January - June, 2014 Vol. 6 No. 1 ISSN 0973-5658

BIJIT

Impact Factor 0.605 and Index Copernicus Value (ICV) 4.75

Indexed with INSPEC (UK), Index Copernicus (Poland), ProQuest (UK), EBSCO (USA), Google Scholar (USA)

BVICAM's International Journal of Information Technology

CONTENTS

Special Section on Fuzzy Logic

1. **Estimation of f-validity of Geometrical Objects with OWA Operator Weights** 655
Abdul Rahman and M.M. Sufyan Beg

2. **Optimal Enactment of a Stand-alone Hybrid Wind-Fuel Cell Based Distributed Generation System Through Fuzzy Logic Control** 664
Mohammad Saad Alam

3. **Use of Wavelet-Fuzzy Features with PCA for Image Registration** 672
Safia Sadruddin and Rashid Ali

4. **Zadeh-Deshpande Approach for Fuzzy Description of Air and Water Quality** 677
Ashok Deshpande, Jyoti Yadav and Vilas Kharat

5. **Business Analysis and Decision Making Through Unsupervised Classification of Mixed Data Type of Attributes Through Genetic Algorithm** 683
Rohit Rastogi, Saumya Agarwal, Palak Sharma, Uarvarshi Kaul and Shilpi Jain

General Section

6. **Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Adhoc Networks** 690
Alka Chaudhary, V.N. Tiwari and A. Kumar

7. **Fuzzy Classification System by Self Generated Membership Function using Clustering Technique** 697
Shruti S. Jamsandekar and Ravindra R. Mudholkar

8. **A New Extended Kalman Filtering For Shadow/Fading Power Estimation in Mobile Communications** 705
George P. Pappas, Subramaniam Ganesan and Mohamed A. Zohdy

9. **Thermal Model and Characteristics of Double Slope Solar Still** 711
Rahul Dev and Amimul Ahsan

10. **Video Denoising Based on Stationary Wavelet Transform and Center Weighted Median Filter** 722
Soundarya K

Bharati Vidyapeeth's
Institute of Computer Applications and Management
A-4, Paschim Vihar, Rohtak Road, New Delhi-63

Email : bijit@bvicam.ac.in, Website : http://www.bvicam.ac.in

Volume 6, Number 1, January – June, 2014

Special Issue on Fuzzy Logic

BIJIT - BVICAM's International Journal of Information Technology is a half yearly publication of Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4, Paschim Vihar, Rohtak Road, New Delhi – 110063 (INDIA).

ISSN 0973 – 5658

Disclaimer

The opinions expressed and figures provided in the Journal; BIJIT, are the sole responsibility of the authors. The publisher and the editors bear no responsibility in this regard. Any and all such liabilities are disclaimed

All disputes are subject to Delhi jurisdiction only.

Address for Correspondence:

Prof. M. N. Hoda

Editor-in-Chief, BIJIT

Director, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM),

A-4, Paschim Vihar, Rohtak Road, New Delhi – 110063 (INDIA). Tel.: +91 – 11 – 25275055

Fax: +91 – 11 – 25255056; E-Mail: bijit@bvicam.ac.in, Visit us at www.bvicam.ac.in/bijit

Published and printed by Prof. M. N. Hoda, Editor-in-Chief, BIJIT and Director, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4, Paschim Vihar, New Delhi – 110063 (INDIA). Tel.: +91 – 11 – 25275055, Fax: +91 – 11 – 25255056. E-Mail: bijit@bvicam.ac.in; mca@bvicam.ac.in; Visit us at www.bvicam.ac.in/bijit

Our Indexing at International Level

The **INSPEC, IET (UK),** formerly IEE (UK), database is an invaluable information resource for all scientists and engineers, that contains 13 million abstracts and specialized indexing to the world's quality research literature in the fields of physics and engineering. *For further details, click at* http://www.theiet.org/resources/inspec/

Index Copernicus International (Poland) is a journal indexing, ranking and abstracting site. This service helps a journal to grow from a local level to a global one as well as providing complete web-based solution for small editorial teams. ICV 2012 for the BIJIT is 4.75. *For further details, click at* *http://jml2012.indexcopernicus.com/BVICAMs+International+Journal+of+Information+Technology,p4852,3.html*

ProQuest (UK) connects people with vetted, reliable information. Key to serious research, the company has forged a 70-year reputation as a gateway to the world's knowledge – from dissertations to governmental and cultural archives to news, in all its forms. *For further details, click at* *http://www.proquest.co.uk/en-UK/default.shtml*

EBSCOhost Electronic Journals Service (EJS) *is a gateway to thousands of e-journals containing millions of articles from hundreds of different publishers, all at one web site. For further details, click at http://www.ebscohost.com/titleLists/tnh-coverage.htm*

Open J-Gate *is an electronic gateway to global journal literature in open access domain. Launched in 2006, Open J-Gate is aimed to promote OAI. For further details, click at http://informindia.co.in/education/J-Gate-Engineering/JET-List.pdf*

DOAJ *aims at increasing the visibility and ease of use of open access scientific and scholarly journals, thereby promoting their increased usage and impact. For further details, click at*
http://www.doaj.org/doaj?func=issues&jId=87529&uiLanguage=en

Google Scholar *provides a simple way to broadly search for scholarly literature and repositories from across different parts of the world. For further details, click at http://scholar.google.com/scholar?hl=en&q=BIJIT%2BBVICAM&btnG=*

Cabell's Directory *of Publishing Opportunities contains a wealth of information designed to help researchers and academics, match their manuscripts with the scholarly journals which are most likely to publish those manuscripts. For further details, click at https://ssl.cabells.com/index.aspx*

Academic Journals Database *is a universal index of periodical literature covering basic research from all fields of knowledge. For further details, click at http://journaldatabase.org/journal/issn0973-5658*

Indian Citation Index (ICI) *is an abstracts and citation database, with multidisciplinary objective information/knowledge contents from about 1000 top Indian scholarly journals For further details, click at*
http://www.indiancitationindex.com/htms/release_notes.htm

and many more..., for more details click at http://www.bvicam.ac.in/BIJIT/indexing.asp

Editorial Board

Editorial

It is a matter of both honor and pleasure for us to put forth the eleventh issue of BIJIT; the BVICAM's International Journal of Information Technology. It presents a compilation of ten papers that span a broad variety of research topics in various emerging areas of Information Technology and Computer Science. Some application oriented papers, having novelty in application, have also been included in this issue, hoping that usage of these would further enrich the knowledge base and facilitate the overall economic growth. This issue again shows our commitment in realizing our vision "to achieve a standard comparable to the best in the field and finally become a symbol of quality".

As a matter of policy of the Journal, all the manuscripts received and considered for the Journal, by the editorial board, are double blind peer reviewed independently by at-least two referees. Our panel of expert referees posses a sound academic background and have a rich publication record in various prestigious journals representing Universities, Research Laboratories and other institutions of repute, which, we intend to further augment from time to time. Finalizing the constitution of the panel of referees, for double blind peer review(s) of the considered manuscripts, was a painstaking process, but it helped us to ensure that the best of the considered manuscripts are showcased and that too after undergoing multiple cycles of review, as required.

Over the years, Fuzzy Logic has changed our lives in a significant way. In a narrow-sense, Fuzzy Logic is considered to be a logical system, which is a generalization of multi-valued logic. A very important distinguishing feature of Fuzzy logic is that in Fuzzy logic, everything is, or is allowed to be, a matter of degree. Furthermore, the degrees are allowed to be fuzzy. In a broader sense, however, Fuzzy Logic is much more than a logical system. In fact, Fuzzy Logic is a precise system of reasoning and computation in which the objects of reasoning and computation are classes with unsharp boundaries. What is not widely recognized, within the scientific community and the general public, is that

Fuzzy Logic has become a vast enterprise. There are over 2,80,000 papers in the literature with Fuzzy in title. There are 25 journals with fuzzy in title. There are close to 25,000 Fuzzy-Logic related patents issued or applied for in the United States of America and Japan. There is a long list of applications ranging from digital cameras to fraud detection systems. Particularly worthy of note, on one end, is the Fuzzy Logic subway system in Sendai, a city of over 1 million in Japan. On the other end, numerically, is Omron's 120 million fuzzy logic blood pressure meters.

Some critics have been saying that Fuzzy Logic is a passing fad. This assessment of Fuzzy Logic fails to recognize that the world we live in is, in large measure, a world of Fuzzy classes, and that science has much to gain from shifting its foundation from classical Aristotelian logic to fuzzy logic.

It is on the above note that we decided to bring out a Special Issue, ***to celebrate the Golden Jubilee Year of Fuzzy Logic****, in the year 2014 – the 50th year of the introduction of* ***Fuzzy Logic by its Father, Lotfi A. Zadeh****, in the year 1965. For this noble initiative of BIJIT, Prof. M. M. Sufyan Beg, Professor, Dept. of Computer Engineering, A.M.U., Aligarh, who has done his Post Doctorate under Prof. Lotfi A. Zadeh at University of California, Berkley, has very kindly agreed to be the Guest Editor of this Special Issue on Fuzzy Logic. A very special thanks to him for having taken pain and finalized the papers for the Special Issue.*

The Ten papers, including the five papers under the Special Section, that were finally published were chosen out of eighty seven papers that we received from all over the world for this issue. We understand that the confirmation of final acceptance, to the authors / contributors, sometime is delayed, but we also hope that you concur with us in the fact that quality review is a time taking process and is further delayed if the reviewers are senior researchers in their respective fields and hence, are hard pressed for time.

We further take pride in informing our authors, contributors, subscribers and reviewers that the journal has been indexed with some of the world's leading *indexing / bibliographic agencies like* ***INSPEC*** *of IET (UK) formerly IEE (UK),* ***Index Copernicus International*** *(Poland) with* ***IC Value 4.75*** *for 2012,* ***ProQuest*** *(UK),* ***EBSCO*** *(USA),* ***Open J-Gate*** *(USA),* ***DOAJ*** *(Sweden),* ***Google Scholar, WorldCat*** *(USA),* ***Cabell's Directory*** *of Computer Science and Business Information System (USA),* ***Academic Journals Database, Open Science Directory, Indian Citation Index, etc.*** *and listed in the libraries of the world's leading Universities like* ***Stanford University, Florida Institute of Technology, University of South Australia, University of Zurich****, etc. Related links are available at http://www.bvicam.ac.in/bijit/indexing.asp. Based upon the papers published in the year 2012, its* ***Impact Factor*** *was found to be* ***0.605****. These encouraging results will certainly further increase the citations of the papers published in this journal thereby enhancing the overall research impact.*

We wish to express our sincere gratitude to our panel of experts in steering the considered manuscripts through multiple cycles of review and bringing out the best from the contributing authors. We thank our esteemed authors for having shown confidence in BIJIT and considering it a platform to showcase and share their original research work. We would also wish to thank the authors whose papers were not published in this issue of the Journal, probably because of the minor shortcomings. However, we would like to encourage them to actively contribute for the forthcoming issues.

The undertaken Quality Assurance Process involved a series of well defined activities that, we hope, went a long way in ensuring the quality of the publication. Still, there is always a scope for improvement, and so, we request the contributors and readers to kindly mail us their criticism, suggestions and feedback at bijit@bvicam.ac.in and help us in further enhancing the quality of forthcoming issues.

Editors

Guest Editor's Page

Fuzzy Logic or No Fuzzy Logic?

Human life has undergone a considerable change with the advent of Fuzzy Logic. However, there has been a stiff resistance to this change all this while. Interestingly, the issue of Fuzzy Logic has seen a crisp division between those who believe in it completely and those who do not believe in it at all. In one of his posts dated November 5, 2009, the Father of Fuzzy Logic, Prof. Lotfi Asker Zadeh says the following.

Dear members of the BISC Group:

Fuzzy logic is associated with a long history of discussion, debate and questioning. It is of interest to take a look at two contrasting views expressed at a time when fuzzy logic was in its early stages of development, and to judge, in retrospect, which view was right.

In l975, Professor William Kahan, a colleague of mine, a superior intellect and a Turing Prize winner, had this to say.

"Fuzzy theory is wrong, wrong and pernicious. I cannot think of any problem that could not solved by ordinary logic. What Zadeh is saying is the same sort of things "Technology got us into this mess and now it can't get us out." "Well, technology did not get us into this mess.Greed and weakness and ambivalence got us into this mess. What we need is more logical thinking not less.The danger of fuzzy theory is that it will encourage the sort of imprecise thinking that has brought us so much trouble."

In l976, in Japan, Hitachi and Kawasaki Heavy Industries began to explore the feasibility of employing fuzzy logic to build an automated subway system in the city of Sendai. The project was launched in l979. The fuzzy-logic-based automated subway system began to operate in l987. I rode on the system on the first day of its operation and was greatly impressed by the remarkable achievement of Japanese engineers and scientists. The system is an acknowledged success. A similar system is in operation on the O-edo line in Tokyo. What is particularly remarkable about the Sendai subway system is that it was conceived long before fuzzy logic became well-established. Subway systems are not toys. Those who conceived, designed and built the Sendai subway system deserve a loud applause.

In retrospect, which view was right?

Regards to all.

--
Lotfi A. Zadeh

Professor in the Graduate School and Director, Berkeley Initiative in Soft Computing (BISC)

It is solely to the credit of Prof. Zadeh that he withstood the pressure of his opponents to let Fuzzy Logic see the light of the day finally. This exhibits not only the confidence but also the firmness of the belief that Prof. Zadeh had in the theory founded by him. To me, it appears that the flamboyant success seen by digital computers (which, in turn, were based on binary logic), made it difficult for people to abandon the premises of binary logic and embrace Fuzzy Logic, even partially.

For those who believe that probability theory can effectively be employed to develop any and every model, Prof. Zadeh has the following thought provoking post dated December 11, 2013.

Dear Michael:

Many thanks for your constructive comment. I read with interest what you wrote. I like the way in which you articulate your views. Following is my quick response. In regard to Bayesianism, my view is summarized in a recent presentation. Here it is.

Bayesianism has many incarnations. In my view, a view which has no claim to be the mainstream view, the core of Bayesianism is the Bayesian doctrine: Any kind of uncertainty can and should be dealt with through the use of probability theory. In the spirit of this doctrine, an eminent Bayesian, Dennis Lindley, had this to say in 1987.

The only satisfactory description of uncertainty is probability. By this I mean that every uncertainty statement must be in the form of a probability; that several uncertainties must be combined using the rules of probability; and that the calculus of probabilities is adequate to handle all situations involving uncertainty...probability is the only sensible description of uncertainty and is adequate for all problems involving uncertainty. All other methods are inadequate...anything that can be done with fuzzy logic, belief functions, upper and lower probabilities, or any other alternative to probability can better be done with probability.

What this statement implies is that probability theory is applicable not only to its traditional domain---repetitive random events---but, more generally, to any event, random or not, repetitive or not. To this end, in Bayesian probability theory, the concept of probability is defined very broadly: Given all available information, the probability of an event is the strength of belief that the proposition which described the event is true. This definition has serious flaws. First, the definition is vague, that is, it is insufficiently specific. How can different kinds of information be combined into a numerical value of strength of belief. More importantly, the definition applies only to crisp propositions; it does not apply to fuzzy propositions, that is, propositions which contain fuzzy predicates and/or fuzzy quantifiers, and/or fuzzy probabilities. Typically, a proposition drawn from a natural language is a fuzzy proposition. Example. Most Swedes are tall. When the Bayesian definition of probability is applied to a fuzzy event/proposition, probability becomes confused with truth. The Bayesian definition of probability is not an operational definition in real-world settings. Is it ever used?

I would like to develop a better understanding of your position. To this end, I should like to pose a few basic questions.

First, how do you define probability? Second, how do you apply your definition of probability to fuzzy/vague events? Third, how would you apply your definition of probability to a simplified version of one of my test problems: X is a real-valued variable. What is known about X is that X is much larger than a and much smaller than b. What is the probability that X is approximately c? Much larger, much smaller and approximately are labels of fuzzy relations with specified membership functions. If you prefer, you can assume that fuzzy, vague words are labels of specified probability distributions.

I would appreciate your concrete answers. Comments are welcome.

Regards,
--
Lotfi A. Zadeh

It may not be correct to dismiss Fuzzy Logic as just being a Multi-valued Logic either. In fact, it is much more than that, as is vividly described by Prof. Zadeh in his following post dated May 23, 2014.

Dear Apostolos,

Thank you for your comment. A clarification is in order. Fuzzy logic is much more than a logical system. Basically, fuzzy logic is a system of reasoning and computation in which the objects of reasoning and computation are classes with unsharp (fuzzy) boundaries. The conceptual structure of fuzzy logic is much broader than the conceptual structure of multivalued logic. In multivalued logic, truth is a matter of degree. In fuzzy logic, everything--including degrees--is or is allowed to be a matter of degree. Truth values and degrees can be linguistic. Examples. Quite true, more or less true, high, low, medium, etc. In multivalued logic there is no concept of a fuzzy quantifier, e.g., most, many, few, many more. In fuzzy logic, such quantifiers and relations are numbers which can be computed with. In multivalued logic, there is the concept of a generalized quantifier(Mostowski 1957), but generalized quantifiers cannot be computed with. In fuzzy logic, quantifiers expressed as $most^2$, $most^{1/2}$ *are admissible; in multivalued logic, they are not. In fuzzy logic, we have the concept of a linguistic variable--a concept which is not in multivalued logic. In fuzzy logic, we have the concept of linguistic probabilities and possibilities, e.g., likely, very likely, usually, quite possible, etc. Lacking these concepts, multivalued logic does not have the capability to represent the meaning of propositions drawn from a natural language. Examples. Most Swedes are tall. It is likely that Vera is middle-aged. Usually temperature is low. These concepts are lacking not just in multivalued logic, but in all existing logical systems other than fuzzy logic. For this reason, no existing logical system--other than fuzzy logic--has the capability to reason and compute with propositions drawn from a natural language. This limitation of existing logical systems--other than fuzzy logic--is unrecognized within the natural language communities. This basic limitation of existing logical systems--other than fuzzy logic is unrecognized within the natural language communities.*

Your statement that "fuzzy logic is multivalued logic," is in need of clarification. The word "is" has two different meanings--everyday meaning and technical meaning. More concretely, technically the proposition, X is A, means that X takes values in A or, equivalently, that A is the possibility distribution of X. In this sense, fuzzy logic is not multivalued logic. Your statement is correct if it is interpreted as: Truth values in fuzzy logic are a matter of degree, as they are in multivalued logic. What should be underscored, as pointed out earlier, in fuzzy logic everything is or is allowed to be a matter of degree.

Comments are welcome.

Regards,

Lotfi

Hence, the purpose of this editorial is basically to bring out the aforementioned thought provoking posts from the very founder of Fuzzy Logic, for the academic community to ponder and hence agree to the utility of Fuzzy Logic, if possible.

--

(Prof. M. M. Sufyan Beg)
B.Tech. (AMU Aligarh, Gold Medalist), M.Tech. (IIT Kanpur), Ph.D. (IIT Delhi), Postdoc (UC Berkeley)
Department of Computer Engineering
A.M.U. Aligarh – 202002, Uttar Pardesh (INDIA)
mmsbeg@eecs.berkeley.edu

CONTENTS

Special Section on Fuzzy Logic

1. **Estimation of f-validity of Geometrical Objects with OWA Operator Weights** 655
Abdul Rahman and M.M. Sufyan Beg

2. **Optimal Enactment of a Stand-alone Hybrid Wind-Fuel Cell Based Distributed Generation System Through Fuzzy Logic Control** 664
Mohammad Saad Alam

3. **Use of Wavelet-Fuzzy Features with PCA for Image Registration** 672
Safia Sadruddin and Rashid Ali

4. **Zadeh-Deshpande Approach for Fuzzy Description of Air and Water Quality** 677
Ashok Deshpande, Jyoti Yadav and Vilas Kharat

5. **Business Analysis and Decision Making Through Unsupervised Classification of Mixed Data Type of Attributes Through Genetic Algorithm** 683
Rohit Rastogi, Saumya Agarwal, Palak Sharma, Uarvarshi Kaul and Shilpi Jain

General Section

6. **Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Adhoc Networks** 690
Alka Chaudhary, V.N. Tiwari and A. Kumar

7. **Fuzzy Classification System by Self Generated Membership Function using Clustering Technique** 697
Shruti S. Jamsandekar and Ravindra R. Mudholkar

8. **A New Extended Kalman Filtering For Shadow/Fading Power Estimation in Mobile Communications** 705
George P. Pappas, Subramaniam Ganesan and Mohamed A. Zohdy

9. **Thermal Model and Characteristics of Double Slope Solar Still** 711
Rahul Dev and Amimul Ahsan

10. **Video Denoising Based on Stationary Wavelet Transform and Center Weighted Median Filter** 722
Soundarya K

Special Issue of BIJIT on Fuzzy Logic
BIJIT - BVICAM's International Journal of Information Technology
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Estimation of f-validity of Geometrical Objects with OWA Operator Weights

Abdul Rahman[1] and **M. M. Sufyan Beg**[2]

Submitted in October 2013; Accepted in February, 2014

Abstract - In the age of sophisticated crimes and terrorism, there is a requirement to develop a perception based multi criteria decision making system that can reveal the hidden clues in the environment of uncertainty. The crime has no fixed dimension to be carried out in. But, still there remain some imprecise clues for the crime site investigation team and precise interpretation of these clues is impossible. That is the place where role of proposed extended fuzzy logic (FLe) comes into play. Moreover, for decision making, crime site investigation team has to consider multi criteria with different weights under the uncertainty. So, Extended Fuzzy Logic and Ordered weighted Averaging (OWA) may be taken together as a double folded milestone in revealing the uncertainty in the world of computational forensic. The concept of unprecisiated fuzzy logic (Flu) was introduced by Zadeh. When a perfect solution cannot be given or process falls excessively costly then the role of concept of Flu comes into play. This novel concept provides the basis for FLe. In order to have a better understanding of Flu, the concept of fuzzy geometry (f-geometry) is introduced. The proposed work is based on Sketching with Word technique. We have introduced some fuzzy theorems (f-theorems) in proposed work. These f-theorems can be used for estimating the membership value of fuzzy objects in f-geometry. These f-objects may play vital role for identifying clues in computational forensic.

