutional function of television advertising. Understanding of television advertising, recall of brands from slogans and understanding of advertising content were clearly related to the cognitive development of children. Advertisers targeting at children should make the TV ad creative funny and interesting so that children like them and remember the brands. Understanding of the key messages of advertising content varies by children's cognitive development and the style of presentation, but urban children of Mumbai definitely understand what advertisements are what is the purpose of TV ads and what benefits TV ads have in their modern life-style.

ACKNOWLEDGEMENTS

I am extremely thankful to all the children respondents and their respective parents for their kind co-operation while interviewing them and for sharing their valuable time and opinion on my subject of research study.

REFERENCES

- [1] ASCI, (2008), *The code for self regulation in Advertising*. The Advertising Standards Council of India
- [2] Bandyopadhyay, S., Kindra, G. and Sharp, L. (2001) Is television advertising good for children? Areas of concern and policy implications. *International Journal of Advertising* **20**(1), 89–116.
- [3] BusinessWorld, (2012-13), *The Marketing White-Book*
- [4] Baron, Rober A., Branscombe, Nyla R., Byrne, Donn., Bharadwaj, Gopa., (2010), *Social Psychology for the University of Mumbai.*, Pearson
- [5] Credit Suisse (2011), *Emerging Consumer Survey*
- [6] Donohue, T.R., Henke, L.L. and Donohue, W.A. (1980) Do kids know what TV commercials intend? *Journal of Advertising Research* **20**, 51–7.
- [7] Euromonitor International
(<http://blog.euromonitor.com/2010/03/emerging-focus-rising-middle-class-in-emerging-markets.html>)
- [8] Lawlor, M. and Prothero, A. (2002), "The Established and Potential Mediating Variables in the Child's Understanding of Advertising Intent: Towards a Research Agenda", *Journal of Marketing Management*, Vol. 18, No. 5-6, pp.481-500
- [9] Piaget, J. (1970) The stages of the intellectual development of the child. In P.H. Mussen, J.J. Conger and J. Kagan (eds) *Readings in Child Development and Personality*. New York: Harper & Row, pp. 291–8.
- [10] Robertson, T.S. and Rossiter, J.R. (1974) Children and commercial persuasion: an attribution theory analysis. *Journal of Consumer Research* **1**, 12–20.
- [11] Roedder, D.L. (1981) Age differences in children's responses to television advertising: an information processing approach. *Journal of Consumer Research* **8**, 144–53.
- [12] Roedder John, D. (1999), "Consumer Socialization of Children: A Retrospective Look at Twenty-five Years of Research", *Journal of Consumer Research*, Vol. 26, December, pp.183-213
- [13] Rubin, R.S. (1974) The effects of cognitive development on children's response to television advertising. *Journal of Business Research* **2**, 409–19.
- [14] Sheikh, A.A., Prasad, V.K. and Rao, T.R. (1974) Children's TV commercials: a review of research. *Journal of Communication* **24**(4), 126–36.
- [15] Smith, P. K., Cowie, H. and Blades, M. (1998), *Understanding Children's Development*, Oxford, Blackwell Publishers
- [16] Swamy, RK, Media Group (2011-12), *Media Market Guide*
- [17] Wood, D. (1988), *How Children Think and Learn: The Social Contexts of cognitive Development*, Oxford, Blackwell
- [18] Yavas, U. and Abdul-Gader, A.A. (1993) Impact of TV commercials on Saudi children's purchase behaviour. *Marketing Intelligence and Planning* **11**(2), 37–43.
- [19] Young, B. (1990), *Television Advertising and Children*, Oxford

Guidelines for Manuscript Submission

- Paper in IEEE two column format can be submitted for possible publication in the “Applied Engineering and Technologies ”Journal of VIT. The template for the IEEE two column format is available at: <http://www.vit.edu.in> in Instructions to authors link
- The paper should be submitted in MS word format to editor.aet@vit.edu.in
- Authors of the accepted papers have to sign the copyright form which is available at: <http://www.vit.edu.in> in Instructions to authors link
- For more information and queries refer <http://www.vit.edu.in> or email at: editor.aet@vit.edu.in

Topic	Pg. No.
• Face Recognition using Block Selection of Wavelet Transform	85
• Bandwidth Enhancement of low profile air fed Micro strip Antennas	90
• Fractal Antenna Analysis & Application Overview	95
• Study of reliable MPPT in PV system using kalman filter	99
• Route Optimization in Mobile MPLS	103
• Image processing approach on detection of brain tumor cells	109
• Seamless Stitching of Images Using Normalized Cross Correlation	113
• Multilevel decision system for the data compression of the xray arm medical images	118
• A New Method for Obtaining an Optimal Solution For Transportation Problems	125
• H.264 Video Encryption Algorithm for High-End Mobile Devices	128
• Asic Verification of learning and aging process in Ethernet bridge	132
• Temperature Prediction using Artificial Neural Network	138
• Genetic Approach for Fuzzy Mining Using Modified K-Means Clustering	142
• A Survey Paper on Management of Risk for Different Models of Software Engineering	148
• A Bayesian Iterative Reconstruction with Computed Tomography for maximum convergence of scattered data	154
• Comparative analysis of Auscultation Technique	162
• Differentiated Instruction and its Implication in motivating L2 writing	168
• Adaptability - A competent skill to harness the power of Effective Communication	177
• Promotional Mix Policies and Practices in Small Scale Industries in Rajasthan	180
• Understanding the perception of Television advertisements among children in Mumbai	185