Index Terms – Fuzzy Theorem, Fuzzy Similarity, Fuzzy Validity, Fuzzy Geometry, OWA

1.0 INTRODUCTION

In computational forensics when exact interpretation of imprecise information at crime site is impossible e.g. clue left behind by criminal such as finger prints, shoe prints, and face-sketch drawn by experts on the basis of onlooker's statement, then role of extended fuzzy logic (FLe) [1][2] comes into play. Face-sketch drawn by experts is a matter of degree of perception rather than measurement. Sometimes forensic experts have to make decision on the basis of multiple criteria. The ordered weighted averaging (OWA) provides a unified decision making platform under the uncertainty [3]. The above said problem is attracting attention of researchers and scientists to merge the concept of extended fuzzy logic and OWA in forensic science. So, Extended Fuzzy Logic and OWA may be taken together as a double folded milestone in revealing the uncertainty in the world of computational forensics. Hence in this work the concept of Flu is used for the fuzzy proof (f-proof) of some complex f-theorems. Further, the results have improved by using the concept of OWA.

Fuzzy Logic, first introduced by L. A. Zadeh, provides a precise conceptual system of reasoning where information at hand is imperfect. When a perfect solution cannot be given or process falls excessively costly then the role of concept of unprecisiated fuzzy logic (Flu) comes into play. The concept of Flu was also introduced by Zadeh. This novel concept provides the basis for FLe. In order to have a better understanding of Flu, the concept of f-geometry is provided in literature [1][2][4][5]. The f-geometry is a counter part of Euclidian geometry in crisp theory. In f-geometry, figures are drawn by free hand. In Flu, there is concept of perception based fuzzy valid (f-valid) reasoning. In [4]-[6] authors apply novel approach for the estimation of perceptions in geometric shapes, which become the basis for the complex shape. For all the geometric shapes, the estimation of membership function is done on the basis of perception. The proposed work is based on sketching with words technique. In [7] authors have applied Yager's OWA weights for aggregating different components of f-objects. Whereas in [8], Minimizing Distance form Extreme Point OWA model is used for estimating the fuzzy validity (f-validity) of fuzzy Rhombus.

This paper is organized as follows. In Section 2, we have briefly looked into the related work. In section 3, we have discussed the basic f-objects of f-geometry followed by their estimation. The section 4 incorporates some of the existing f-theorems. In section 5, we have proposed some f-theorems and their respective f-proofs. The section 6 consists of the experimental work. The final section 7 comprises conclusion and future directions.

2.0 RELATED WORK

In [9], for fuzzy image processing membership values are assigned to some of the properties like brightness, edginess, homogeneity etc. of an image, in case of fuzziness and then in defuzzifying it to appropriate gray levels for computation. In the above work, application of fuzzy geometry is done with feature extraction, feature segmentation, and feature representation. There are number of factors which demarcates f-geometry from Poston's fuzzy geometry [10], coarse geometry [11], Rosenfeld fuzzy geometry [12], Buckley and Eslamis's fuzzy geometry [13], Mayburov's fuzzy geometry [14], and Tzafesta's fuzzy geometry [15]. The major one is that FLe allows f-valid reasoning based on perception in place of allowing p-valid reasoning which is based on measurement.

[1, 2]*Department of Computer Engineering, Jamia Millia Islamia (A Central University), New Delhi-110025, India*
E-mail: [1]*rahman.jhansi@gmail.com and* [2]*mmsbeg@cs.berkeley.edu*

In [16], OWA was applied to find the importance of weight for each of the document viewed by the user in web searching. With many aggregation methods introduced in fuzzy information processing tasks [17,18], the OWA operators are mainly described in detail in [3], has been applied in many applications including fuzzy logic controller [19], market analysis[20], image compression in [21]. In [22], environmental indices were developed using fuzzy numbers OWA operators, for finding similarity with multiple linguistic parameter as inputs. In [23], content based image retrieval from XML-based DBMS unites some features in an image; indexing structure uses Euclidean distance for individual feature is used. Moreover, ordered weighted averaging is used to aggregate the distance function of the features, support nearest neighbor and fuzzy queries. In [24], similarities among images are computed for retrieving similar images from the database that combines weighted averaging, Choquet Integral, and relevance feedback for a better performance. However, there has been much of work in image retrieval but with very little intelligence to recognize fuzzy objects and image. In [4],[7] and [8] OWA is applied for estimating f-validity of different f-objects.

3.0 FUZZY GEOMETRY

Flu introduces the concept of f-geometry. In Euclidean geometry crisp concept C corresponds to fuzzy concept f-C, in f-geometry. The fuzzy geometric shapes have the following elements.

3.1 fuzzy Points

In f-geometry, a point is said to be f-point, if radius is not exactly zero, but has haziness please refer to Fig 1(a). The membership function of f-point is shown in Fig. 1(b). The membership value of a f-point decreases with an increase in the length of radius d [4].

In equation (1) 'r' is the radius. From Fig.1(b) we can conclude that if the value of r is equal to 0 then we can conclude that given f-point is an exact point with validity index 1.

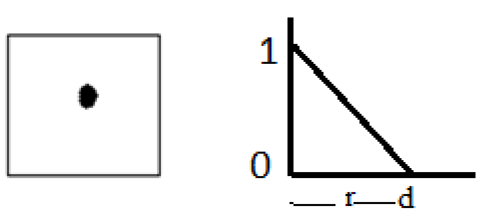


Figure 1: (a) f-point (b) Membership function

$$\mu(\text{f- point}) = \begin{cases} \dfrac{d-r}{d} & \text{if } 0 \le r \le d \\ 0 & \text{if } d \le r \end{cases} \quad (1)$$

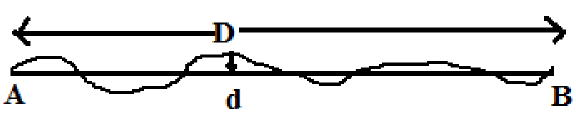


Figure 2: f-line

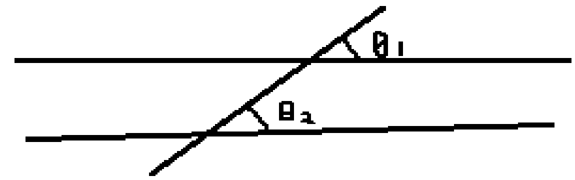


Figure 3: f-parallel

3.2 fuzzy Line

Let us consider an f-line as shown in Fig.2, which is like a curve that passes through a straight line *AB,* such that the distance between any point on the curve and the straight line *AB* is very small or negligible. With reference to Fig. 2, this implies that we are assigning a small value to the distance d, with the understanding that d is inferred as an imprecise value [4].

$$\mu(\text{f- line}) = \begin{cases} \dfrac{c-d}{c-b} & \text{if } b \le d \le c \\ 0 & \text{if } c < d \end{cases} \quad (2)$$

Equation (2) is membership function of f-line , where d << D, and 'd' is the maximum distance between the f-line and straight line and 'D' is the length of the straight line AB. Any line with little increase in difference from the straight line results in a decrease of the membership value.

3.3 fuzzy Parallel

In f-geometry, two lines are said to be f-parallel, if its membership value is closer to the membership value of parallel lines, and the membership function decreases with an increase in the difference of the corresponding angle θ as given by (3).

$$\mu(\text{f-similar angle}) = \begin{cases} \dfrac{c-h}{c-b} & \text{if } b \le h \le c \\ 0 & \text{if } c < h \end{cases} \quad (3)$$

h is the difference between angles θ_1 and θ_2.

Since the membership function decreases with the increasing value of h, we have estimated the membership function for two lines as f-parallel in (3).

3.4 fuzzy Triangle

In f-geometry a shape is said to be fuzzy triangle (*f*-triangle) if its membership value is closer to the membership value of triangle. As shown in Fig.4(b).

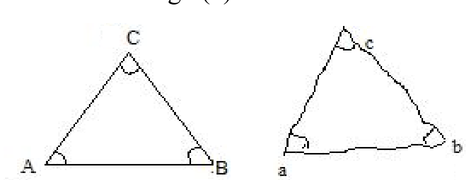


Figure 4: (a) Crisp-triangle (b) f-triangle

Figure 5: Estimating the *f*-interior angle of *f*-trianlge

The membership function of f-triangles is given by using some rules of geometry. Let us prove *f*-triangle by constructing it using (i) three *f*-lines, and (ii) three *f*-angles, as shown in Fig 5. We find the interior *f*-angle of an *f*-triangle by drawing f-altitude opposite to the angle.

Thereafter, we find the length of the base (x) and *f*-altitude (y) as shown in Fig. 5.

$$\theta_1=\tan^{-1}(y/x)$$

We substitute the values of x and y for estimating the *f*-interior angle θ_1 . The same method is followed for finding the other two *f*-interior angles θ_2 and θ_3.

In (4) μ_{SIA} denotes the membership value of sum of internal angles, where θ is given by $\theta=\theta_1+\theta_2+\theta_3$.

$$\mu(\text{f- SIA}) = \begin{cases} \dfrac{\theta-a}{b-a} & \text{if } a\leq\theta\leq b \\ \dfrac{c-\theta}{c-b} & \text{if } a\leq\theta\leq b \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Where θ, a, and b are real numbers in equation 4.

The membership of f-triangle is given by (5).

$$\mu\,(\text{f-triangle})=\mu_{d1}*\mu_{d2}*\mu_{d3}*\mu_{SIA} \quad (5)$$

Here μ_{d1}, μ_{d2}, and μ_{d3} denote the membership values of f-line with distance d_1, d_2, and d_3 from the reference straight line given by (2) respectively. The μ_{SIA} denotes the membership value of sum of internal angles given by equation (4).

4.0 F-Theorem

The f-theorem in f-geometry is f-transform of a theorem in Euclidean geometry. In f-theorem, we are trying to formalize the f-concept in f-geometry, generally in the form of membership functions, e.g. by transforming of some rules of crisp geometry into f-geometry.

4.1 f-similarity and f-validity

In f-geometry any two f-objects are said to be f-similar, if both of them have same shape. Very specifically, by uniform scaling one must be congruent to other. Conversely, f-similar polygons may be of same f-angles and scaling of f-sides may be proportionate. This section illustrates the concept of f-similarity by using well known triangle postulates. There are three well known postulates. All the interior angles are same (AAA), all the sides are same in proportion (SSS), and two sides are same in proportion with a same angle (SAS). These foresaid postulates prove triangles to be similar. Assuming a formal illustration of the concept of the f-theorem, let us consider that ABC constitutes a triangle with three straight lines and three interior angles, as shown in Fig. 4(b).

4.1.1 Angle Angle Angle (AAA): In f-geometry, two triangles are said to be f-similar if their membership function has high validity index to the property of similar triangles (AAA) and the membership values decrease with the increase of difference of the corresponding angles. The high validity index in this context refers to the degree of closeness. It is mathematically represented as

$$\mu(\text{f-similar})=\mu_{A1}*\mu_{A2}*\mu_{A3} \quad (6)$$

Where μ_{A1}, $\mu_{A2,}$andμ_{A3} are membership of angle1, angle2, angle3 are based on the difference in the corresponding angles.

4.1.2 Side Side Side (SSS): In f-geometry, two triangles are said to be f-similar if their membership function has high validity index to the property of similar triangles (SSS), with all the three corresponding sides are equal in f-proportion and the membership values decrease even in slight difference in proportion of the sides. Mathematically represented as

$$\mu(\text{f-similar})=\mu_{S1}*\mu_{S2}*\mu_{S3} \quad (7)$$

Where μ_{S1} , μ_{S2} , and μ_{S3} are memberships of f-proportions of corresponding side1, side2, side3 respectively.

4.1.3 Side Angle Side (SAS): In f-geometry, two triangles are said to be f-similar if their membership function has high validity index to the property of similar triangles (SAS) and the membership values decreases in difference in corresponding angle and difference in proportion of two corresponding sides. Mathematically represented as

$$\mu(\text{f-similar})=\mu_{S1}*\mu_{A2}*\mu_{S3} \quad (8)$$

Where $\mu_{S1,}$ μ_{A2}, and μ_{S3} membership functions of side1, angle2, side3 respectively. In case of SAS, we assume that AB/DC' *=BC/AD' *= k (A constant) i.e. corresponding sides of the two triangles are in the same ratio as in geometry. Here AB/DC' and BC/AD' takes the fuzzy proportion values $k_1$and k_2 respectively. Point to be noted here is AB/DC' *= BC/AD means AB/DC' is approximately equals to BC/AD [4-6]. The membership function of the f-similar side and f-similar angle are computed by (9) and (10) respectively.

$$\mu(\text{f-similar side})= \begin{cases} \dfrac{c-j}{c-b} & \text{if } b\leq j\leq c \\ 0 & \text{if } c\leq j \end{cases} \quad (9)$$

Where j is given by k-k1 and k-k2 .

$$\mu(\text{f-similar angle})= \begin{cases} \dfrac{c-l}{c-b} & \text{if } b\leq l\leq c \\ 0 & \text{if } c\leq l \end{cases} \quad (10)$$

Here $l=\theta_1-\theta_2$ is the difference between angles.

The f-similarity of triangles is given by

$$\mu_{SAS}\text{(f- similarity)}=\text{f-val1}*\text{f-val2}*\mu_{diff1}*\mu_{diff2}*\mu_{diff3} \quad (11)$$

Where f-val1 and f-val2 are the f-validities of corresponding triangles and given by (5). The μ_{diff1} and μ_{diff2} are membership of the difference between corresponding f-interior angles is calculated by (10).

5.0 PROPOSED F-Theorem

In this section we have introduced the definition of fuzzy parallelogram (f-parallelogram) and fuzzy rhombus (f-rhombus). Further we have discussed f-theorems and their f-proof for f-parallelogram and f-rhombus by using validation principle. By using the proposed methodology we can introduce more complex f-objects. These f-objects may play vital role for identifying clues in computational forensic**.** This section further comprises of some proposed f-theorems and their f-proofs. In f-geometry, f-proof may be either empirical or logical. The empirical f-proof involves experiments while the logical f-proof is the f-transform of their counterpart of Euclidean geometry. An important principle in f-geometry is validation principle, "Let *p* be a *p*-valid conclusion drawn from a chain of premises $p^1,...,p^n$. Then, using the star notation,$*p$ is an *f*-valid conclusion drawn from$*p^1,...,*p^n$ and $*p$ has a high validity index. It is this principle that is employed to derive *f*-valid conclusions from a collection of *f*-premises"[1].The validation principle leads to the following assertion in *f*-geometry. If ABC and *ACD* are *f*-similar *f*-triangles then ABCD be f-parallelogram. In f-theorem 1 to 4 Fig. 6 is taken as an example. In theorem 5 Fig. 7 is taken as an example.

5.1 f-parallelogram

A shape ABCD as shown in Fig.6, is called f-parallelogram, if both pairs of opposite f-sides are f- parallelogram and represented in terms of membership function as [7][8]

$$\mu(\text{f-p})=\mu_D*\mu_{diff} \quad (12)$$

$$\mu(\text{f-p})=\mu_{d1}*\mu_{d2}*\mu_{d3}*\mu_{d4}*\mu_{diff1}*\mu_{diff2} \quad (13)$$

Where μ_{d1},μ_{d2},μ_{d3}, and μ_{d4} are the individual membership values of f-Sides AB, BC, CD, and AD respectively as given by (2). The μ_{diff1} and μ_{diff2} are the membership values of difference of the corresponding angles given by (10) .

5.2 f-rhombus

In a shape (as shown in Fig. 7) if all of its f-sides are f-similar such that the difference between all four sides are very small or negligible, then it is called f-rhombus [7][8]. The f-rhombus is represented by (14) in terms of membership function.

$$\mu(\text{f-rhombus })=\mu_{d1}*\mu_{d2}*\mu_{d3}*\mu_{d4} \quad (14)$$

Where μ_{d1},μ_{d2},μ_{d3}, and μ_{d4} are the individual membership values given by (9).

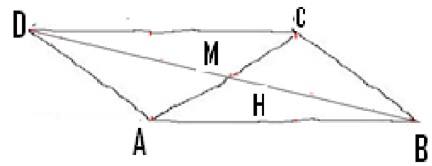


Figure 6: f-parallelogram

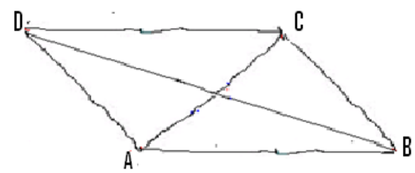


Figure 7: f-rhombus

5.3 f-theorems

5.3.1 f-theorem1: If ABCD be a f-parallelogram with a diagonal f-line AC. Then f-line AC divides f-parallelogram into two f-similar triangles ABC and ACD.

Above f-theorem has been derived by using validation principle. Let ABCD be the f-quadrilateral and consist of two f-triangles ABC and ACD. If f-triangles ABC and ACD are f-similar then ABCD will be the f-parallelogram. Higher degree of similarity of f-triangles ABC and ACD leads to f-valid conclusion, "The quadrilateral ABCD has higher degree of validity index of f-parallelogram".

5.3.1.1 f-proof : The f-similarity of f- triangles ABC and ACD is calculated by (11).

$$\mu_{SAS}\text{(f-similarity)} =\text{f-val}_{ABC} * \text{f-val}_{ACD}*\mu_{diff1}*\mu_{diff2}*\mu_{diff3} \quad (15)$$

Where f-val_{ABC} and f-val_{ACD} are the f-validities of triangle ABC and ACD respectively. The μ_{diff1} and μ_{diff2} are the differences of the corresponding sides given by (9). The μ_{diff3} is difference of the corresponding f-interior angle which is calculated by (10). The f-val1 and f-val2 are evaluated by (5) as follows.

$$\mu(\text{f-validity1})=\mu_{d1}*\mu_{d2}*\mu_{d3}*\mu_{SIA}$$

Where μ_{d1},μ_{d2}, and μ_{d3} denote the membership values of f-sides AB, BC, and CD with distance d_1, d_2, and $d_3$from the reference straight line respectively. The μ_{SIA} denotes the membership value of sum of internal angles.

$$\Sigma\theta= \angle CAB+ \angle ABC+\angle ACB,$$

Where $\angle$ CAB, $\angle$ ABC, and $\angle$ ACB are the internal angles of triangle ABC.

$$\mu(\text{f-validity2})=\mu_{d1}*\mu_{d2}*\mu_{d3}*\mu_{SIA}$$

Where μ_{d1},μ_{d2}, and μ_{d3} denote the membership values of f-line AC, AD, and DC with distance d_1, d_2, and $d_3$from the reference straight line respectively. The μ_{SIA} denotes the membership value of sum of internal angles

$$\Sigma\theta = \angle CAD+\angle ADC+\angle ACD$$

Where $\angle$ CAD, $\angle$ ADC, and $\angle$ ACD are the internal angles of triangle ADC.

The membership function of each f-line and sum of interior angles are calculated by using (4) and (5) respectively.

5.3.2 f-theorem2: In f-quadrilateral, if difference of each pair of opposite f-angle is either very small or negligible then ABCD is f-parallelogram..

5.3.2.1 f-proof : The quadrilateral ABCD has higher degree of validity index of f-parallelogram if the value of $\delta\theta_1$ and $\delta\theta_2$ are small or negligible .

Where $\delta\theta_1$ and $\delta\theta_2$ are given by $\delta\theta_1 = \angle DCB - \angle DAB$

$$\delta\theta_2 = \angle ADC - \angle ABC$$

The validity index is given by (16)

$$\mu(\text{f-validity}) = \mu_{diff1} * \mu_{diff2} * \mu_{d1} * \mu_{d2} * \mu_{d3} * \mu_{d4} \quad (16)$$

The μ_{diff1} and μ_{diff2} denote the membership values of difference of opposite angles $\delta\theta_1$ and $\delta\theta_2$ respectively. The value of μ_{diff1} and μ_{diff2} is evaluated by (3). Where $\mu_{d1}, \mu_{d2}, \mu_{d3}$, and μ_{d4} are memberships of f- lines AB, BC, DC, and AD with distance d_1, d_2, d_3, and d_4 from the reference straight line respectively.

5.3.3 f-theorem3: If the diagonals (f-line) of f-quadrilateral bisect each other then it will be a f-parallelogram.

5.3.3.1 f-proof: The quadrilateral ABCD has higher degree of validity index of f-parallelogram if the value of distance *diff* between intersection point H of diagonals AC and BD from midpoint of any diagonal is either small or negligible .

$$diff = H - M$$

Where validity index given by

$$\mu(\text{f-validity}) = \mu_{diff1} * \mu_{d1} * \mu_{d2} * \mu_{d3} * \mu_{d4} * \mu_{d5} * \mu_{d6} \quad (17)$$

The value of μ_{diff} denotes the membership value of the distance given by (2). The μ_{d1}, μ_{d2}, μ_{d3}, and μ_{d4} are membership values of f-lines AB, BC, CD, and AD respectively. The μ_{d5} and μ_{d6} denotes the membership values of f-lines AC and DB with distance d_5 and d_6 from the reference straight line respectively.

5.3.4 f-theorem4: A f-quadrilateral is f-parallelogram if a pair of opposite sides is f-equal and f-parallel.

5.3.4.1 f-proof : The quadrilateral ABCD has higher degree of validity index of f-parallelogram if the difference between opposite f-sides AB/CD and BC/AD and difference between opposite angle are either small or negligible. The validity index is given by (18).

$$\mu(\text{f-validity}) = \mu_{diff1} * \mu_{diff2} * \mu_{diff3} * \mu_{diff4} * \mu_{d1} * \mu_{d2} * \mu_{d3} * \mu_{d4} \quad (18)$$

Where μ_{diff1} and μ_{diff2} are denoting the membership value of f-similar side by (9) . μ_{diff3} and μ_{diff4} denoting the membership value of f-parallel line by (3). $\mu_{d1}, \mu_{d2}, \mu_{d3}$, and μ_{d4} are memberships of f-line AB, BC, DC and AD with distance d_1, d_2, d_3, and d_4 from the reference straight line respectively.

5.3.5 f-theorem5*:* If diagonals (f-lines) of f-quadrilateral are perpendicular to each other then f-quadrilateral will be a f-rhombus.

5.3.5.1 f-proof*:* The f-quadrilateral ABCD has higher degree of validity index of f-rhombus, if the difference between interior angle made by diagonals from right angle is either small or negligible .

The validity index is given as

$$\mu(\text{f-validity}) = \mu_{IA1} * \mu_{IA2} * \mu_{d1} * \mu_{d2} \quad (19)$$

Here μ_{IA1} and μ_{IA2} are the individual membership values of interior angle1 and interior angle2 from the right angle respectively. μ_{d1} and μ_{d2} denote the membership values of diagonal f-lines AC and DB with distances d_1 and d_2 respectively.

The f-similarity and f-validity of f-objects is estimated by using multiplication and OWA methods. It is shown in results OWA method has significant improvement over multiplication method.

5.4 Simple Multiplication Method

This method is multiplication of all the membership values.

5.5 Ordered Weighted Averaging Method

Ordered Weighted Averaging (OWA) is the central concept of information aggregation, originally introduced by Yager[3]. OWA facilitates the means of aggregation in solving of problems that arises in multi criteria decision making. Furthermore, OWA operator provides a parameterized family of aggregation operators, including well-known operators such as maximum, minimum, arithmetic mean, k-order statistics, and median. Sometimes, exact **AND**-ness is required for multi-criteria decision making, which offers minimum value and sometimes exact **OR**-ness which offers maximum value. The OWA aggregation operator lies somewhere in between the two extremes of **AND**-ness and **OR**-ness. Two extremes are restricted to mutually exclusive probabilities for multiplication (like **AND**-gate) and summation (like **OR**- gate). Subsequent part discloses a brief account of OWA operators, detail discussion about the behavior of operators is in [3]. The OWA operation involves three following steps.

1) Reordering of inputs, 2) Weight determination related with OWA operators, and 3) Aggregation process.

Definition: "Mapping the OWA operator R from $R^m \rightarrow R$, (where R = [0, 1]), with dimension m, has weighting vector $w = (w_1, w_2, w_3, \ldots w_m)^T$, where $w_j \in [0, 1]$ and $\Sigma w_j = 1$, the summation of individual weights will always found to be one"[3]. Thus, for the input parameter $(x_1, x_2, x_3 \ldots\ldots x_m)$, the size of multicriteria will be m,

In vector $(y_1, y_2, y_3 \ldots . y_m)$ the y_j is the j^{th} largest number in the vector $(x_1, x_2, x_3, \ldots x_n)$, and $y_1 \geq y_2 \geq y_3 \geq \ldots \geq y_m$. However, the weights wj of the operator R are not related with any exact value of xj, instead they are related with the ordinal position of y_j [3]. The minimum and maximum range of values can be decided based upon the concept of **OR**-ness (β), as defined by (21) [3].

$$OWA(x_1, x_2, x_3 \ldots\ldots x_m) = \sum^{m} w_j y_j \qquad (20)$$

$$\beta = \frac{1}{m-1} \sum_{j=1}^{m} w_j (m-1) \qquad (21)$$

Here, β (***OR**-ness)* ranges between [0, 1]. On every occasion the value of β = 1, generates the weight vector as (1, 0, 0,...0). Thus, the maximum value of x_j acquires the entire weight, resulting the OWA operator as *maximum* operator. On the other hand, if β = 0, generates the weight vector as (0,0,0,.., 1). Thus, the minimum value of x_j will acquires the entire weight, resulting the OWA operator as *minimum* operator. When β = 0.5, generates the weight vector as (1/n, 1/n, 1/n,…, 1/n), means that arithmetic mean of weights are evenly distributed among the inputs. The membership function of a relative quantifier can be represented as

Here, β (***OR**-ness)* ranges between [0, 1]. On every occasion the value of β = 1, generates the weight vector as (1, 0, 0,...0). Thus, the maximum value of x_j acquires the entire weight, resulting the OWA operator as *maximum* operator. On the other hand, if β = 0, generates the weight vector as (0,0,0,.., 1). Thus, the minimum value of x_j will acquires the entire weight, resulting the OWA operator as *minimum* operator. When β = 0.5, generates the weight vector as (1/n, 1/n, 1/n,…, 1/n), means that arithmetic mean of weights are evenly distributed among the inputs. The membership function of a relative quantifier can be represented as

$$Q(r) = \begin{cases} 0 & \text{if } r < a \\ \dfrac{r-a}{b-a} & \text{if } b \leq r \leq a \\ 1 & \text{if } r > b \end{cases} \qquad (22)$$

Where a, b, and r∈[0,1].

In [3], Yager calculates the weights w_j of the OWA aggregation from the function *Q* describing the quantifier, with m number of criteria.

$$w_j = Q(j/m) - Q((j-1)/m) \qquad (23)$$

6. EXPERIMENTAL WORK AND RESULTS

In this section the f-validity of f-parallelogram and f-rhombus is computed. The f-theorem1, f-theorem2, f-theorem3, f-theorem4, and f-theorem5 are illustrated in Example1, Example2, Exmaple3, Exmaple4, and Exmaple5 respectively. In Fig. 9 to 13 results of practical work is shown. The sample images which are taking inputs are shown in Fig.8.

Examples

Example1: The f-parallelogram shown in Fig.6 is constituted by two triangles ABC and ACD. The ABC has f-transformation distances for three f-lines AB, BC, and AC are 5, 49, and 9. This results in μ_d as {0.97, 0.74, 0.95}. The sum of interior angles is 177.16^o which in turn results in μ_{SIA} as {0.716}.The *f*-validity is calculated by taking the product of the above membership values {0.97, 0.74, 0.95, 0.716}. The result comes to be 0.4882. For triangle ACD the f-transformation distances for three sides (f-lines) AC, DC, and AD is 9, 6, and 13. This results in μ_d as {0.95, 0.96, 0.93}. The sum of interior angles is 176^o which in turn results in μ_{SIA} as {0.6}.The *f*-validity is calculated by taking the product of the above membership values {0.97, 0.74, 0.95, 0.6}. The result comes to be 0.513. The difference among the corresponding f-interior angles is found as {7} generates the membership values {0.3}.

The differences in proportion of corresponding f-sides are 0.02 and 0.03 generate membership values 0.98 and 0.97. To compute the f-similarity, we go for implementing the above data set by using (15). Then, the f-similarity is computed with the following values as:

$$\mu_{SAS}(\text{ f-similarity}) = \text{f-val}_{ABC} * \text{ f-val}_{ACD} * \mu_{diff1} * \mu_{diff2} * \mu_{diff3}$$

$$\mu_{SAS}(\text{ f-similarity}) = \{\mu_{AB} * \mu_{BC} * \mu_{AC} * \mu_{SIA1} * \mu_{AC} * \mu_{DC} * \mu_{AD} * \mu_{SIA2} * \mu_{diff1} * \mu_{diff2} * \mu_{diff3}\}$$

$$\mu_{SAS}(\text{ f-similarity}) = 0.0714$$

The quadrilateral has 0.0714 validity index of f-parallelogram. In OWA method, we have considered all 11 important parameters by using (15) as inputs, i.e. the size of input vector m = 11.The fuzziness in f-parallelogram is computed using the OWA operator R for the linguistic quantifier "most" i.e. a=0.3 and b= 0.8 by (22). The weight vector generated by (23) is (0,0,0,0.1273, 0.181818, 0.181818, 0.181818, 0.181818, 0.145455) used in (20) with membership values 0.3,0.6, 0.716, 0.74, 0.74, 0.93, 0.95,0.95,0.9,0.97 and 0.98 produces f-similarity of 0.8 which is a significant improvement over 0.07 produced by multiplication method. The corresponding results are shown in Fig. 9.

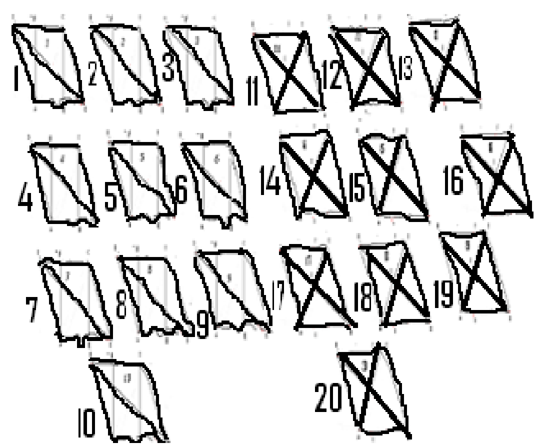


Figure 8: Sample Images

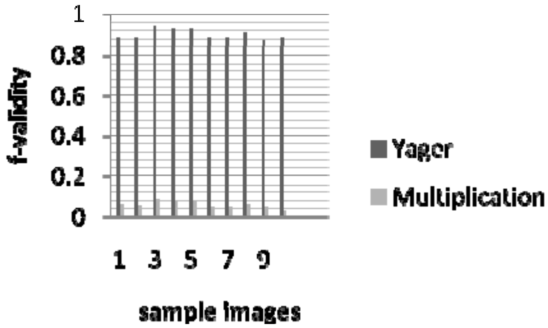


Figure 9: Comparision of f-validity generated by Yager and simple multiplication method by theorem 1

Example 2: In f-parallelogram shown in Fig.6, f-transformation distance for the f-lines AB, BC, CD and AD are 5, 49, 6, and 13. This results in μ_d as {0.97, 0.74, 0.96, 0.93}. The differences of the opposite interior angles $\angle DCB$- $\angle DAB$ and $\angle ADC$- $\angle ABC$ are 0.2 and 0.6. Which in turn results μ_{diff} as {0.98, 0.94}.The values of angle $\angle DCB$, $\angle DAB$, $\angle ADC$, and $\angle ABC$ are 56.4^0, 56.8^0, 123.6^0 and 123^0 respectively.

The f-validity is given by (16).

μ(f-validity)=μ_{diff1}* μ_{diff1}*μ_{d1}*μ_{d2}*μ_{d3}*μ_{d4}

μ(f-validity)=0.98∗0.94∗0.97∗0.74∗0.9681∗0.9309

μ(f-validity)=0.595

In OWA method, we consider 6 important parameters by using (16) as inputs, i.e. the input value m = 6.The fuzziness in f-parallelogram will be computed using the OWA operator R for the linguistic quantifier "most" i.e. a=0.3 and b= 0.8 by (22). The weight vector generated by (23) is (0, 0, 0.333, 0.166, 0.166, 0.333, 0).The ordered membership values are (0.98, 0.97, 0.9681, 0.94, 0.9309,0.74).Produces the f-validity 0.953414 by (20).The significant improvement over multiplication method in the result can be seen in Fig. 10.

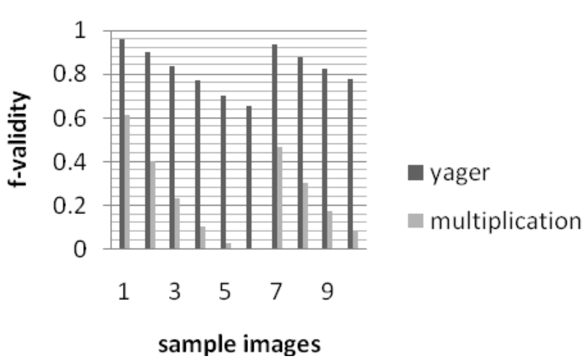


Figure 10: Comparision of f-validity generated by Yager and simple multiplication method by theorem 2

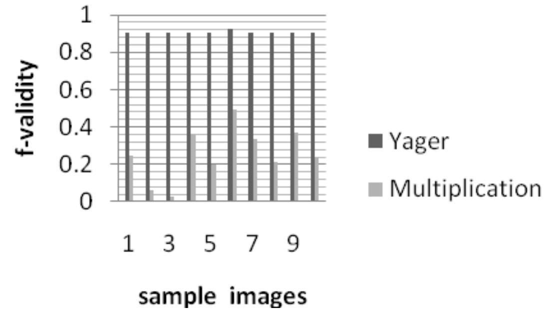


Figure 11: Comparision of f-validity generated by Yager and simple multiplication method by theorem 3

Example 3: In f-parallelogram shown in Fig.6, f-transformation distance for the f-lines AB, BC, CD, and AD are 5, 49, 6, and 13. This results in μ_d as {0.97, 0.74, 0.9681, 0.9309 }. The f-transformation distances of intersection points of diagonal AC and BD are 5 and 49.This results in μ_d as {0.97, 0.74}.

μ(f-validity)=μ_{diff1}*μ_{d1}*μ_{d2}*μ_{d3}*μ_{d4}*μ_{d5}*μ_{d6}

The value of difference is 11.5 which generate μ_{diff} 0.42. Then, the f-validity is computed by (17).

μ(f-validity)=0.42* 0.97* 0.74*0.9681* 0.9309*0.97*0.74

μ(f-validity)=0.247454

To compute the level of f-validity by OWA in f-parallelogram, here we consider 7 important parameters by (19) as inputs, i.e., with *m* = 7, the fuzziness in an f-parallelogram will be computed using the OWA operator R for linguistic quantifier "most" a=0.3 and b= 0.8.The weight vector is (0,0,0.2571,0.2857,0.2857,0.1714,0) with membership values (1,0.95,0.93,0.640) produces f-validity 0.9072 over 0.247454. Please refer to Fig.11 which is showing the results after applying theorem 3 on sample images of Fig.8.

Example4: In f-parallelogram shown in Fig.6, f-transformation distance for the f-lines AB, BC, CD and AD are 5, 49, 6, and 13. This results in μ_d as {0.97, 0.74, 0.9681, 0.9309}. The product of membership values of μ_d is 0.646259.The proportion of opposite f-lines CD and AB is 1.05757566.The proportion of f-lines BC and AD is 0.849276. Then, the f-validity is computed by (18)

μ(f-validity)=μ_{diff1}* μ_{diff2}*μ_{diff3}*μ_{diff4} *μ_{d1}*μ_{d22}*μ_{d3}*μ_{d4}

μ(f-validity)=0.367

To compute the level of f-validity by OWA in f-parallelogram, 8 important parameters has been considered as inputs, i.e., with *m* = 8. The fuzziness in an f-parallelogram will be computed by using (20). The OWA operator R for linguistic quantifier "most" a=0.3 and b= 0.8. The weight vector is (0,0, 0..2571,0.2857,0.2857,0.1667 ,0.1667, 0.3333,0) with the membership values (0.97, 0.97,0.74, 0.9681, 0.93, 0.894,0.915)

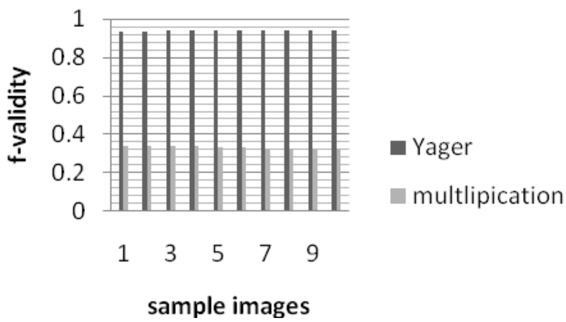


Figure 12: Comparision of f-validity generated by Yager and simple mutiplication method by theorem 4

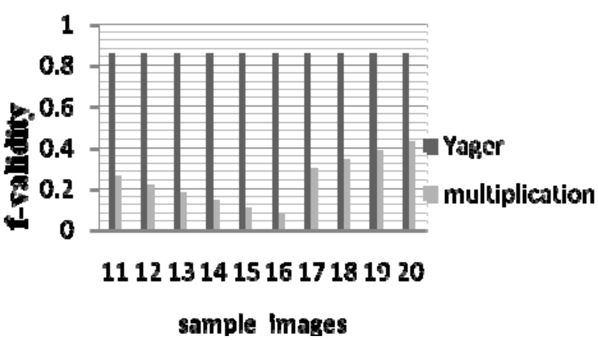


Figure 13: Comparision of f-validity generated by Yager and simple multiplication method by theorem 5

produces f-validity 0.9026.The experimental results are shown in Fig.12.

Example 5: In f-rhombus shown in Fig. 7, f-transformation distance of f-lines AC and DB are 9 and 13 respectively. The membership values of f-lines AC and DB are 0.95 and 0.93. The values of internal angles DEF and CEF are 102.254, 94.90173 respectively. Which in turn results μ as {0.38, 0.75}.Then, the f-validity is computed by (19)

μ (f-validity)=μ_{IA1}*μ_{IA2}*μ_{d1}*μ_{d2}= 0.258.

To compute the level of f-validity by OWA in f-parallelogram, here we consider 4 important parameters by (19) as inputs, i.e., with m =4, the fuzziness in an f-parallelogram will be computed using the OWA operator R for linguistic quantifier "most" a=0.3 and b= 0.8.The f-validity is 0.86.The weight vector generated by (23) is (0.0,0.4,0.5,0.1).The ordered membership values (0.38,0.75,0.93,0.95) produces f-validity 0.862.Comparison of results of Yager's and simple multiplication method is shown in Fig. 13.

7. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have given the f-definitions and f-theorems of f-geometry. Afterwards, we have estimated the f-similarity of f-geometric objects by using membership values, such as f-parallelogram and f-rhombus by using f-theorems. By using the proposed methodology we can introduce more f-objects. These f-objects may play vital role for identifying clues in computational forensic. The OWA operators are employed to aggregate the membership values of individual feature and produced more improved result. In future, the investigation will be on for new and more effective multi criteria decision making methods.

REFERENCES

[1]. L.A. Zadeh, "From Fuzzy Logic to Extended Fuzzy Logic—The Concept of f-Validity and the Impossibility Principle", Plenary session, FUZZ-IEEE 2007, July 24 , Imperial College, London, UK,(2007)

[2]. L.A.Zadeh, "Towards Extended Fuzzy Logic – A First Step", Fuzzy Sets and Systems, *p.p.*3175-3181 Information Sciences (2009) .

[3]. R. R. Yager, "On OWA aggregation operators in multi-criteria decision making", IEEE Transactions on Systems, Man, and Cybernetics, 18:183–190, (1988).

[4]. B.M.Imran, M.M.S.Beg, "Elements of Sketching with Words", in Proc of IEEE International Conference on Granular Computing, (GrC2010), San Jose, California, USA, August 14-16, p.p.: 241-246, (2010)

[5]. B.M.Imran, M.M.S.Beg, "Towards Computational forensics with f-geometry", World Conference on Soft Computing' 2011, San Francisco State University, California, USA (2011)

[6]. Abdul Rahman, M. M. S. Beg, "Face Sketch Recognition Using Sketching With Words", In Springer 's International Journal of Machine Learning and Cybernetics ISSN 1868-8071 (2014)

[7]. Abdul Rahman, M.M.S. Beg. "Estimation of f-validity of Geometrical Objects With OWA Operator Weights", Fuzz-2013 IEEE International Conference on Fuzzy Systems on July 7-10 , Hyderabad India, (2013)

[8]. Abdul Rahman, M.M.S. Beg, "Investigation of OWA Operator Weights for Estimating f-validity of Geometrical Objects", in Proceeding of Third Annual World Conference on Soft Computing, (WCSC-2013) San Antonio, Texas, USA, December 16-18 p.p. 15-24 (2013)

[9]. H.R. Tizhoosh, "Fuzzy Image Processing: Potentials and State of the Art", IZUKA'98, 5th International Conference on Soft Computing, Iizuka, Japan, October 16-20, 1998, vol. 1, pp. 321-324 (1998).

[10]. T. Poston, Fuzzy geometry, Ph.D. Thesis, University of Warwick, (1971).

[11]. J. Roe, "Index theory, coarse geometry, and topology of manifolds", in: CBMS: Regional Conf. Ser. in Mathematics, The American Mathematical Society, Rhode Island,(1996)

[12]. A. Rosenfeld, "Fuzzy geometry: an updated overview," Information Science 110 (3–4) pp. 127–133. (1998)

[13]. J. Buckley, "E. Eslami, Fuzzy plane geometry I: points and lines", Fuzzy Sets and Systems 86 pp.79–187. (1997)

[14]. S. Mayburov, “Fuzzy geometry of phase space and quantization of massive fields”, Journal of Physics A: Mathematical and Theoretical 41 p.p. 1–10 (2008)

[15]. S.G. Tzafestas, C.S. Chen, T. Fokuda, F. Harashima, G. Schmidt, N.K. Sinha, D. Tabak, K. Valavanis (Eds.), “Fuzzy logic applications in engineering science”, in: Microprocessor-based and Intelligent Systems Engineering, Vol. 29, Springer, Netherlands, , pp. 11–30 (2006)

[16]. M. M. S. Beg, “User Feedback Based Enhancement in Web Search Quality,” International Journal of Information Sciences, Elsevier, vol. 170, no. 2-4, pp. 153-172, (2005).

[17]. G. Beliakov, A. Pradera, T. Calvo. “Aggregation Functions: A Guide for Practitioners,” Springer, (2007)

[18]. V. Torra, Y. Narukawa, “Modeling Decisions: Aggregation Operators and Information Fusion”, Springer, (2007).

[19]. R. R. Yager, “Connectives and quantifiers in fuzzy sets”, Fuzzy Sets and Systems, 40:39–75, (1991).

[20]. R. R. Yager, L. S. Goldstein, and E. Mendels, “Fuzmar: an approach to aggregating market research data based on fuzzy reasoning,”, Fuzzy Sets and Systems, 68(1), Pg: 1–11, (1994).

[21]. H. B. Mitchell and D. D. Estrakh, “A modified OWA operator and its use in lossless DPCM image compression”, International Journal of Uncertain Fuzziness, Knowledge Based Systems, 5, pp 429–436,(1997).

[22]. G. Bordogna, M. Fedrizzi, G. Pasi, “A linguistic modeling of consensus in group decision making based on OWA operators”, IEEE Transactions on Systems, Man and Cybernetics - Part A, 27(1):126– 133, (1997).

[23]. R. Sadiq, S.Tesfamariam, “ Developing environmental indices using fuzzy numbers ordered weighted averaging (FN-OWA) operators, Stochastic Environmental Research and Risk Assessment”, Springer 22(5):p.p. 495-505 (2008).

[24]. Y.S.Choi, D.Kim, R.Krishnapuram, “Relevance feedback for content-based Image retrieval using the choquet integral”, in Proc. of IEEE International Conference on Multimedia and Expo, p.p.: 1207 – 1210, USA, (2000)

Special Issue of BIJIT on Fuzzy Logic
BIJIT - BVICAM's International Journal of Information Technology
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Optimal Enactment of a Stand-alone Hybrid Wind-Fuel Cell Based Distributed Generation System Through Fuzzy Logic Control

Mohammad Saad Alam

Submitted in August 2013; Accepted in February, 2014

Abstract - In this work, a hybrid distributed power generation (DG) system composed of two renewable energy sources, viz. a wind turbine and a fuel cell is proposed. A fuzzy logic controller has been introduced for optimal power management to provide electric supply to a residential load on a continuous basis based on the feasibility of economic power generation. This controller directs power to a fixed voltage bus in the power conditioning unit (PCU). The fixed voltage bus supplies the load, while the excess power is directed to the energy storage bank first and then to an electrolyzer, which is used to generate hydrogen for the fuel cell. Complete system modeling and simulation has been carried out through HOMER software, and hybrid controller has been simulated in Simulink/MATLAB environment. The simulation results proved the effectiveness of the hybrid fuzzy logic controller for real-time applications of intelligent methods in sustainable power and energy systems.

Index Terms – Distributed Generation, Fuel cell, Fuzzy logic controller, Wind energy

NOMENCLATURE

P_m : Mechanical output power of the turbine
Cp: Performance coefficient of the wind turbine
λ: Tip speed ratio of the rotor blade to wind speed
β: Blade pitch angle (^{o})
ρ: Air density (kg/m ^{3})
A: Turbine swept area (m ^{2})
V_{wind} : Wind speed (m/sec)
ηt, ηe, η r : Thermal, Electric and Reaction efficiency
n_{H2}: hydrogen produced (moles/sec)
,η_F : Faraday efficiency = $96.5e^{(0.009/i_e - 75.5/i_e^2)}$,
n_c : Number of electrolyzer cells in series,
F: Faraday constant (Ck/mol)
i_e : Electrolyzer current
u: Controller output; K_P: Proportionality constant;
K_I: Integral constant; k : is the k^{th} sampling time;
e(k): Error in the controller input i.e. $e(k)= V_{ref} - V(k)$;
$\Delta e(k)$:Change of error in input signal i.e.$\Delta e(k)= e(k)-e(k-1)$;
$\Delta u(k)$:Change of control output $u(k)$ i.e. $\Delta u(k)= u(k)-u(k-1)$.

Lead R&D Systems Engineer at Magneti Marelli SpA (Fiat Chrylser Automobiles) Auburn Hills, MI, USA.
Email: hybridvehicle@gmail.com

1.0 INTRODUCTION

Energies like fossil fuel and electricity are the essence of the world. Nations have always tried to find ways to make them more efficient. The volume of fossil fuel available is limited; hence experts are trying to use different energy resources. The vision of reducing the world's reliance on fossil fuels is challenging [1]. Alternative energy industries, such as nuclear energy, hydroelectric energy, solar energy, wind energy and geothermal energy exist, but these energy sources currently only account for a combined 14 percent of energy consumed worldwide [2].

In the present day world, there are ample renewable energy resources, and a tremendous potential to boost the ability to handle the further expansion of the technology in the power generation, transmission, storage and distribution of energy and to identify large-scale application of these systems [3-6]. Average annual growth rate from 2003-2012 of different power sources is plotted in figure 1.

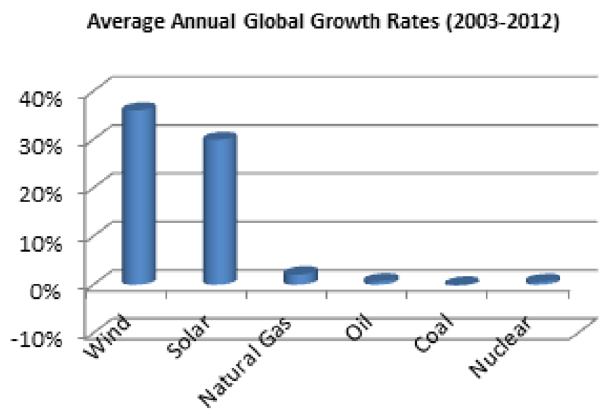


Figure 1: Average Annual Growth Rate [2].

As evident from Figure 1, there is a progressive growth in the deployment of renewable energy power generating systems like solar and wind energy systems. However, these DG systems need to address the applied aspects of what seems feasible from a commercial and cost-effective perspective [7]. The incompetence to pledge reliable, uninterrupted output at a cost that can be analogous to conventional nuclear, steam and coal based electric power generation has been the drawback of renewable based distributed generation systems [3-7].

To mitigate the availability of required wind speed or solar radiations, a number of off-grid hybrid systems, partaking attention from the worldwide energy community. Numerous proto type hybrid DG systems were installed and tested in the past decades [4-6]. Many literature references have discussed how to determine the optimum combination of a hybrid energy system from the meteorological data for small loads (range from 10 W to 1kW) in a given location [5-6, 8-9]. The results clearly indicate that off-grid hybrid DG systems can compete with the grid power, in isolated locations where the grid is either not viable or nonexistent. Hybrid DG systems, are an effective option to solve the problem of power-supply for remote and isolated areas when compared to grids. [10-16]

The key challenge in operating the hybrid DG system is the optimum power management through the available renewable energy [6-13]. The main objective of this proposed work is to design an optimally controlled hybrid wind fuel cell power system. Two renewable energy based power generating sources, a wind turbine and a fuel cell are chosen to provide power to a typical house load. An energy storage unit is also implied for storing the excess energy and to supply the power during the peak load demand periods. Figure 2 shows the power flow diagram of the proposed hybrid wind -fuel cell DG.

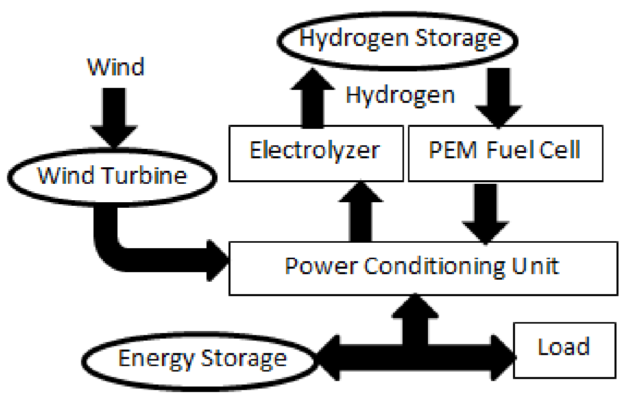


Figure 2: Power Flow Diagram of the Proposed Hybrid DG System.

As a first step towards the realization of this methodology, the proposed DG system is simulated in Hybrid Optimization Model for Electric Renewables (HOMER) environment [17]. Due to constraint limitations, reformer was not included in the preliminary simulation; electrolyzer is considered as the only source for the production of hydrogen. Simulation results are reported in this paper.

Based on a typical house load profile and the wind availability at various locations, it has been estimated that fuel cell in conjunction with energy storage unit, can provide supplementary power to meet the load demand in the absence of the required wind speed. Due to ambiguous and imprecise nature of the wind, house load demand, state of the charge of the energy storage unit, ambient temperature to operate fuel cell etc., it is impossible to provide a set of realistic control data to a conventional controller for optimal power management [18]. A conceivable option is to define the range of the control data and to set up the logic rules for the system and formulate a fuzzy logic based control. The proposed control system will be implied to handle the system level power managements under different load conditions.

Preliminary simulation results in Simulink/ MATLAB® environment show the aptitude of the approach. It is implied that this research will lead to enhancement in the operation of hybrid DG systems considering real-time weather conditions and intelligent control for robust load matching.

This manuscript is divided into five sections. System Architecture is defined in Section 2. Integration of the DG system and simulation in HOMER® is outlined in Section 3. In Section 4, proposed control strategy is developed, simulated and the obtained results are discussed. In section 5, conclusion is drawn based on the results of the proposed work.

2.0 SYSTEM ARCHITECTURE

The key components of proposed hybrid wind- fuel cell DG system are a 20KW wind turbine, 5KW PEM fuel cell, 20KW battery bank, a power conditioning unit (PCU), a 30KW electrolyzer and a hydrogen tank of 0.50 Kg capacity. The ratings of these equipment are considered optimistically in the initial stage. The list of equipment is tabulated in Table 1

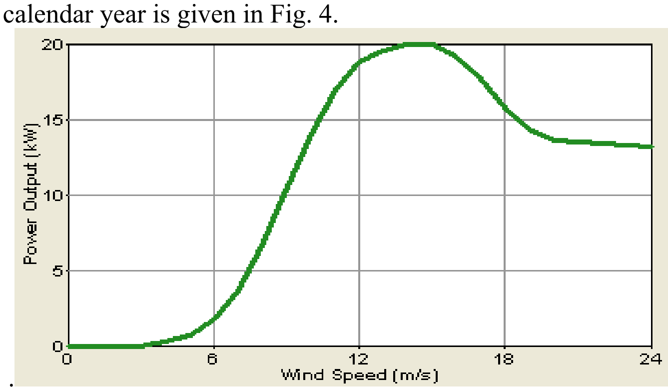
.

Equipment	Rating
Wind Turbine	20 KW
Fuel Cell	5 KW
Battery	20 KW
PCU (Converters)	20 KW
Electrolyzer	30 KW
Hydrogen tank	0.5 Kg

Table I: List of Hybrid DG System Components.

2.1 Energy from the Wind

The wind data considered in this simulation is based on the availability of wind at Mt. Clemens city, located at the Lake Huron front with coordinates 42° N and 82° W. The basic idea for this simulation was derived from a sample model on the HOMER webpage [17]. The power curve for the G20 wind turbine for the location considered is obtained from the wind resource data base of HOMER for a twenty four hour span and is given in Figure 3. The monthly wind profile for the whole calendar year is given in Fig. 4.



.

Figure 3: Power Curve of the G20 Wind Turbine.

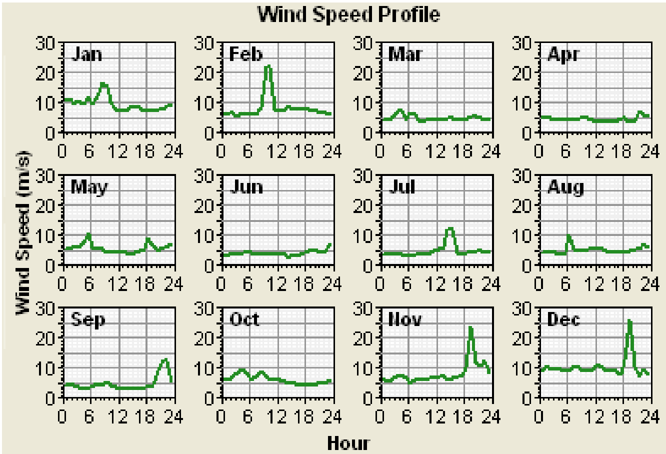


Figure 4: Annual Wind Speed Profile.

The output power of the wind turbine is given by (1) [18].

$$P_m = Cp\ (\lambda, \beta)\ (0.5)\ \rho\ A\ v^3_{wind} \tag{1}$$

The AC power generated from the wind turbine will be transmitted to PCU (to the AC bus in PCU which is then converted through converter in PCU for transmission).

2.2 Energy Storage Unit

Energy Storage unit, like a battery bank cab, stores the power when DG system is producing more than the load demand, and will serve as the backup power supply during the peak load demand periods. For cost effectiveness, conventional lead acid batteries are considered [19].

2.3 Fuel Cell

Primarily, a fuel cell is an electro- chemical device wherein the chemical energy of hydrogen is directly converted into electric energy. The fuel cell is much more efficient than thermal power plants, converting up to 60% of the chemical energy in the fuel cell into electricity, whereas the normal maximum efficiency is 40% for conventional power plants. The over-all efficiency of a fuel cell in terms of generated power can be estimated through (2) [20-24].

$$\eta = \eta t * \eta e * \eta r \tag{2}$$

In this modeling, fuel cell is working as back-up power source when the load demand is more than the available wind energy production and the stored energy in the energy storage unit.

2.3.1 Hydrogen Production and Storage

Hydrogen can be produced by the decomposition of water into its elementary components by passing the electric current. According to Faraday's law, hydrogen production of an electrolyzer can be estimated through (4) [22].

$$n_{H2} = 0.5\ \eta F * nc * ie\ /F \tag{3}$$

2.4 Load

An average demand of a residential load of 8.4 kWh/day is considered. In this analysis, the load is modeled with a few peak demands of almost 8.9 kW and a load factor of 0.5 over a span of 24 hours. The residential load profile of a similar load is obtained through HOMER database and is shown in Figure 5

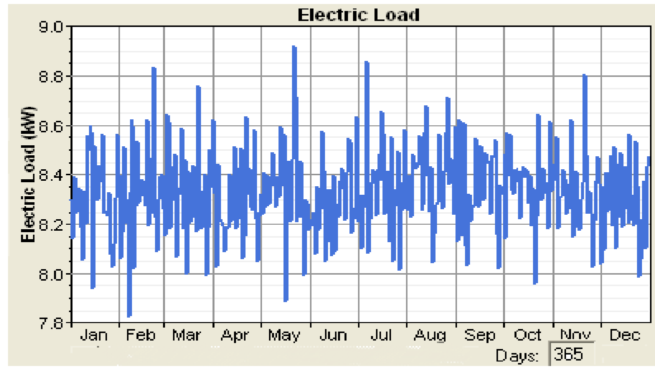


Figure 5: Load Variation of an Average House over a Year.

3.0 INTEGRATION OF HYBRID DG SYSTEM

The integration of hybrid wind-fuel cell DG system can be analyzed as a stand-alone Wind-Fuel cell system. The DC power produced from fuel cell is converted into AC power and fed to the AC Bus. The AC power generated from the wind turbines is directly fed to the AC Bus. The power conversion in the controller model is packaged in the power conditioning unit (PCU). Excess power is stored in the battery bank and to the electrolyzer, which generates hydrogen for storage in the hydrogen tank for the utilization by fuel cell in case of lack of generated power from wind source. The architecture of the generating side of the DG system simulated in HOMER is shown in Figure 6.

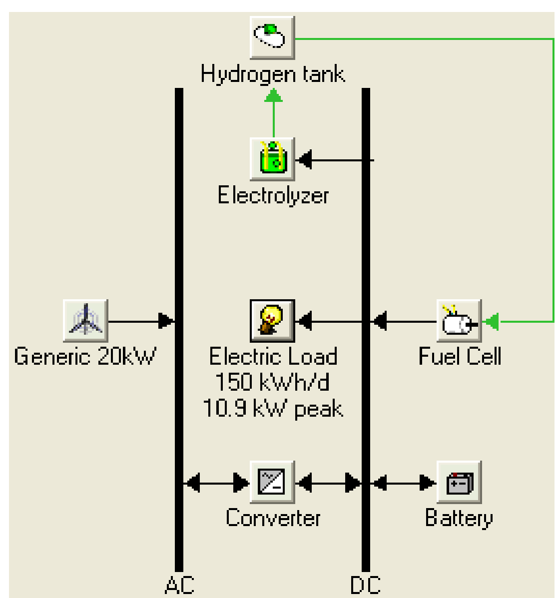


Figure 6: System Architecture of the Hybrid DG Systems.

The annual electric energy production and annual electric energy consumption is tabulated in Table II and III respectively. The production of power by individual renewable source is stated here as the percent fraction.

Renewable Energy Source	Energy Production (kWh/ yr.)	Annual Energy Fraction
Wind Energy	4368.2	82.9 %
Fuel Cell	898.1	17.1%
Total Energy	5266.3	100%

Table II: Annual Electric Energy Production.

Constraints	Energy Value (kWh/yr)
Excess Electricity	1065.2
Unmet Load	301.4
Capacity Shortage	263.3

Table III: Annual Electric Energy Consumption.

The results obtained through HOMER simulation can be considered promising. The excess load was 1065 kWh per year, unmet load was only 301.4 kWh and there is a capacity shortage of only 263.3 kWh annually. The major share in power production is of wind power with an annual estimate of 82.19%. The fuel cell has only a 17.1% share in the power production, which is obvious because only a 5 kW fuel cell was considered and also fuel cell is meant only for back-up power. However, as the consideration of equipment was done optimistically for the desired house load, further detailed economic analysis is required for practical implementation

4.0 DESIGN OF CONTROLLER

In order to transfer maximum power from the hybrid system at any time, a specific load has to be applied at that time. The flow chart depicting the control action is drawn in Figure 7.

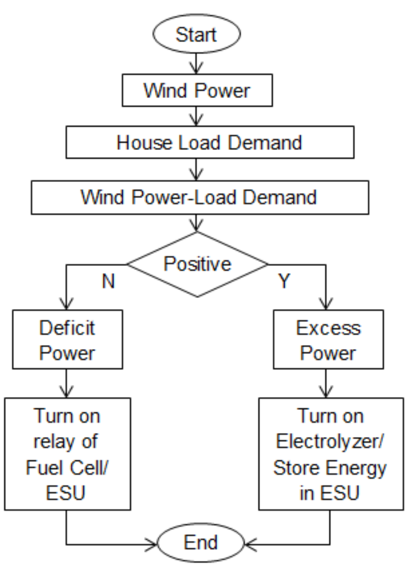


Figure 7: Flow Chart Governing the System.

The control output needs to control and monitor the power supply to the daily energy loads, activate the fuel cell and energy storage as a backup power supply and be able to turn on electrolyzer and maintain hydrogen flow from electrolyzer to fuel cell.

Due to the inconsistency of the wind energy, a power conditioning unit which consists of converters that provide a fixed output voltage by properly adjusting its duty cycle is implied to interface the hybrid system and the load. The power conditioning unit mainly consists of AC/DC and DC/DC and DC/AC converters. The detailed overview of the mechanism of the converters is outlined in [25]. For optimal power management, the control method implied to regulate such a system should have the following characteristics:

1. To minimize the error in the output voltage *V(k)* by appropriate control of the duty cycle *D(k)*,T
2. To provide a smooth control process near the reference point such that the transients in the controller output, i.e. duty cycle, should not affect the output of the dc-dc converter.

The control action of a conventional PI controller can be expressed as

$$u= K_P * e + K_I * \int_t edt \qquad (4)$$

Upon differentiation, the discrete-time based description of the above equation can be described as

$$\Delta u(k)= K_P * \Delta e(k)+K_I *e(k) \qquad (5)$$

The essential characteristics of the desired control system mentioned above must have variable gains over the range of operation near the reference point. This leads to a multi-PID controller causing chattering effect. This conventional multi-PID control system is not suitable for this application because of the continuous nature of variables and high-frequency switching requirement of dc-dc converter. A practical alternative for various challenging control applications for uncertain nonlinear dynamical systems is Fuzzy Control [26]. Fuzzy Control is based on Fuzzy set theory [27]. Basics of FLC are described in [28]. Through the processing of heuristic information, a fuzzy logic controller (FLC) interpolates among the consequent of all the rules according to their firing strength. Therefore, a FLC can be seen as multiple PID/PI controllers with smooth interpolation capability without chattering phenomena for real-time applications. For an effective micro grid, the successive-time state transition probabilities of the renewable energy sources are required to control in order to avoid any mismatch between the power supply and demand.

The block diagram in Fig.8 represents the proposed scheme for optimal power management of hybrid wind-PV fuel cell system under uncertain environmental conditions. The block diagram of the proposed hybrid wind fuel cell DG system consists of three power sources, supplementary hardware components and the buses. The power sources are (P1) wind energy, (P2) fuel cell and (P3) energy storage unit (ESU), while supplementary hardware components are electrolyzer and hydrogen storage. Functional buses are the power flow buses in black; control input bus in green, control output buses in red, and their relays R_1- R_3 for activating the respective power sources based on the controller output.

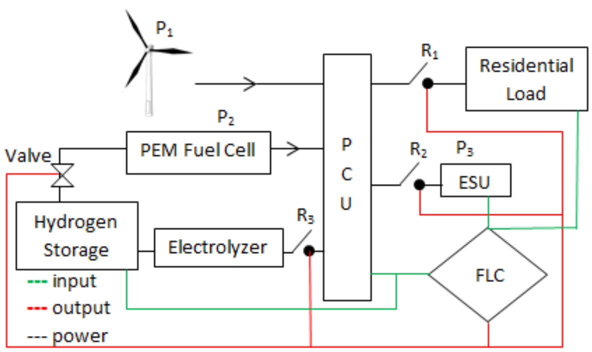


Figure 8: Schematic Diagram of Hybrid Fuzzy Logic Control

As evident from figure 8, the inputs to the FLC are the power available through the wind, fuel cell, energy storage unit and the power demand from the house. The instantaneous output voltage of the PCU ($V_{out}=V_k$) is compared with the reference voltage (V_{ref}). The error in voltage ($\Delta V_k= V_{ref} - V_{Boost}$) is input as error ($e_k$) and change in voltage error ($d\ (\Delta V_k)/dt = \Delta V_k - \Delta V_{k-1}$) is input as change of error (ce_k). The error 'e_k' and change in error 'ce_k' are input to the FLC. The output of F LC is the change in duty cycle 'du_k'. At any instant, duty cycle D_k is expressed as $u_k= u_{k\text{-}1}+ du_k$, which is used as control signal for switching of the converters in PCU and the relays for determining the actual load as seen by the PCU.

4.1 Execution of the Control Strategy

The Knowledge base of FLC consists of rule-base and database. The schematic diagram of an FLC is shown in Fig 9.

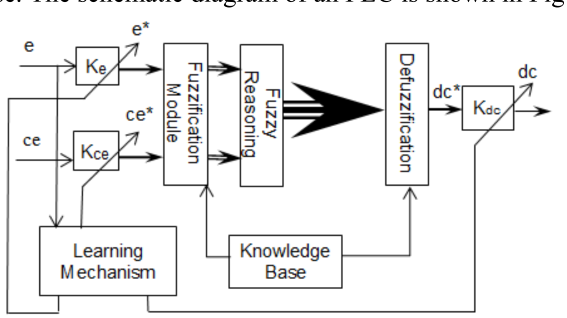


Figure 9:Schematic Diagram of Fuzzy Logic Controller [31].

The rule-base for an FLC can be developed by the observation of error and change of error of the system. Further, these rules have to be associated with the proper membership functions [28-35]. The universe of discourse (*UoD*) membership functions are normalized over the range *-1 to +1* as shown in Figure 10 and 11.

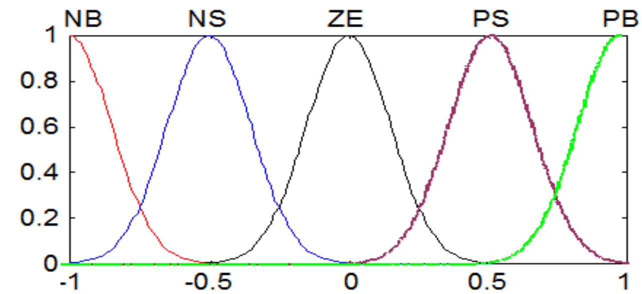


Figure 10: Gaussian Membership Functions for Inputs e & ce.

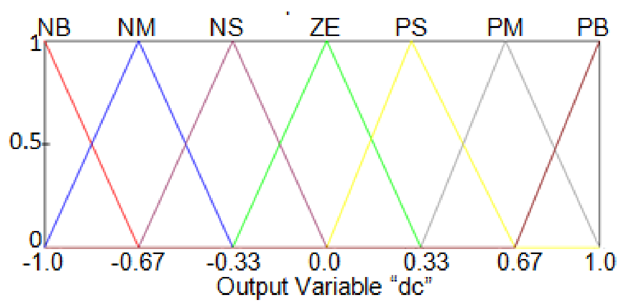


Figure 11: Triangular Membership Functions for Output du*.

In the proposed FLC, the error e_k , the change in error ce_k are partitioned into five fuzzy sets that lead to a maximum of twenty five rules, and the output variable du_k is partitioned into seven fuzzy sets. The choice of this number of portioning is to achieve tradeoff between the explosion of fuzzy rules and high *discretization* [28-31]. The comprehensive set of fuzzy membership function is plotted in Figure 12.

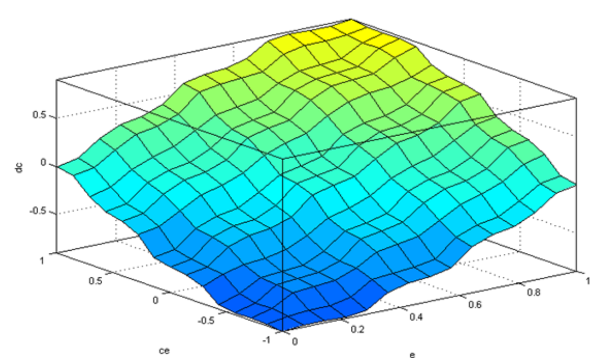


Figure 12: Surface of the Fuzzy Membership Function.

Based on the selection of input and out membership functions, a 25 rule base is developed. Entries in the set of rules are fuzzy sets associated with consequent variable *du* .The "product-sum" inference mechanism is used for mapping the rule-base to the consequent fuzzy set. The center-of sums method is used for the defuzzification of the FLC output. A sample rule of the rule-base is of the form:

IF e(k) is………AND ce(k) is……..THEN du is………. (6)

4.5 Simulation, Results and Discussion

A system level model based on the schematic in Figure 8 is developed in MATLAB/Simulink® environment. The component models for wind turbine, fuel cell, battery and converter available in the Simpower® toolbox of Simulink® are selected. The Fuzzy logic controller is then developed in conjunction with the components models based on the FKBC developed in section 4.4. A sample test case simulation is presented to demonstrate the efficacy of the proposed hybrid fuzzy logic control scheme.

Firstly the behavior of the load demand with the power generating capacity of the hybrid system based on the environmental condition is studied with HOMER and to analyze the performance of the FLC, simulation of the system model is observed for a duration of 2 hours with an assumption of constant temperature in fuel cell and electrolyzer. For a rated wind speed of 12.5 m/sec and 50% state of the charge of the battery bank and 3 phase ac load of a typical household, the phase voltage at the residential load obtained is shown in Figure 13. The balanced 3 phase input ac voltage available at the house load in Figure 13 through the adequate operation of the relays by the controller shows the successful operation of the proposed FLC.

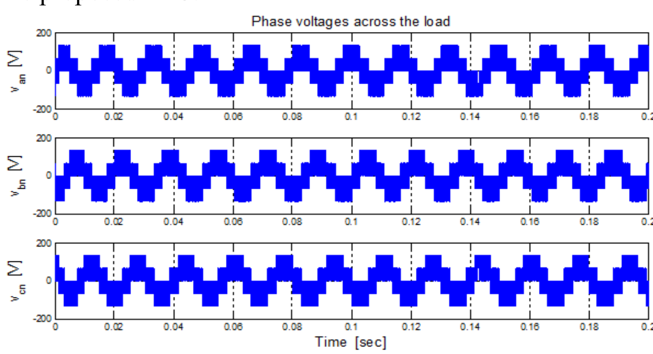


Figure 13: Phase Load Voltages.

Further efficacy of the FLC controller is evident from the concurrence of the phase a, b, c output load currents and the reference input current as shown in figures 14, 15 and 16 respectively.

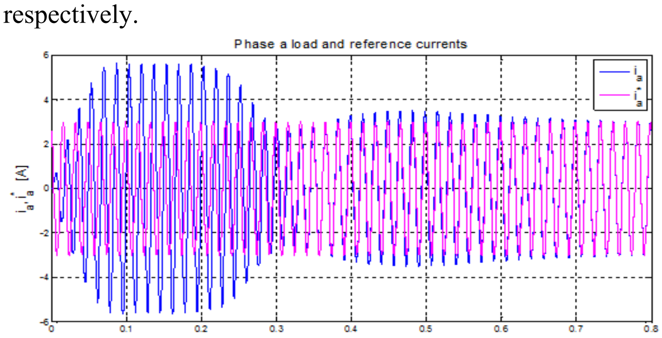


Figure 14: Phase *a* Load and Reference Current.

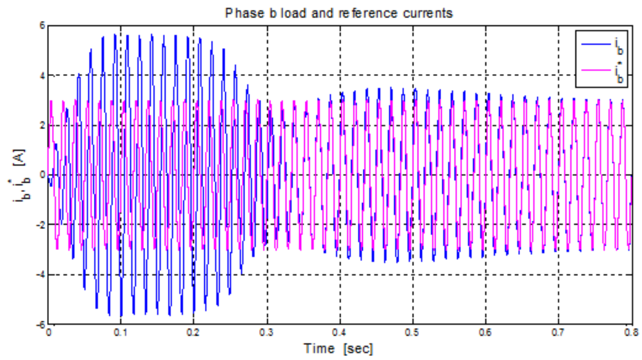


Figure 15: Phase *b* Load and Reference Current.

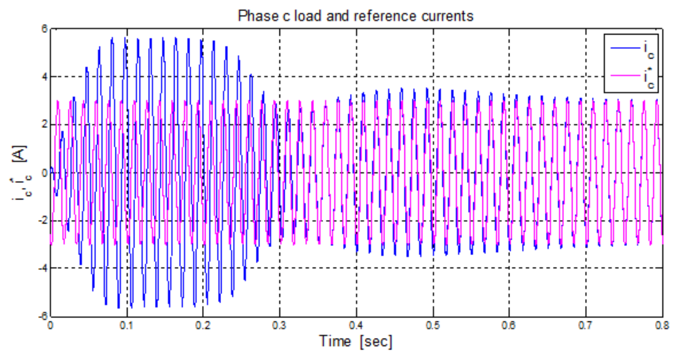


Figure 16: Phase *c* Load and Reference Current.

5.0 CONCLUSION

The application of hybrid wind-fuel cell based power system for consumption as stand-alone distributed generation systems, for a residential customer was studied. The feasibility study of the system was performed for various locations on the globe and a specific location of Michigan was presented. Based on the wind availability, hydrogen production and battery bank storage, it is concluded that such a system is a viable option for remote areas where there is no power supply available or it is very expensive to install transmission lines.

To optimize the power management of the hybrid wind-fuel cell system, a fuzzy logic based controller is developed. During the periods of adequate power availability from the renewable energy sources, FLC has exhibit a satisfactory performance in the power management. Further the control of the switching of the relays through FLC was successful by having a balanced 3 phase voltage and current available for the house load.

Hence, the proposed hybrid DG system is a promising option for remote areas and the proposed FLC has a tremendous potential in real-time applications of intelligent methods in sustainable power and energy systems.

REFERENCES

[1]. F. Farret, and M. Simões. Integration of Alternative Sources of Energy. Wiley-IEEE Press, 1st edition2006.

[2]. Statistics from International Energy Agency , http://www.iea.org/statistics/, accessed on April 4th, 2013

[3]. M. S. Alam, and A. T. Alouani, "Dynamic modeling of photovoltaic module for real-time maximum power tracking," J. of Renewable Science and Energy, AIP, NY, vol. 2, no. 3, pp: (043102)1-16, 2010

[4]. M.S. Alam, R.C. Chaban, "Realization of Hybrid-Electric Powertrain System for a Three Wheeler Auto Taxi," Buletin Teknik Elektro dan Informatika, ISSN: 2089-3191, Vol.1 No.2, pp: 131-138, June 2012.

[5]. M.S. Alam, B. Barnes, J. Hakel, A. Alqhtani, "Feasibility of Off-Grid Residential Photovoltaic System Utilizing Hydrogen Storage," Michigan Academy of Science Arts and Letters Annual Conference, University Center, MI, pp: 3-6, March 11, 2011.

[6]. B. H. Khan, "Non- conventional energy technologies", McGrawHill Publications, 2005.

[7]. M.S. Alam, D.W. Gao, "Modeling and Analysis of a Wind/PV/Fuel Cell Hybrid Power System in HOMER," proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA 2007), pp.: 1594-1599, Hurbin, China, May, 2007

[8]. T. Markvart, "Sizing of hybrid photovoltaic-wind energy systems," Journal of Solar Energy, Vol. 57, No.4, pp. 277-281, 1996.

[9]. R. Ramakumar, I. Abouzahr, K. Ashenayi, "A knowledge based approach to the design of integrated renewable energy systems," IEEE Trans. Energy Conversion, vol. 7, pp. 648–659, Dec. 1992.

[10]. D. B. Nelson, M. H. Nehrir, C. Wang, "Unit Sizing and cost Analysis of standalone hybrid wind/PV/fuel cell power generation systems," J. Renewable Energy, no. 31, pp.1641–1656, 2006.

[11]. H. Polatidis, Di. A. Haralambopoulos, "Renewable energy systems: A societal and technological platform," J. of Renewable Energy ,Vol. 32, 2007, pp:329–341

[12]. E. K. Stuart, "Energizing the Island Community: a Review of Policy Standpoints for Energy in Small Island States and Territories," J. of Sust. Dev. Vol. 14, 2006.

[13]. M.S. Alam, M.F. Azeem, "Economics and Sustainability of Generation, Storage and Distribution of Electric Power by a Hybrid PV-Wind System," Mid-South Area Engineering and Sciences Conference, Oxford, MS, USA, May 17-18, 2007.

[14]. Martin Pehnt, "Dynamic life cycle assessment (LCA) of renewable energy technologies", J. of Renewable Energy vol. 31, 2006, pp: 55–71

[15]. Michael Jefferson, "Sustainable energy development: performance and prospects," J. of Renewable Energy vol. 31, 2006, pp:571–582

[16]. Jukka V. Paatero, Peter D. Lund, "Effects of large-scale photovoltaic power integration on electricity distribution networks," J. of Renewable Energy, Vol. 32, 2007, pp: 216–234

[17]. Home page of HOMER: http://www.nrel.gov/homer/.

[18]. E.S. Abdin, W. Xu, "Control design and dynamic performance analysis of a wind turbine-induction generator unit," IEEE Trans. on Energy Conversion, vol. 15, no. 1, pp. 91–96, 2000.

[19]. M.S. Alam, M.F. Azeem, "Potential Role of Battery Energy Storage System in the Competitive Electricity Market," Mid-South Area Engineering and Sciences Conf. (MAESC 2007), Oxford, MS, USA, May 17-18, 2007.

[20]. Kaushik Rajashekara, "Hybrid Fuel Cell Strategies for Clean Power Generation," IEEE Transactions on Industry Applications, Vol. 41, No. 3, May/ June 2005.

[21]. M. S. Alam, T. Moeller, and A. Maly, "Converion of an Indian three wheeler scooter into a hybrid fuel cell Ni-MH battery vehicle and validation of the vehicle model for the Bajaj three wheeler scooter," in Proc. of the IEEE Conf. on Electric and Hybrid Vehicles,Pune 2006.

[22]. M.S. Alam, M.F. Azeem, "Facts, Issues and Future Perspectives of Hydrogen Technology," Mid-South Area Engineering and Sciences Conf. (MAESC 2007), Oxford, MS, USA, May 17-18, 2007.

[23]. M.S. Alam, D.W. Gao, "Conversion of a Golf car into Hybrid Fuel Cell Li-ion Battery Vehicle and Validation of the Vehicle Model for the 2006 GEM eL Golf Car Using PSAT," IEEE IES Conference, Mumbai, India, Dec. 15-17, 2006.

[24]. M.S. Alam, T. Taher, M. Khader, A. Lateef, R. Kizilel, "Design, Analysis and Hardware Development of a Novel Prototype Hybrid Hydrogen Fuel Cell and Lithium-ion Battery ZIPiR3 Senior Disability Vehicle," IEEE International Conference on Hybrid and Electric Vehicles, Pune, India, Dec. 18-20, 2006.

[25]. N.Mohan, T. Undeland, W. Robbins, "Power Electronics: Converters, Applications and Design," Wiley Pub.,3rd ed., ISBN-10:0471226939, 2003

[26]. E. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," Proceeding of the Inst. Elect. Eng., vol.121, no. 12, pp. 1585–1588, 1974.

[27]. L.A. Zadeh, "Outline of a new Approach to the Analysis of complex systems and decision process", IEEE Trans. Syst. Man. Cybern., vol. 3, no. 1, pp.28-44, 1973 & vol. 15, pp. 15-30, 1979

[28]. D.Driankov, H. Hellendoorn, M.Reinfrank, "An Introduction to Fuzzy Control," Springer Verlag, 2nd edition, ISBN-10: 3540606912, 1996

[29]. R.Li, Y.Zhang, "Fuzzy logic controller based on genetic algorithms," Elsevier Journal of Fuzzy Sets and Systems, vol. 83, no.1.pp.1-10, 1996

[30]. M. S. Alam, M. Azeem, "Modified Queen-Bee Based Evolution of GA for the tuning of scaling factors of FKBC", Proceeding of the IEEE India Annual Conference "INDICON 2004, India, pp.299-303, Dec. 2004.

[31]. M.S. Alam, M.F. Azeem, A. Alouani, "Modified Queen-Bee Algorithm-Based Fuzzy Logic Control for Real-Time Robust Load Matching for a Solar PV System," IEEE Transaction on Sustainable Energy- July 2013, doi: 10.1109/TSTE.2013.2271449 (Web version of

the publication is available online at http://ieeexplore.ieee.org)

[32]. A.K. Verma, Anil R, Om Prakash Jain, "Fuzzy Logic Based Revised Defect Rating for Software Lifecycle Performance Prediction using GMR" BIJIT-BVICAM's International Journal of Information Technology, Issue 1: (January-June, 2009) Vol.1 No.1; ISSN 0973-5658

[33]. Devendra K. Tayal, Amita Jain, Vinita Gupta," Fuzzy Expert System for Noise Induced sleep Disturbance and health effects", BIJIT-BVICAM's International Journal of Information Technology, Issue 3: (January-June, 2010) Vol.2 No.1; ISSN 0973-5658

[34]. Raj Gaurang Tiwari, Mohd. Husain, Raees Ahmad Khan, "Application of Fuzzy Relations in Convalescing Link Structure" BIJIT-BVICAM's International Journal of Information Technology, Issue 4: (July-December, 2010) Vol.2 No.2; ISSN 0973-5658

[35]. P.C. Jha, Shivani Bali, P.K. Kapur, " Fuzzy Approach for Selecting Optimal COTS based Software Products under Consensus Recovery Block Scheme" BIJIT-BVICAM's International Journal of Information Technology, Issue 5: (January-June, 2011) Vol.3 No.1; ISSN 0973-5658

Disclaimer:

This manuscript is based on author's individual research and it does not represent directly or indirectly the opinion of any business including author's employer.

Special Issue of BIJIT on Fuzzy Logic
BIJIT - BVICAM's International Journal of Information Technology
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Use of Wavelet-Fuzzy Features with PCA for Image Registration

Safia Sadruddin[1] and **Rashid Ali**[2]

Submitted in October, 2013; Accepted in February, 2014

Abstract- In this paper, we discuss an Image Registration system based on neural network, which uses Wavelet-fuzzy features of an image. In this system, Wavelet-fuzzy features are extracted from an image and then reduced using Principal Component Analysis (PCA). The reduced feature set is then used for training the neural network for image registration. The geometric transformation between the reference and sensed image sets are evaluated using affine transformation parameters. The trained neural network produces registration parameters (translation, rotation and scaling) with respect to reference and sensed image. Two parameters namely Mean Absolute Registration Error and Mutual Information are used as evaluation parameters. Experimentally, we show that the proposed technique for image registration is accurate and robust for distorted and noisy inputs.

Index Terms – Fuzzy Logic, Gaussian Noise, Image Registration, Neural Network, Principal Component Analysis (PCA), Wavelet Transform,

1.0 INTRODUCTION

Image Registration is required as a pre-processing step in many image processing tasks [1]. Image registration is required for comparing two or more images of the same scene taken at different time or from different viewpoints or from different sensors [2]. It plays an important role in disease diagnosis from medical images. The goal of image registration is to determine the geometric transformation that aligns the reference and sensed image. The alignment and integration of images, obtained from different sources and environment, helps to gain complementary information, which is not available from independent images. Registration of distorted and noisy images is a difficult task. Different techniques have been proposed for registration of noisy and distorted images. I. Elhanany et al. [3] proposed feed forward neural network using DCT (Discrete Cosine Transformation) for feature extraction. They extracted DCT coefficients from the lowest frequency band and fed them as input feature vectors to neural network. Feed forward neural network has also been used for image registration using Fourier transform for medical image registration in [4].

[1]*Department of Information Technology, Terna Enginnering College, Nerul, Navi-Mumbai, India.*
[2]*College of Computers and Information Technology, Taif University, Taif, Saudi Arabia.*
E-mail:[1] safia2007@gmail.com

Wavelet transform has been commonly used in image processing. P. Ramprasad et al. [5] proposed wavelet technique for matching noisy and poor contrast dental x-rays based on optimization of correlation coefficient of wavelet coefficients. In this, single level wavelet decomposition is performed and then correlation coefficients are calculated. The optimization is performed by maximizing the correlation coefficients. There technique performs better when compared to conventional and manual technique. It registers the images that differ only by rotation and translation but not scaling. Wavelet techniques has been used for registering multi-resolution [6] and multi-modal images [7].

Neural network based systems are reported to perform well when the feature sets are reduced using Principal Component Analysis (PCA) [8].

Fuzzy logic has been used in process control, management and decision making, operation research, economies, pattern recognition and classification. Lu et al. [9] used fuzzy set in multi-modal image registration. In this, Generalized Fuzzy Operator (GFO) is utilized to detect edge. Edges are detected using distance transform (DT). A distance transform of a binary image measures the distance of non-edge pixels to the nearest edge pixel while the edge pixels get the value zero. Fuzzy similarity measure is used to compute similarity measure between the reference and the target images.

In [10], Li et al. proposed fuzzy-wavelet feature extraction approach for pattern recognition. In [11], the authors presented a neural network based image registration technique that used wavelet-fuzzy features. In this paper, we present a neural network based image registration technique that used wavelet-fuzzy features along with PCA. The proposed technique is well suited for noisy and distorted images. Here, the wavelet-fuzzy features are extracted using wavelet transform and fuzzy set. The extracted features are then reduced using PCA. The reduced feature vectors are fed as input vectors to feed forward neural network (NN) and neural network is trained for image registration.

This paper is organized as follows. In Section 2.0, we discuss the proposed registration technique. In Section 3.0, we present our experimental results for noiseless and noisy images. Finally, we conclude in section 4.0.

2.0 PROPOSED REGISTRATION TECHNIQUE

The proposed technique for image registration works in two stages. In the first stage, affine transformation is applied and some noise is appended to the reference image. Additive White

Gaussian Noise (awgn) is added to improve the performance for distorted inputs. Wavelet transform and fuzzy set is then used to extract the features. In [10], wavelet-fuzzy has been used for feature extraction. These extracted features are then fed to feed forward neural network (NN) as an input. The feed forward NN with one hidden layer and four output layer nodes is then trained by considering the four affine transformation parameters as the four target outputs. We then save the layer weights of the trained network. This stage is called as training phase. The second stage is test phase. In this phase, we extract the features and feed them to the trained NN, and we get the estimated parameter values as output. The overall process is illustrated in Figure1.

The wavelet-fuzzy feature extraction process involves two steps namely (i) Pre-processing step and (ii) Wavelet-fuzzy feature extraction step.

Pre-processing step: To extract the features using the proposed feature extraction method, some preprocessing of images is done. Suppose that an image has *M x N* pixels. The image matrix can be reshaped to 1 x (*M.N*) vector, given as-

$(a_{1,1},a_{1,2},\ldots,a_{1,N},a_{2,1},a_{2,2}\ldots a_{2,N},\ldots\ldots\ldots,a_{M,1},a_{M,2},\ldots a_{M,N})_{1\mathrm{x}(M.N)}$

where $a_{i,j}$, i =1,2,3,…M and j=1,2,3,….N. This reshaped vector is to be used for feature extraction.

Fuzzy logic → PCA
g

Wavelet-Fuzzy Feature Extracti decomposed to a single level approximation signal and detail signal using 'haar' wavelet transform as shown in figure 2.

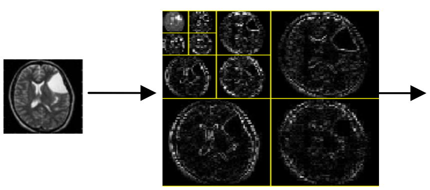


Figure 2

The data size of both the approximation and detail signals at the resolution 2^j are N_0 x 2^j after the j^{th} decomposition. The approximation-detail pairs ($h_{x_A,2^j}(k), h_{x_D,2^j}(k)$), k=1, 2,.N_0 x 2^j is used in wavelet-fuzzy feature extraction.

The linguistic value sets for x_A , x_D – the approximation and detail signals respectively, are given as-

$$T_A = \begin{pmatrix} v_{A,1} \\ v_{A,2} \\ \vdots \\ v_{A,c_A} \end{pmatrix}, \quad T_D = \begin{pmatrix} v_{D,1} \\ v_{D,2} \\ \vdots \\ v_{D,c_D} \end{pmatrix} \qquad (1)$$

where T_A and T_D are the term sets for the approximation and detail signal respectively; $v_{A,k}$ and $v_{D,l}$, k = 1,2,…….c_A and l=1,2,……c_D are the linguistic values for x_A,x_D , respectively; c_A and c_D are cardinalities for T_A and T_D, respectively. The corresponding membership function is denoted as follows:

$$\mu_A(h_A(k)) = \begin{pmatrix} \mu_{A,1}(h_A(k)) \\ \mu_{A,2}(h_A(k)) \\ \vdots \\ \mu_{A,c_A}(h_A(k)) \end{pmatrix}, \mu_D(h_D(k)) = \begin{pmatrix} \mu_{D,1}(h_D(k)) \\ \mu_{D,2}(h_D(k)) \\ \vdots \\ \mu_{D,c_D}(h_D(k)) \end{pmatrix} \qquad (2)$$

With the two parameters K_A and K_D for tuning sensitivity and robustness, here we have taken K_A = K_D =1, the feature vector of a signal can be expressed as follows:

$$F = \begin{bmatrix} \sum_{k=1}^{2^j N_0} \wedge\left(\mu_A^1\left(\frac{h_{x_A,2^j}(k)}{K_A}\right), \mu_D^1\left(\frac{h_{x_D,2^j}(k)}{K_D}\right)\right) \\ \sum_{k=1}^{2^j N_0} \wedge\left(\mu_A^2\left(\frac{h_{x_A,2^j}(k)}{K_A}\right), \mu_D^2\left(\frac{h_{x_D,2^j}(k)}{K_D}\right)\right) \\ \vdots \\ \sum_{k=1}^{2^j N_0} \wedge\left(\mu_A^{c_A}\left(\frac{h_{x_A,2^j}(k)}{K_A}\right), \mu_D^{c_D}\left(\frac{h_{x_D,2^j}(k)}{K_D}\right)\right) \end{bmatrix} \qquad (3)$$

After extracting features from image using wavelet-fuzzy, the next step is to reduce the features using Principal Component Analysis (PCA). PCA helps to reduce redundant data and transforms correlated variables into smaller number of uncorrelated variables using largest eigenvectors of the correlation matrix. Following algorithm is used to reduce the feature vectors. The reduced feature vectors are used as an input to feed forward NN.

PCA Algorithm:

Let X be a data set of dimension M x N [12]

Step 1: Calculate the empirical mean:

$$u[m] = \frac{1}{N}\sum_{n=1}^{N} X[m,n]$$

Step 2: Calculate the deviations from the mean:

$$B=X-uh$$

where h is a 1xN row vector of all 1's:

h[n]=1 for n=1……N

Step 3: Find the covariance matrix C:

$$C = \frac{1}{N} B.B*$$

Step 4: Find the eigenvectors and eigenvalues of the covariance matrix C:

$$[D,V]=V^{-1}CV$$

where, D is the diagonal matrix of eigenvalues of C and V is the eigenvector matrix. Matrix D will take the form of an MxM diagonal matrix, where

$D[p,q]=\lambda_m$ for $p=q=m$

Step 5: Rearrange the eigenvectors and eigenvalues: Sort the columns of the eigenvector matrix V and eigen value matrix D in order of decreasing value.

Step 6: Select a subset of the eigenvectors as basis vectors: Save the first L columns of V as the M x L matrix W:
W[p,q]=V[p,q] for p=1……….L where $1 \le L \le M$

3.0 EXPERIMENTAL RESULTS

A single MRI image of brain is used to produce the training data set using different parameter values for rotation, scaling, vertical translation and horizontal translation. The transformation parameter values for training images are given in Table I. A total of 256 images are generated, each of size 128 by 128 pixels. Features are extracted using wavelet-fuzzy feature extraction steps. These feature vectors are used to train the neural network. The test data set contains 81 images. Features are extracted from the test images and fed to the trained NN.

The transformation parameter values for test images are given in Table II. The transformation values for training and test sets are same that has been used by H. Sarnel et al. [13]. Figure 3 shows the original image and transformed images after applying rotation, scaling and translations and also a noisy image.

The designed neural network has 7 neurons in one hidden layer and 4 neurons in the output layer. Hidden layer has a tangent sigmoid transfer function and linear function for the output layer neurons, and Levenberg Marquardt method was used for training.

Mean Absolute Registration Error (MAE) and Mutual Information (MI) is calculated for 81 images in test set for image registration with PCA and without PCA. Table III shows MAE values and Table IV shows MI values for without noise images, 5db noise images and 20db noise images. From these results, it can be concluded that using wavelet-fuzzy features after feature reduction with PCA improves the performance for noisy and distorted images when compared to using the wavelet-fuzzy features directly without PCA.

Values used for training set	Transform parameter
0.9,0.965,1.035,1.1	Scale
-5,-2,2,5	Rotation (degrees)
-5,-2,2,5	Vertical translation (pixels)
-5,-2,2,5	Horizontal translation (pixels)

Table 1: Affine transformation parameter values for training set

Values used for test set	Transform parameter
0.93,1,1.07	Scale
-3,1,4	Rotation (degrees)
-4,0,3	Vertical translation (pixels)
-3,1,4	Horizontal translation (pixels)

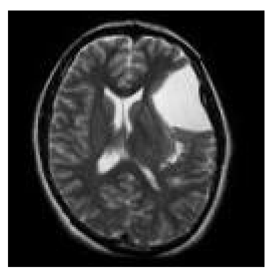
Table 2: Affine transformation parameter values for test set

Feature Extraction Technique	Without Noise	5db Noise	20db Noise
Wavelet+fuzzy	1.28416	1.318225	1.308850
Wavelet+Fuzzy+PCA	1.26834	1.264930	1.2553912

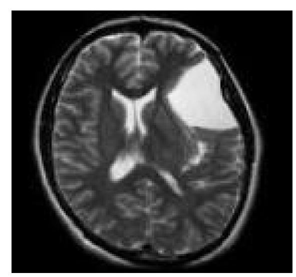
Table 3: Mean Absolute Registration Error

Feature Extraction Technique	Without Noise	5db Noise	20db Noise
Wavelet+fuzzy	0.874757	0.936966	0.930331
Wavelet+Fuzzy+PCA	1.012543	1.008405	1.010163

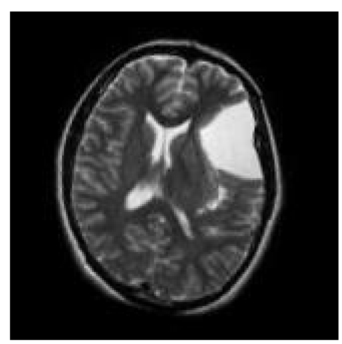
Table 4: MI values for Image Registration



(a)



(b)



(c)

(d) (e)

Figure 3: (a) Original Image (b), Transformed, rotated and scaled Image (c) Transformed, rotated and scaled Image (d) Translated, rotated, scaled images with 5db noise (e) Translated, rotated, scaled images with 20db noise

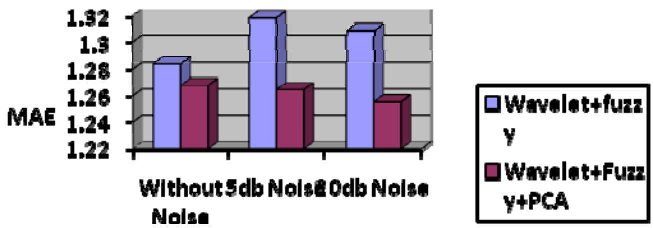


Figure 4: Mean Absolute Registration Errors

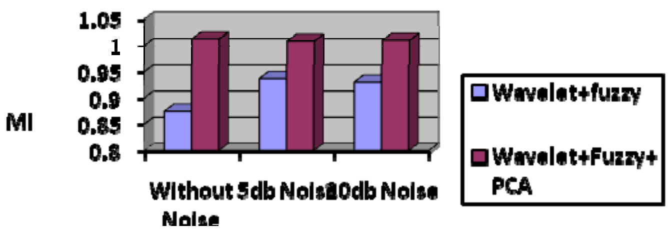


Figure 5: MI values for Image Registration

4.0 CONCLUSION

In this paper, we have discussed the use of wavelet transform and fuzzy set for feature extraction for image registration. The extracted features are reduced using PCA. The reduced feature vectors are then used to train the NN. According to the experimental results, the proposed method is more efficient and robust to noise and distorted inputs. We compare the results of the proposed NN based image registration system with the results of the system where PCA is not used to reduce the extracted wavelet-fuzzy features. Experimentally, we find that reducing the extracted wavelet-fuzzy features using PCA as in the proposed system improves the performance of the neural network based image registration with smaller mean absolute registration error and larger mutual information values.

REFERENCES

[1]. L.G.Brown, "A survey of Image Registration Techniques", ACM Computing Survey, vol. 24, pp. 325-376, 1992.

[2]. Barbara Zitova and Jan Flusser, "Image Registration Methods: a survey", Image and Vision Computing 21, pp.977-1000, 2003.

[3]. I. Elhanany, M. Sheinfeld, A. Beck, Y. Kadmon, N. Tal and D. Tirosh, Robust Image Registration based on FeedForward Neural Networks", IEEE International Conference on Systems, In Man and Cybernetics, vol.12, pp.1507-1511, 2000.

[4]. A.B. Abche, F. Yaacob, A. MaalouKaram,"Image Registration based on Neural Network and Fourier Transform", Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, pp.4803-4806,2006.

[5]. P. Ramprasad, H. C. Nagaraj and M. K.Parasuram, "Wavelet based Image Registration Technique for Matching Dental X-rays", International Journal of Computer Science 4:2 2009.

[6]. N. Kaewkamnerd, K.R. Rao, "Wavelet Based Watermarking Detection Using Multiresolution Image Registration", TENCON 2000 Proceedings, vol.2,pp.171-175,2000.

[7]. Ghantous, M., Ghosh, S., and Bayoumi, M., "A Multi-Modal Automatic Image Registration Technique based on Complex Wavelets", 16thIEEE International Conference on Image Processing (ICIP'09), pp.173-176, 2009.

[8]. Safia Sadruddin and Rashid Ali, "Neural Network Based Image Registration Using Wavelet Transform and PCA," Proc. of the International Conference on Intelligent Systems and Data Processing (ICISD 2011), Vallabh Vidyanagar, Gujarat, India, ISBN: 978-1-6123-3002-0, pp. 259-263, January 2011.

[9]. Zhentai Lu, Minghui Zhang, Qianjin Feng, Pengcheng Shi and Wufan Chen, "Image Registration Based on Fuzzy Similarity", IEEE/ICME International Conference on Complex Medical Engineering, pp. 517 – 520, -2007.

[10]. Chunsihen Li, Jyh-Yann Huang and Chih-Ming Chen, "Soft computing approach to feature extraction", Fuzzy sets and Systems 147 2004, pp. 119-140, 2004.

[11]. Safia Sadruddin and Rashid Ali, "Neural Network Based Image Registration Using Wavelet-Fuzzy Features", Proc. of the 7th National Conference on Advancement of Technologies Information Systems & Computer Networks (ISCON-2012), Mathura, India, March 2012.

[12]. El-Sayed Ahmed El-Dahshan , Tamer Hosny and Abdel-Badeeh M. Salem,"Hybrid intelligent techniques for MRI brain images classification", Digital Signal Processing, vol.20, pp. 433–441, 2010.

[13]. H.Sarnel, Y.Senol and D. Sagirlibas, "Accurate and Robust Image Registration based on Radial Basis Neural Networks", IEEE International Conference on Systems, Man and Cybernetics, pp. 204-207, 2008.

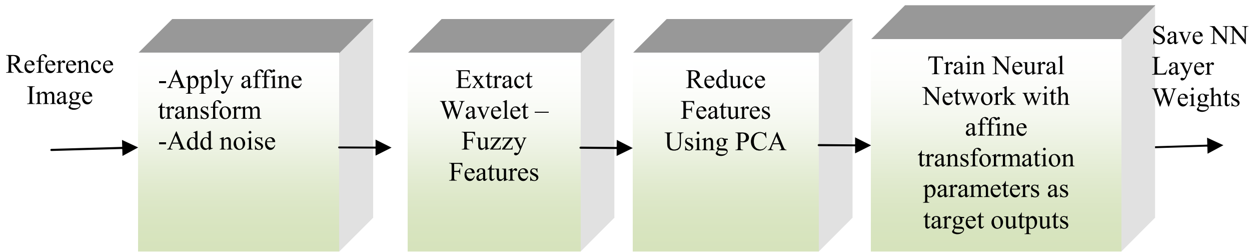


Figure 1(a): Training Phase

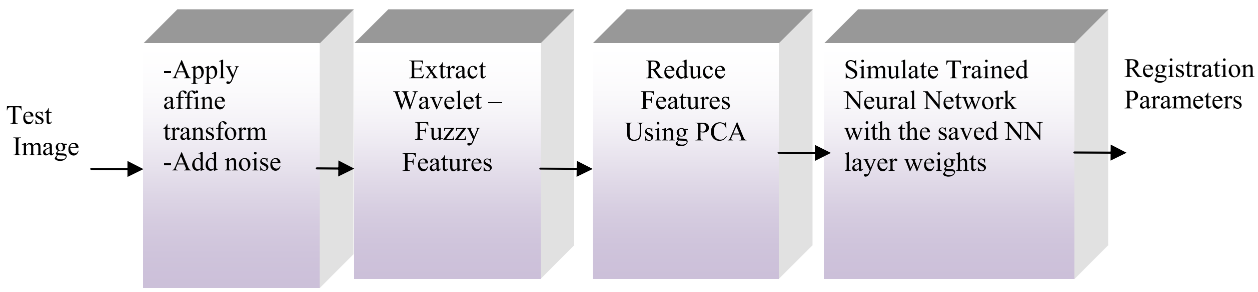


Figure 1(b): Test Phase

Figure 1: Proposed Image Registration

Special Issue of BIJIT on Fuzzy Logic
BIJIT - BVICAM's International Journal of Information Technology
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Zadeh-Deshpande Approach for Fuzzy Description of Air and Water Quality

Ashok Deshpande[1], Jyoti Yadav[2] and Vilas Kharat[3]

Submitted in September, 2013; Accepted in February, 2014

Abstract - Ever increasing pollution levels due to rapid urbanization and industrialization, especially in many developing countries with minimal focus on adequate pollution abatement strategies, has resulted in impairing the natural environment such as river, air, land, and alike. It is, therefore, important to classify the environmental quality. The practice in vogue for classifying air or water quality for variety of usage is by computing Air Quality Index (AQI) or Water Quality Index (WQI). Why compute a numeric AQI or WQI, and then describe air/ water quality linguistically? Human brain does not compute any numbers. Why not describe air or water quality, for the defined usage, straightway in linguistic terms with some linguistic degree of certainty attached to each linguistic description?

In this paper, we present two need based research studies. The first case study refers to fuzzy air quality description in Pimpri-Chinchwad Municipal Corporation (PCMC) monitoring location, while the other relates to linguistic classification of water quality with degree of certainty in PCMC area, India,

Index Terms – Air quality, water quality, uncertainty, Linguistic term, degree of match, fuzzy logic, fuzzy rule base system, degree of certainty

1.0 INTRODUCTION

Information on the status of environmental quality is necessary to formulate sound public policies and effective implementation of environmental quality programs. One of the effective ways of communicating such information in general, and air and water quality in particular to policy makers and ultimately to the end users, is with indices.The efforts on the development of water quality indices -expressed in an interval scale in numeric terms were initiated in 1990's in USA. Air/ water quality index (AQI or WQI) is in essence a function used to simplify large quantities of data into a more useful form which might convey an image of overall air/water quality to variety of users.

[1]*Berkeley Initiative in Soft Computing (BISC) - Special Interest Group (SIG) - Environment Management Systems (EMS), College of Engineering Pune (COEP) Pune India*
[2,3]*Department of Computer Science, University of Pune*
E-mail: [1]ashokdeshpande@hotmail.com,
[2]jyo_yadav@yahoo.co.in and [3]laddoo1@yahoo.com

The limitation of the selected indices is their lack of power in dealing with uncertainty and subjectivity present in the expert's perception in describing air/water quality in linguistic terms such as: *Good, Fair, Poor and so on* for the defined purpose. In spite of the best efforts, there has been no general acceptance of AQI/WQI for the intended usage.

There exists *imprecision / fuzziness* in the perception of domain experts' about rating environmental quality parameters in various linguistic terms as *epistemic uncertainty* -an unavoidable feature of most humanistic systems. Uncertainty is also present due to inherent variability in randomness in environmental quality parametric data and is known as - statistical or *aleotary uncertainty*. We believe that fuzzy logic concepts could be considered as a useful formalism for modeling, both, *aleotary and epistemic uncertainty.*

The paper is organized as follows: Section 1 is Introduction covering the relevance of the proposed need based research. The write up in Section 2 describes fuzzy logic based formalism used in the study. Application of fuzzy logic concepts in two case studies is presented in Section 3. The first part of the study refers to air quality description in PCMC with linguistic degree of certainty attached to each linguistic description while the second part describes water quality in Pawana river near PCMC intake well straightway in linguistic terms with some degree of certainty The results and discussion are also included in this section. Section 4 relates to the conclusion and the need for further research.

2.0 THE METHODOLOGY

Fuzzy logic based approach used in describing air and water quality in PCMC area is detailed in this section.

2.1 Type 1 Fuzzy Inference System (FIS) with Degree of Match (DM)

Randomness in air quality data or for that matter any parametric data, can be represented using Gaussian distribution with probability density function given by

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

where x is a random variable taking values on real line, μ is the mean and σ is the standard deviation. The continuous random variable can be discretised using distribution

$$\Phi(x) = P[X \le x] \quad (2)$$

such that probability of any x_i in the parametric domain, denoted by $p(x_i)$ is :

$$p(x_i) = \phi\left(x_i + \frac{dx}{2}\right) - \phi\left(x_i - \frac{dx}{2}\right) \quad (3)$$

where dx is a small interval about x_i for i = 1,2,3,4,……n.
Thus if x_i is some point on the parametric domain for which p (x_i) is maximum, then define function $\mu_A(x)$ as:

$$\mu_A(x) = p(x)/p(x_i) \quad (4)$$

The operation in equation 6 transforms a random variable into a convex normalized fuzzy number A with membership grade $\mu_A(x)$, thereby characterizing the dynamic behavior of the governing parameter. The construction of fuzzy number or convex fuzzy sets for modeling perception of the experts' in classifying each parametric domain linguistically, involves:

- Selection of linguistic terms such as Very Good, Good, Fair etc which allows for referencing all possible parametric values to be described.
- Classification of the parametric domain and assigning linguistic terms to each class linearly by the experts', reflecting imprecision in their perception. The set of values for which all the experts' assign the same linguistic term are given membership value as $\mu = 1.0$ while none of the expert assigning that term are given membership value as $\mu = 0.0$. The breakeven point membership grade 0.0 and 1.0 are connected by continuous monotonic function which presupposes that the degree of consensus amongst the experts' goes on increasing as the parametric values approach the core of the fuzzy number for the specified linguistic term.

2.2 Matching between two fuzzy values

The fuzzy number for pollutant data (A) on parameters and the fuzzy number characterizing linguistic terms ($\tilde{A}$) are matched together to arrive at a measure called Degree of Match (DM) (Figure 1 and 2) defined by :

$$DM_{ff}(A, A') = \frac{\int \mu_{A\cap A'}(x)dx}{\int \mu_A(x)dx}, \; x \in X \quad (5)$$

in which X denotes the universe, and $\mu_{A\cap\tilde{A}}(x)$ is the membership grade for $A \cap \tilde{A}$. A (fact or the data) and $\tilde{A}$ (expert's perception) are the possibility distributions the measure is defined as:

$$DM_{ff}(A, A') = \frac{\sum \mu_{A\cap A'}(x)dx}{\sum \mu_A(x)dx}, \; x \in X \quad (6)$$

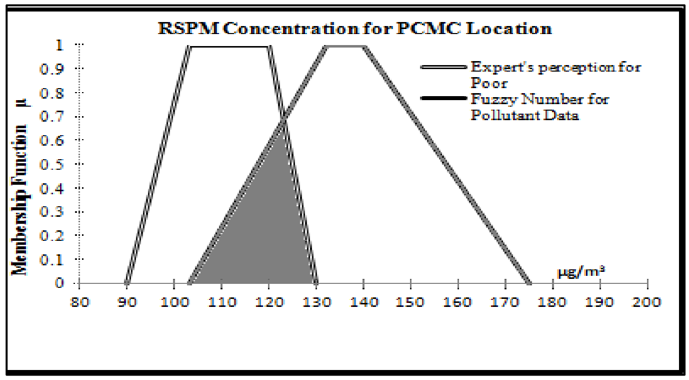


Figure 1: Degree of Match of RSPM with Linguistic Term Poor for PCMC location

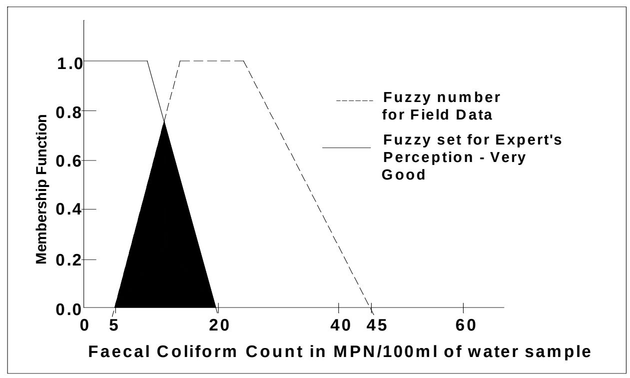


Figure 2: Fuzzy matching between A and A'

2.3 Fuzzy Rule Based System

The computational scheme of Degree of Match (DM) can be used with a view to estimate matching between expert's perception and the antecedent part of the rule, in order to describe air quality fuzzily with certain degree of certainty [3]. The degree of match of each classification rule indicates the certainty value of classification, in the present case: air and water quality. The greater the degree of match, the greater is the possibility that air and water quality is classified in that class. A fuzzy rule based system was developed for knowledge representation or reasoning process. A set of fuzzy rules is constructed for classifying air quality as: *very good, good, fair, poor and very poor* in order to aggregate the set of attributes. These linguistic descriptions are invariably imprecise / vague / fuzzy keeping in view the inadequate information on the health implications of each parameter on the users and the aggregated effect of all the parameters on human health. Sample rules displayed in Table 1 were stored in the knowledge base. The rules are processed using conjunction and disjunction operators. The optimal acceptance strategy is usually that for which the degree of assertion is the maximum [4]. Figure 3 presents the hierarchical structure for water quality classification [6] and Figure 4 represents the Fuzzy Inference System for Air Quality classification.

.

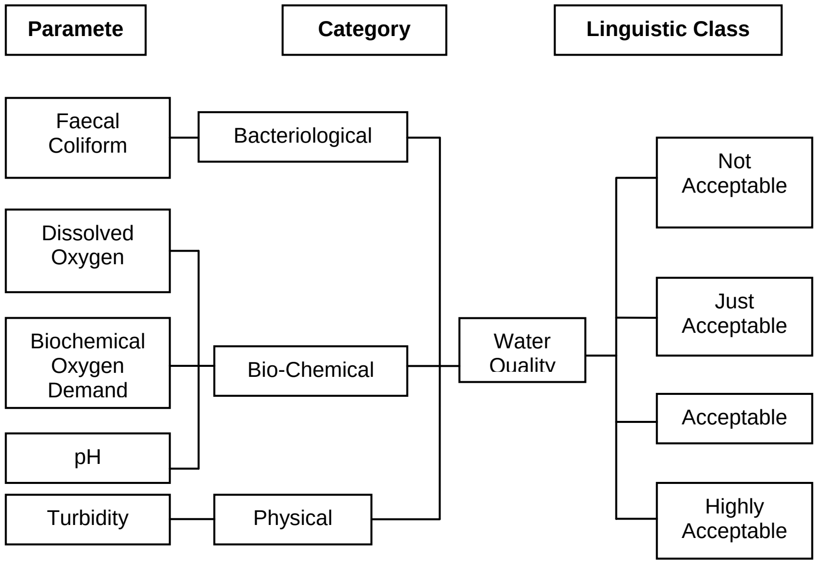


Figure 3: Hierarchical structure for water quality classification

Sr. No.	Air Quality Fuzzy Rules	Water Quality Fuzzy Rules
1.	If RSPM is Good AND NO_x is Poor AND SO_x is Good THEN Air Quality is Good	IF BOD is *Good* AND DO is *Very Good* AND pH is *Very Good* THEN Biochemical status of water is *Good*
2.	If RSPM is Fair AND NO_x is Good AND SO_x is Fair THEN Air Quality is Fair	If Bacteriological status of water is *Fair* AND Biochemical status of water *Very Good* AND Physical status of water *Very Good* THEN over all water quality at water intake well is *Fair*.

Table 1: Fuzzy Rules by Air and Water Quality Domain Expert's

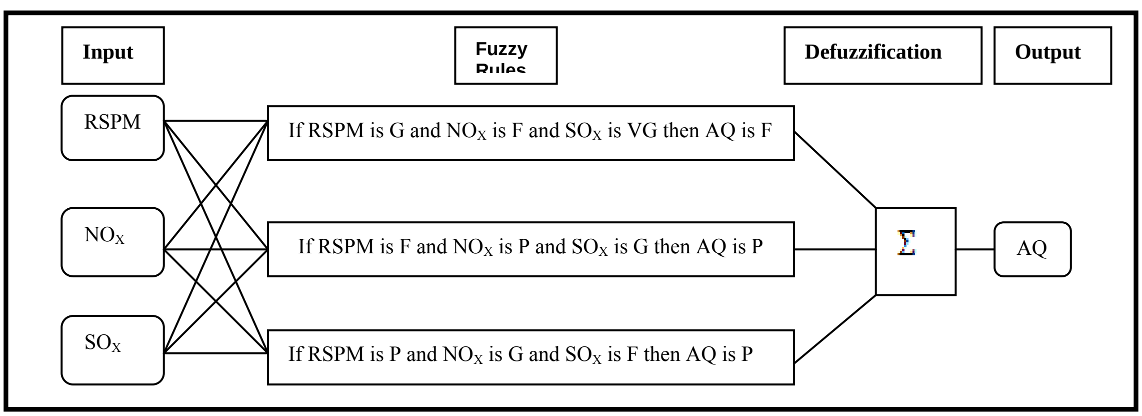


Figure 4: Fuzzy Inference System for Describing Air Quality

3. 0 CASE STUDY

Application of fuzzy set theory has been a long standing need while dealing with engineering system in general and air and water quality classification in particular.

3.1 Linguistic description of Air Quality with Linguistic Degree of Certainty

Air pollutants are added in the atmosphere from variety of sources that change the composition of atmosphere and affect the biotic environment. The concentration of air pollutants depends not only on the quantities that are emitted from air pollution sources but also on the ability of the atmosphere either to absorb or disperse these emissions. The air pollution concentration vary spatially and temporarily causing air pollution pattern to change with different locations and time due to changes in meteorological and topographical condition. The sources of air pollutants include: vehicular traffic, emissions of hazardous gases from industries, domestic sources and natural sources [7].

Air pollution is the *fifth* leading cause of death in India after high blood pressure, indoor air pollution, tobacco smoking and poor nutrition, with about 620,000 premature deaths occurring from air pollution related diseases (Times of India News dated 14/2/2013). Especially in the developing countries, a large portion of vehicles on city roads use diesel fuel, and contribute greatly to the emissions of particulates, especially those that are less than 10 microns in size and are respirable. In many cities, transportation fuels contain lead and high amounts of sulphur and use older engine designs that emit more toxic pollutants than modern ones. Various other factors like predominance of old outdated vehicles, lack of maintenance, limited use of emission control technologies, poor traffic management and poor road conditions helped accentuate the level of automobile pollution. The number of two/three wheeler vehicles which use two stroke engines and produce up to 10 times size hydrocarbons are on increase [7].

Air quality monitoring network design needs structured approach which involves : identification of pollutants, selection of locations, frequency and duration of sampling, sampling techniques, infrastructural facilities, man power operation, maintenance costs, and alike. The design also depends on type of pollutants in the atmosphere through various common sources, called common urban air pollutants- such as Suspended Particulate Matter (SPM), Respirable Suspended Particulate Matter (RSPM), Sulphur Dioxide (SO_2), Oxides of Nitrogen (NO_x), and Carbon Monoxide (CO) etc. Factors to be considered for the identification of air quality monitoring locations include: high traffic density, industrial growth, human population and its distribution, emission source, public complaints, if any, the land use pattern etc. Realising the increase in air pollution levels, it was considered relevant to initiate a study to classify the air quality in PCMC.

The case study relates to fuzzy air quality description with the available air quality data. The locations of air quality monitoring stations are: 1. Pimpri-Chinchwad Municipal Corporation, and 2. Bhosari in PCMC. Auto exhaust pollution is on increase in PCMC area. From air quality standpoint, winter months are more important due to temperature inversion phenomenon. November 2010 air quality parametric data is considered for the assessment of ambient air quality at PCMC location. Following three pollutants monitored by the Maharashtra State Pollution Control Board (MPCB) viz. Oxides of Sulphur (SO_x), Oxides of Nitrogen (NO_x), and Re-Suspended Particulate Matter (RSPM) is used for the assessment of air quality following the concept of fuzzy inference system with degree of match as detailed in this paper.

3.2 Fuzzy description of Pawana river water quality near PCMC intake well

The developing countries have been witnessing pollution of water resources which has assumed a serious threat to mankind due to increase in the incidence of water related diseases. The World Health Organization (WHO) in their report (2002) states that around 21% of communicable diseases in India are water related, and out of these, diarrhoeal diseases alone killed over 7,00,000 Indians in the year 1999 [2]. The present study refers to the water quality near intake well of water treatment plant of PCMC India. PCMC depends mainly upon Pawana river water. Water quality surveillance from the source to consumer tap is routine activity of major water supplies. In India, the issue of sample size using statistical methods is invariably ignored. How many samples to be collected from a sampling location? In this study, we have implemented the concept which is described below:

Total coli form is one of the important parameters in describing water quality for drinking purpose. A monitoring study was intended to estimate the mean concentration of total coli form near the intake well of PCMC near Water Treatment Plant. A preliminary survey consisting of fourteen (n=14) representative observations, were: 1200, 1300, 1600, 1100, 400, 600, 1800, 300, 1000, 440, 350, 120, 460, 1250 expressed in MPN/100ml of sample. Biochemical Oxygen Demand (BOD) expressed as mg/l were also in the range of 6- 18 mg/l, indicating pollution at the intake point on Pawana river. Looking at these high values, the then Municipal Commissioner (MC) who is also an engineer, requested the first author on suggesting the sample size as he was keen to direct PCMC officials to carry out sampling based on basic statistics. The purpose of these investigations was to initiate action on the defaulters -a sugar mill owner, sewage from nearby colonies, villages to take necessary corrective measures.

1. What sample size is needed to estimate the true mean within ± 200 units? Also, suggest a remedial measure to bring down drastically the bacterial load i.e. say, ≥ 500 MPN/100 ml of water sample.

Sizing the Experiment or sample size

Assume: data will follow a normal distribution.

$$\text{Number of Samples } N = \frac{(Z_{\alpha/2}\,\sigma)^2}{E^2} \tag{7}$$

where, σ is the standard deviation(524 MPN/100ml of water sample); Z is $[(x_i - \bar{x})/\sigma]$ =1.96 ; for α/2 equals 0.025 and E (true mean within ± 200 units) = half length E=200. The computed number of samples will be = 26.37, say a total of 30 samples should be collected. Therefore PCMC officials should collect additional (30-14) = 16 sample for the estimation of total coli form.

Similar procedure has been followed for deciding the number of samples for all the other water quality parameters, viz. Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), pH, Turbidity and was used in the deciding the water quality near PCMC intake well using fuzzy logic based formalism.

4.0 RESULTS AND DISCUSSION

The Air Quality description with Degree of Match method describes air quality straightway in linguistic terms with *linguistic* degree of certainty. Uncertainty of uncertainty is modelled using type I fuzzy logic. Firstly the uncertainty in the expert's perception is modelled and then the second uncertainty related to Degree of Certainty (DC) is modelled by defining fuzzy sets for DC as Very Good, Good, Fair, Poor and Very Poor.

Location	CAQI	AQ Description and DC in Linguistic terms			
		Expert 1	Expert 2	Expert 3	Expert 4
PCMC	151.13 (VP)	Very Poor (1.00-VG)	Very Poor (0.92-VG)	Poor (0.92-VG)	Poor (0.88-G)
Bhosari	100.38 (P)	Poor (0.53-F)	Poor (0.30-P)	Fair (0.30-P)	Fair (0.42-F)

Table 2: CAQI and Fuzzy description of Air Quality with DC in PCMC location by 4 AQ Experts'

From Table 2 it can be seen that the air quality computed using degree of certainty method is at par with the air quality description using the Conventional Air Quality Index (CAQI). We can say that air quality at *PCMC location is Very Poor with Very Good degree of certainty* .Variability in experts perception is also modelled using fuzzy logic. The four air quality experts are almost close in their opinion about describing air quality at Pimpri-Chinchwad monitoring location.

Linguistic Description	Degree of certainty (DC)
Very Good	0
Good	0.2 (Fair)
Fair	0.75 [Good]
Poor	0.1
Water Quality at PCMC is Fair with DC = 0.75, Good with DC = 0.2 & Poor with DC = 0.1	

Table 3: Fuzzy Description Of Water Quality With Degree Of Certainty

- It could be inferred from fuzzy modelling that the water quality at Intake well of Water Treatment Plant is *fair* with degree of certainty 0.75 (Good) and is tending towards *good* with degree of certainty 0.2 (fair). It is necessary to improve Pawana river water quality as *Very good with high degree of certainty*. The strength of fuzzy logic concepts lies in defining, in this case, river water quality at intake, straightway in linguistic terms with some degree of certainty attached to each linguistic terms [6].
- There is visible organic pollution due to domestic waste water from the nearby villages. A few industrial discharges are also located in about 2 kms of the intake well. Therefore, there will be increase in Bio -chemical oxygen demand resulting in high bacteriological load in Pawana river water.
- In case the pollution remains unchecked, the possibility of introducing pre chlorination at the intake resulting into possible trihalomethane (THM) formation can not be overruled. PCMC, therefore, must to undertake pollution abatement measures on a war footing. The team of experts will look into the issue and locate the sources of pollution for further action.

5. 0 CONCLUDING REMARKS

The two valued logic and probability theory is the basis of prospect theory. It is a deep-seated tradition in science to employ the conceptual structure of bivalent logic and probability theory as a basis for formulation of definitions and concepts.

What is widely unrecognized is that, in reality, most concepts are fuzzy rather than bivalent, and, in general, it is not possible to formulate a co-intensive definition of a fuzzy concept within the conceptual structure of bivalent logic and probability theory. Fuzzy logic via computing with words is a human centric logic which has been successfully used where the perceptions of the human or domain experts are of primary concern. This is more relevant in decision making wherein the decision makers have no or less numerical information on the governing parameters. It is well known that the expert knowledge base uses only linguistic terms. Why not compute with words and why always numbers?

Over the past few decades, soft computing tools such as fuzzy-logic-based methods, neural networks, and genetic algorithms have had significant and growing impacts. But we have seen only limited use of these methods in environmental fields, such as risk assessment, cost–benefit analysis, and life-cycle impact assessment. Because fuzzy methods offer both new opportunities and unforeseen problems relative to current methods, it is difficult to determine how much impact such methods will have on environmental policies in the coming decades.

The research study has conclusively demonstrated the utility soft computing techniques with focus on fuzzy logic in water and air quality management. It could be an eye opener for those who are engaged in improving urban water and air quality

For the types of complex and imprecise problems that arise in environmental policy, the ability to model complex behaviors as a collection of simple if–then rules makes fuzzy logic via computing with words an appropriate modeling tool. Would decision makers and the public accept expressions of water or air quality goals in linguistic terms with computed degrees of certainty? Resistance is likely. In many regions, such as the

United States and European Union (EU), both decision makers and members of the public seem more comfortable with the current system—in which government agencies avoid confronting uncertainties by setting guidelines that are crisp and often fail to communicate uncertainty. Perhaps someday a more comprehensive approach that includes exposure surveys, toxicological data, and epidemiological studies coupled with fuzzy modeling will go a long way toward resolving some of the conflict, divisiveness, and controversy in the current regulatory paradigm.

ACKNOWLEDGMENT
Assistance from Asheesh Sharma and Parvin Ladkat in providing necessary data is gratefully acknowledged.

REFERENCES
[1]. WHO World Health Organization Health Report, 2002, centre for disease control, Geneva, Switzerland.
[2]. Ersoy, K., Cho, S., 1992. An algorithm to compute the degree of match in fuzzy systems. Journal of Fuzzy Sets and Systems. Vol. 49(1992), 285- 299.
[3]. Deshpande, A. W., Raje, D.V., Khanna, P., 1996. Fuzzy Description of River Water Quality. International Conference-EUFIT Germany.
[4]. Sharma, A., Yadav, J., Madlik, P., Ladkat, P., Deshpande, A., 2012.
[5]. Fuzzy logic in Water Supply System Management – A Case Study. Paper presented in NAFIPS, Berkeley, 2012.
[6]. Zadeh, L.A.,1973. The concept of a linguistic variable and its application to approximate reasoning. ERL-M411, Berkeley, USA.
[7]. Mckone, T. E., Deshpande, A.W., 2010. Can Fuzzy Logic Bring Complex Environmental Problems into Focus? IEEE. 11-17.
[8]. Yadav, J. Y., Kharat V., Deshpande, A. W., 2013. Evidence Theory and Fuzzy Relational Calculus in Estimation of Health Effects due to Air Pollution. International Journal of Intelligent Systems. Vol. 22(1), 9-22.

Business Analysis and Decision Making Through Unsupervised Classification of Mixed Data Type of Attributes Through Genetic Algorithm

Rohit Rastogi[1], Saumya Agarwal[2], Palak Sharma[3], Uarvarshi Kaul[4] and **Shilpi Jain[5]**

Submitted in July, 2013; Accepted in February, 2014

Abstract - Grouping or unsupervised classification has variety of demands in which the major one is the capability of the chosen clustering approach to deal with scalability and to handle the mixed variety of data set. There are variety of data sets like categorical/nominal, ordinal, binary (symmetric or asymmetric), ratio and interval scaled variables. In the present scenario, latest approaches of unsupervised classification are Swarm Optimization based, Customer Segmentation based, Soft Computing methods like Fuzzy Based and GA based, Entropy Based methods and hierarchical approaches. These approaches have two serious bottlenecks…Either they are hybrid mathematical techniques or large computation demanding which increases their complexity and hence compromises with accuracy.

It is very easy to compare and analyze that unsupervised classification by Genetic Algorithm is feasible, suitable and efficient for high-dimensional data sets with mixed data values that are obtained from real life results, events and happenings.

Index Terms – Clustering, Clustering Algorithms, Categorical Dataset, Numerical Dataset, Data Mining, Pattern Discovery, Genetic Algorithm.

Core Idea of Our Paper

The proposed methodology deals with this problem with a new and efficiently generated Genetic Algorithm approach. It shapes into a better, lesser complex and computationally demanding pseudo codes which may in future lead into a revolutionary approach. We work upon multivariate data sets. In case of nominal variables, we calculate the Genetic Fitness Function criterion of each data item to decide their parent cluster and quantify the dataset by different methods to construct combined category quantifications and also plot the object scores. Here we propose an iterative procedure to calculate the cluster centers. To support our approach, a numerical experiment has been demonstrated. For all mixed variety of attributes Binary, Interval and Ratio scaled with ordinal type attributes, an efficient methodology by GA approach has been designed and has been suggested in this paper

[1, 2, 3, 4, 5]ABES Engineering College, Ghaziabad (U.P.), India
E-mail: [1]rohit.rastogi@abes.ac.in,
[2]mail4somya.ag@gmail.com,[3]sharma.palak595@ymail.com,
[4]saniakaul27@gmail.com and [5]shilpijain474@gmail.com

1.0 INTRODUCTION

The basic operation in Data Mining is partitioning a set of objects in database into homogeneous groups or clusters. It is useful in a number of tasks, such as unsupervised classification, image processing, sequence analysis, market research, pattern recognition, spatial analysis, economics etc. Clustering is a popular and widely used approach to implement the partitioning operation. It helps to partition objects-set into similar groups called as clusters in such a manner that objects in the same cluster are more similar to each other than objects in different clusters as per the same defined criteria.

Its major requirement is in the data mining concept to deal with high dimensionality for unsupervised classification distinct from other traditional applications of cluster analysis which may have thousands or millions of records with tens or hundreds of attributes. By this characteristic, many existing clustering algorithms stopped from being used in data mining applications. Data may be collected in the data mining sector which may contain all types of mixed attributes in real life practical applications. The traditional way to treat categorical, nominal, ratio-scaled or ordinal attributes as numeric with the help of dissimilarity matrices (after calculating Euclidean/Manhattan/Minkowaski distances and applying Normalization (standard deviation/ Z-score or min-max normalization on the results) and applying the related algorithms for numeric values, but the drawback of this process is that it may produce useless results sometimes because most of the categorical attributes are not ordered.

Most already available unsupervised classification algorithms either can handle both data types but are not efficient when clustering large data sets or can handle only the numeric attributes efficiently. Few algorithms can perform both well, such as k-prototypes and etc.

In order to handle large data sets with variety of mixed numeric and categorical and other data values, a new cost function may be defined for clustering by modification of the commonly used trace of the within cluster dispersion matrix. For minimizing the cost function (to get optimal solution), we introduce genetic algorithm (GA) in clustering process. Since GA uses search strategy globally and fits for implementing in parallel, the benefit of high search efficiency is achieved in GA based clustering process, which is very much favorable for unsupervised classification of large data sets.

The rest of the paper is organized as follows. Forthcoming Section covers some literature survey and concluding remarks over some contemporaries unsupervised classification algorithms. Then next section, gives some mathematical preliminaries of the algorithm. Then we discuss the Genetic-Algorithm briefly with modified and efficient cost function for all the data sets. Last section summarizes the discussion.

2.0 LITERATURE SURVEY

LI Jie et al.[1] The authors writers have depicted in the paper and proposed a novel clustering algorithm for the mixed data sets. They have successfully shown the modification of the common cost function, trace of the within cluster dispersion matrix.

- The results have demonstrated the effectiveness of the algorithms in discovering structures in data. The scalability tests have shown that the algorithm is efficient when clustering large complex datasets in terms of both the number of records and the number of clusters. These properties are very important to data mining.
- The emphasis of this paper is put on the issue that employs the genetic algorithm to solve the clustering problem.
- However, in using this algorithm to solve practical data mining problems, they still faced the common problem:
- How many clusters are in the data? This will be a challenging topic for further research.
- A novel clustering method with network structure based on clonal algorithm.
- By analyzing the neurons of the obtained network with minimal spanning tree, one can easily get the cluster number and the related classification information. The test results with various data sets illustrate that the novel algorithm achieves more effective performance on cluster analyzing the data set with mixed numeric values and categorical values.
- The distance measure of their clustering algorithm indicated that the smaller the distance measure is, the better the clustering partition. For this case, the clonal algorithm asks for a bigger affinity value. Hence, they have defined the Ab-Ag affinity function by using the dissimilarity measure.

$$f(x_j, p_{ig}) = \frac{1}{1+\sum_{l=1}^{t} |x_{jl}^r - p_{ig,l}^r|^2 + \lambda \cdot \sum_{l=t+1}^{m} \delta(x_{il}^c, p_{ig,l}^c)} \qquad i=1,2,\cdots c, \quad g=1,2,\cdots i_n$$

……..Eq. 2.1

The Ab-Ab affinity is defined as:

$$D_{ij} = \|p_{ig} - p_{jl}\| \qquad i,j=1,2,\cdots c \quad g=1,2\cdots i_n \quad l=1,2,\cdots j_n$$

Where, $\|\cdot\|$ is any a norm; $D=(D_{ij})_{N\times N}$ is antibody-antibody affinity matrix, and $N=\sum_{i=1}^{c} i_n$ is number of neurons of networks. **…….**

.Eq. 2.2

- Since the new algorithm combines the clonal selection algorithm and the forbidden clone operator, the obtained network has not only the specificity but also the tolerance of immunity. The experimental results illustrate that the novel algorithm can effectively explore the cluster structures of the data set.
- Moreover, it does not depend on the prototype initialization and the priori information of cluster number, which makes it as a real unsupervised learning.
- Very popular and effective methodology in this context is a CSA-based clustering algorithm for large data sets with mixed numeric and categorical values.
- They have successfully shown the modification of the common cost function, trace of the within cluster dispersion matrix. The clonal selection algorithm (CSA) can be used to optimize the new cost function. By the Experiments, results illustrate that the CSA-based new clustering algorithm is suitable for the large data sets with mixed variety of the numeric and categorical values.
- For clustering analysis on the large data set with mixed numeric and categorical attributes, the CSA-based algorithm not only has a high convergence speed, but also is independent on the initialization of the prototypes and can converge to the global optimum with the probability of 1.These properties are very important to the applications of data mining.
- In this research article, authors have tried to present a data mining algorithm based on supervised clustering to learn data patterns and use these patterns for data classification. By this approach, they have proposed a scalable incremental learning of patterns from data with both numeric and nominal variables. There are two different methods of combining numeric and nominal variables in calculating the distance between clusters are investigated.
- One method suggests, numeric and nominal variables are combined into an overall distance measures and now separate distance measures are calculated for numeric and nominal variables, respectively.
- Second method says, nominal variables are converted into numeric variables, and then a distance measure is calculated using all variables.
- The prediction accuracy and reliability of the algorithm were analyzed, tested, and compared with those of several other data mining algorithms and then analyzed the computational complexity, and thus, the scalability, of the algorithm, and tested its performance on a number of data sets from various application domains.
- CCAS is a supervised clustering and classification algorithm which has been extended a scalable, incremental,

and—into ECCAS that has the capacity of handling data with both numeric and nominal variables.

- Two different methods of handling mixed data types are developed. The two methods of ECCAS are tested and compared on a data set with mixed variable types for intrusion detection. Both methods produced comparable performance to that of the winning algorithm in a data mining contest on the same data set.
- The ECCAS algorithm and the distance measure could be used in common data mining applications. The authors have developed methods to adaptively and dynamically adjust the parameters during training, including the grid-interval configuration and the threshold-controlling outlier removal.
- ECCAS (A) was also tested on two other data sets for medical diagnosis and salary prediction applications with comparable performance to those of other data mining algorithms applied to these data sets. The performance on different data sets showed the reliability of ECCAS. The testing results for one data set also showed that the five phases of ECCAS reduces the impact of the data presentation order on the prediction accuracy. The number of grid intervals showed the impact on the prediction accuracy of ECCAS. In this study, they tested different numbers of grid intervals empirically.[1]

JiangXi et al. [2] the authors represented their research ideas.
Strength- K-Prototype is one of the important and effective clustering analysis algorithms to deal with mixed data types. This article discussed fuzzy clustering algorithm based on K-Prototype in detail and made improvements to solve its initial value problems. The proposed method is simple, easy to understand and can be achieved easily.

Technology Gist
In order to improve the randomness of selecting initial cluster centers and enhance the stability of algorithm results, one selection method is proposed as follows: assume it will be divided into k clustering, consider in two steps:
(1) Value- The normalized treatment of source data set makes the numerical data in interval [0, 1]. The purpose is to prevent too much weight paralleling numerical attribute with categorical attribute, and reduce differences of the dissimilarity between the two attributes. Select equal division interval point for numerical attribute: 1 / (k + 1), 2/(k + 1), 3 / (k + 1), ... , k / (k + 1) corresponding to k cluster centers of numerical attributes.

(2) Categorical Attributes- The aim is to divide the dataset into k portions. The data number in each portion is [n / k] and put the rest into the last division directly. Then, find the value with the highest frequency in each division. It is also the mode corresponding to the categorical attributes of K initial cluster centers. Finally, combine the two parts into k initial cluster center.

Advantages and Disadvantages with Other Approaches
The problem of selecting initial cluster center is improved by analyzing fuzzy K-Prototype. This algorithm solves the problem of clustering of mixed data, reduces iteration times in the clustering process and improves the quality and efficiency.

Limitations-
Further research need to be done, such as the existing local convergence to optimize the clustering result, distributed process about massive data and so on.

3.0 BETTERMENT BY THE USE OF GENETIC ALGORITHM

With the basic features of GA like encoding, crossover, mutation, appropriate fitness function and reproduction with survivor selection, the GA can be able to design better clustering and unsupervised classification operations.
The proposed approach can be described with experiments and their results. To test its clustering performance against other algorithms, the algorithm can be run on real-life datasets. At the same time, its properties are also empirically studied. One property from the above analysis is that our algorithm‘s computation complexity is determined by the component clustering algorithms. So far, many efficient clustering algorithms for large databases are available, to implicate that our algorithms will effective for large-scale data mining applications, too.

4.0 THE DEFINITION OF COST FUNCTION

Cost function is a function that determines the amount of residual error in a comparison and needs to be minimized in optimization experiment. Let X={$x_1, x_2, \ldots, x_n$} define a set of n objects and x_i= $\{x_{i1}, x_{12}, \ldots \ldots, x_{im}\}^T$ be an object and there are m attribute values. Let k be a positive integer. The objective of clustering **X** is to find a partition that divides objects in **X** into k disjoint clusters. A simple way to solve it is to choose a clustering criterion to guide the search for a partition. A clustering criterion is called cost function**.**
Let the no. of possible partitions of the definite but highly large objects be n. Here we have to investigate every partition in order to find a better one for a classification problem.

5.0 COST FUNCTION FOR NUMERIC DATA CLUSTERING

The widely used cost function is the trace of the within cluster dispersion matrix. One way to define the cost function is

$$C(W) = \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^2 (\mathrm{d}\,(x_j,\ x_i))^2 \ , w_{ij} \in \{0,1\} \quad \textbf{Eq. 5.1}$$

Here,w_{ij} is the membership degree of x_j belonging to cluster i.
W is a $k \times n$ order partition matrix. The function d(.) is a dissimilarity measure often defined as the
Euclidean distance. For the data set with real attributes, i.e., X$\subset R^m$, we have

$$d(x_j, x_i) = \left(\sum_{l=1}^{m} | x_{jl} - x_{il} | 2\right)^{1/2} \quad \textbf{Eq. 5.2}$$

Since, w_{ij} indicates x_j belonging to cluster i, and $w_{ij} \epsilon [0,1]$, we call **W** to be a hard k-partition.

6.0 COST FUNCTION FOR MIXED DATA CLUSTERING

6.1Max-Min Normalization for numeric data

For clustering the numeric data, first we will normalize numeric data so as to prevent the dominance of particular attribute. For which the normalization formula is as follows:-

$$n_i = \frac{x_i - min(x_i)}{max(x_i) - min(x_i)} \times (Rh - Rl) + Rl \quad \textbf{Eq. 6.1}$$

Where, x_i is the i-th object Rh and Rl are the higher and lower ranges respectively. N is the new normalized matrix containing all types of data.

6.2 Normalizing ratio-scaled values:-

First, we will take log of the ratio-scaled values, given as

$$f(n) = \log(n) \quad \textbf{Eq. 6.2}$$

6.3Normalizing ordinal values:-

First we assign ranks to the values as, better the value higher the rank and vice versa. Now, based on their ranks we will normalize them. Give 1 to the highest rank and 0 to the lowest one and other ranks get the value as:

$$\measuredangle(r) = \frac{1}{no.ofdifferentordinalvalues - 1} \times (r-1) \quad \textbf{Eq. 6.3}$$

6.4 Normalizing categorical values:-

If the two values match put value 1 and otherwise 0.

$$\delta(a, b) = \begin{cases} 0 & a = b \\ 1 & a \neq b \end{cases} \quad \textbf{Eq. 6.4}$$

6.5 Re-defining cost function:-

When N has attributes with numeric and mixed values, assuming that each object is denoted by=[, n_i

$$[n_{i1}^r, \ldots, n_{it}^r, n_{i,t+1}^c, \ldots, n_{im}^c, n_{i,m+1}^b, \ldots, n_{i,y}^b, n_{i,y+1}^o, \ldots, n_{iu}^o, n_{i,u+1}^{rs}, \ldots, n_{iz}^{rs}$$

],

the dissimilarity between two mixed-type objects n_i and n_j can be measured by the following **Eq. 6.5**

$$d(n_i, n_j) = [\left(\sum_{l=1}^{t} |n_{il}^r - n_{jl}^r|)2 + \measuredangle_1 . (\sum_{l=t+1}^{m} |n_{il}^c , n_{jl}^c|)2 + \measuredangle_2 . \left(\sum_{l=m+1}^{y} |n_{il}^b - n_{jl}^b \right|)2 + \measuredangle_3 . (\sum_{l=y+1}^{u} |n_{il}^o - n_{il}^o \left|\right|)2\right) + \measuredangle_4 . \sum_{l=y+1}^{u} |n_{il}^{rs} - n_{il}^{rs} |2)]^{1/2}$$

Where all the terms are squared Euclidean distance measure on the mixed attributes.

Using Eq. 6.5, for mixed-type objects, we can modify the cost function of Eq. 5.1, for mixed data clustering. In addition, to extend the hard k-partition to fuzzy situation, we further modify the cost function for fuzzy clustering as:

$$C(W) = \sum_{i=1}^{K} \left(\sum_{j=1}^{n} w_{ij}^2 \sum_{l=1}^{t} | x_{jl}^r - p_{il}^r |2 + \measuredangle_1 \sum_{j=1}^{n} w_{ij}^2 \sum_{l=t+1}^{m} |x_{jl}^c - p_{il}^c|2 + \measuredangle_2 . \sum_{j=1}^{n} w_{ij}^2 \sum_{l=m+1}^{y} |n_{il}^b - n_{jl}^b |2 + \measuredangle_3 . \sum_{j=1}^{n} w_{ij}^2 \sum_{l=y+1}^{u} |n_{il}^o - n_{il}^o |2\right) + \measuredangle_4 . \sum_{l=y+1}^{u} |n_{il}^{rs} - n_{il}^{rs} |2)$$

$, w_{ij} \epsilon [0,1]$ **Eq. 6. 6**

Let $C_i^r = \sum_{j=1}^{n} w_{ij}^2 \sum_{l=1}^{t} | x_{jl}^r - p_{il}^r | 2$

$$C_i^c = \measuredangle_1 \sum_{j=1}^{n} w_{ij}^2 \sum_{l=t+1}^{m} |x_{jl}^c - p_{il}^c |2 C_i^b$$

$$= \measuredangle_2 . \sum_{j=1}^{n} w_{ij}^2 \sum_{l=m+1}^{y} |n_{il}^b - n_{jl}^b |2 \quad C_i^o$$

$$= \measuredangle_3 . \sum_{j=1}^{n} w_{ij}^2 \sum_{l=y+1}^{u} |n_{il}^o - n_{il}^o |2$$

Eq. 6.7

We rewrite Eq.6.6 as:

$$C(W) = \sum_{i=1}^{k} (C_i^r + C_i^c + C_i^b + C_i^o) \qquad \textbf{Eq. 6.8}$$

7.0 GA-BASED CLUSTERING ALGORITHM FOR MIXED DATA

To obtain the optimal fuzzy clustering of the large data set with mixed values, genetic algorithms are employed to minimize the cost function. Since GA is a global search strategy in random fashion, it has high probability to achieve the global optima. Moreover, GA is fit for implementation in parallel, so GA-based clustering algorithm will be suitable for large data set.

8.0 GENETIC ALGORITHM

Genetic algorithm is a search strategy based on the mechanism of natural selection and group inheritance in the process of biology evolution. It simulates the cases of reproduction, mating and mutation in reproduction. GA looks each potential solution as an individual in a group (all possible solutions), and encodes each individual into a character string. By a pre-specified objective function, GA can evaluate each individual with a fitness value. In the beginning, GA generates a set of individuals randomly, then some genetic operations, such as crossover, mutation and etc., are used to perform on these individuals to produce a group of offspring. Since these new individuals inherit the merit of their parents, they must be better solution over their predecessors. In this way, the group of solution will evolve toward more optimal direction.

9.0 GA-BASED CLUSTERING ALGORITHM

9.1 Algorithm:

Step1. Begin

Step2. Define pop-size as desired population size

Step3.Randomly initializes pop-size population
Step4.While (Ideal best found or certain number of generations met)
O Evaluate fitness
O While (number of children=population size)
O Select parents
O Apply evolutionary operators to create children
O End while
Step5. End While
Step6. Return Best solution
Step7. End

To employ GA to solve the clustering, the following three problems should be settled first.
(1) How to encode the clustering solution into the gene string?
(2) How to design a reasonable fitness function for our clustering problem?
(3) How to select or design genetic operators including their parameters to guarantee fast convergence.

9.2 Encoding: From the cost function in Eq. 5.1 and Eq. 6.6, it is clear that the objective of clustering is to obtain a (fuzzy) partition matrix **W.** Then using the fitness function (stated below) we can improve the chances of a particular data point to be chosen. Then after selecting that particular cluster we can further subdivide the data points in the cluster, based on their fitness values.
Note that since we process data with mixed attributes, besides the numeric parameters, there are other mixed parameters in gene string. Due to this, it is not ordered for the binary attributes; they can be directly encoded rather than should be normalized first.

9.3 Fitness function: We are taking the fitness function such that fitness value is inversely proportional to the cost function value, i.e., the smaller the cost function is, the better the fuzzy clustering partition. For this case, the GA asks for a bigger fitness value. Hence, we define the fitness function by using the clustering cost function. Exponentially increased cost function will sharply reduce the fitness function.

$f(g) = \frac{1}{1+e^{C(W)}}$ **Eq. 9.1**

9.4 Genetic operators: Our GA-based clustering algorithm involves all the basic genetic operators, such as selection, reproduction, crossover and mutation. What we need to do is to specify the probability parameters for each operator. For the N individuals in a generation of population, we sort their fitness value in ascending order and label each individual with its order. The selection probability is specified as:

$P_s(g_{(i)}) = \frac{f(g_i)}{\sum_{i=1}^{n} f(g_i)}$ **Eq. 9.2**

The operation probabilities for crossover and mutation are adaptively assigned as Eq. **Eq. 9.3**

$$P_c(g_i, g_j) = \begin{cases} \frac{\alpha_1(f_{max} - f')}{f_{max} - \bar{f}} f' \geq \bar{f} \\ \alpha_2 otherwise \end{cases}$$

$$P_m(g_i) = \begin{cases} \frac{\alpha_3(f_{max} - f(g_i))}{f_{max} - \bar{f}} f(g_i) \geq \bar{f} \\ \alpha_4 otherwise \\ \textbf{Eq. 9.4} \end{cases}$$

where, $f_{max} = max_{i=1}^{N}\{f(g_i)\}$

$\bar{f} = \sum_{i=1}^{N} f(g_i), \; f' =, f(g_j)\}$,and$\alpha_i$€[0, 1]

Apart from above operators, we define a new operator for the clustering algorithm,
Gradient Operator. The changes in the existing weights are done as per the formula:
The gradient operator includes two steps iteration as:

$$w_{ij} = \sum_{i=1}^{k} \frac{(d(x_j,x_i))2}{(d(x_j,x_i))2}, i, j \quad \textbf{Eq. 9.5}$$

10. A PRACTICAL SAMPLE DATA TABLE OF MIXED VARIETY OF DATA TYPES:
We are representing the real life concept of our approach by taking data of 5 employess working in a company. Here we will use every kind of data (related to all data types) to show that our method works for every kind of data.In this example :

We are taking the weighted matrix (W_{ij}) as:

	1	2	3	4	5
1	0				
2	0.4	0			
3	0.2	0.2	0		
4	0.1	0.3	0.2	0	
5	0.5	0.2	0.1	0.4	0

Table 10.1: The Weighted Matrix W_{ij}

Test-1 contains salary of an employee (numeric data)
Test-2 shows whether the employee is male or female(binary data- Male=1/ Female=0)
Test-3 shows the department to which employee belongs (categorical data)
Test-4 depicts the ability of an employee (ordinal values) Exc.-Excellent, Fair or Good
Test-5 shows avg. credit points alloted according to their performance (ratio-scaled values)
Last Column shows the log value of ratio-scaled data type.

Object-id	Test-1	Test-2	Test-3	Test-4	Test-5	Log
1	25K	M	Code-A	Exc.	445	2.65
2	40K	F	Code-B	Fair	22	1.34
3	55K	M	Code-C	Good	164	2.21
4	27K	M	Code-A	Exc.	1210	3.08
5	53K	F	Code-B	Fair	38	1.58

Table 10.2: The Real Life Practical Dataset

Table 10.2 is converted into the normalized matrix using the above equations Eq. (6.1), Eq. (6.2), Eq. (6.3), and Eq. (6.4)

	1	2	3	4	5
1	0	0	0	0	0
2	0.5	1	1	1	1.31
3	1	0	1	0.5	0.44
4	0.0666	0	0	0	0.43
5	0.9333	1	1	1	1.07

Table 10.3: The Normalized Matrix of Table 10.2

We calculate the value of the expression (stated below) to be further used in Eq.6.6

$$\sum_{j=1}^{n} d(x_j, x_i)$$

	1	2	3	4	5
1	0				
2	4.9961	0			
3	2.4436	2.2569	0		
4	0.1893	3.962	2.1213	0	
5	5.0159	0.2453	1.6153	4.1607	0

Table 10.4: The Matrix calculated by Eq. 6.6

Now for Eq.6.6 we are calculating the value of the expression:

$$\sum_{j=1}^{n} w_{ij}^2 \left(d(x_j, x_i)\right) 2$$

	1	2	3	4	5
1	0				
2	0.0799	0			
3	0.0977	0.0903	0		
4	0.0018	0.3566	0.0848	0	
5	1.2539	0.0098	0.0165	0.6571	0

Table 10.5: The Matrix for expression of Eq. 6.6

Now we will calculate the expression:

$$\sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^2 \left(d(x_j, x_i)\right) 2$$

1	1.3455
2	0.5366
3	0.2013
4	1.1063
5	1.9373

Table 10.6: The Calculated Expression Value

Now using the Eq. 9.1, we find the fitness value of the above calculated values (above 5 tulles):

1	0.2066
2	0.3689
3	0.4498
4	0.2497
5	0.1259

Table 10.7: The Fitness Values of Calculated Values

First we arrange the above values in ascending order and label each one of them and then using Eq.9.2 calculate the selection probability $P_s(g_{(i)})$

1	0.1474
2	0.2633
3	0.3204
4	0.1782
5	0.0898

Table 10.8: The Selection Probability

11.0 ANALYSIS ON OUR EXPERIMENTAL RESULTS

From the above calculated tables, we can easily verify the dissimilarity matrices of our real life experimental data shown in tabular structure; we can comfortably decide the set of clusters based on the fitness values. We are taking the threshold value for our method to be 0.22. Data item 1 and 5 whose fitness value lie below the threshold value can be grouped together in the cluster and the other three tuples can be grouped in another.

Now these clusters can be improved using GA and using the selection probability.

Result:

So there can be two clusters:

C1:- data items 1 and 5.

C2:- data items 2,3 and 4

12.0 CONCLUSION AND FUTURE SCOPE

Here we have presented the genetic algorithm to cluster large data sets. The emphasis of this paper is put on the issue that employs the genetic algorithm to solve the clustering problem. Though the application is specific for the business, our approach is general purpose and could be used with a variety of mixed-type databases or spreadsheets with categorical, numeric and other data values, and temporal information. With improved metrics, artificial intelligence algorithms and decision analysis tools can yield more meaningful results and agents can make better decisions.

This approach, then, can ultimately lead to vastly improved decision-making and coordinating among business units and agents alike. If a class attribute is involved in the data, relevance analysis between the class attribute and the others (or feature selection) should be performed before training to ensure the quality of cluster analysis. Moreover, most variants of the GA use Euclidean-based distance metrics. It is interesting to investigate other possible metrics like the Manhattan distance or Cosine-correlation in the future. To faithfully preserve the topological structure of the mixed data on the trained map, we integrate distance hierarchy with GA for expressing the distance of categorical values reasonably.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their valuable suggestions. They would also like to thank HOD-CSE, Dr. R. RadhaKrishnan and HOD-IT, Prof. A.K. Sinha (Dean CRAP-ABES-Engineering College, Ghaziabad) for his involvement and valuable suggestions on soft-computing in the early stage of this paper.

REFERENCES

[1]. LIJie, GAO Xinbo, JIAO Li-cheng, "A GA-Based Clustering Algorithm for Large Data Sets with Mixed Numeric and Categorical Values", National Key Lab. of Radar Signal Processing, Xidian Univ., Xi'an 710071, China, 1998.

[2]. Zhou Caiying of Science & Technology Division and Jiang Xi , "The Improvement of Initial Point Selection Method for Fuzzy K-Prototype Clustering Algorithm", 2010 2nd International Conference on Education Technology and Computer (ICETC), University of Science and Technology ,GanZhou, China Zhoucaiying, Longjun Software of Software, Jiang Xi Normal University ,Nanchang, China,2010.

[3]. Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Dept. of Computer Science, the University of British 3. M. R. Anderberg, "Cluster Analysis for Applications", Academic Press, New York,1973.

[4]. B. Everitt, "Cluster Analysis", Heinemann Educational Books Ltd., 1974.

[5]. Zhexue Huang, Michael K.Ng., "A fuzzy *k*-modes algorithm for clustering categorical data", IEEE Trans. on Fuzzy Systems, 7(4): 446-452, August, 1999.

[6]. Zhexue Huang, "A fast clustering algorithm to cluster very large categorical data sets in data mining Columbia, Canada, pp.1-8.

[7]. R. Krovi, "Genetic Algorithm for Clustering",A Preliminary Investigation. IEEE press, Pp.504-544.

[8]. J. H. Holland, "Adoption in Natural and Artificial System", Ann Arbor, MI: Univ. Mich. Press, 1975.

Author's Profile

Mr. Rohit Rastogi received his B.E. degree in Computer Science and Engineering from C.C.S.Univ. Meerut in 2003, the M.E. degree in Computer Science from NITTTR-Chandigarh (National Institute of Technical Teachers Training and Research-affiliated to MHRD, Govt. of India), Punjab Univ. Chandigarh in 2010. He was Asst. Professor at IMS College, Ghaziabad in computer Sc. Dept. He has authored/co-authored, participated and presented research papers in various Science and Management areas in around 40 International Journals and International conferences including prestigious IEEE and Springer and 10 national conferences including SRM Univ., Amity Univ., JP Univ. and Bharti Vidyapeetha etc. He has guided 5 ME students in their thesis work and students of UG and PG in around 100 research papers. He has developed many commercial applications and projects and supervised around 30 B.E. students at graduation level projects.

At present, he is a Sr. Asst. Professor of CSE Dept. in ABES Engineering. College, Ghaziabad (U.P.-India), affiliated to Uttar Pradesh Tech. University, Lucknow.

His research interests include Data Ware Housing and Data Mining, Design Analysis of Algorithm, Theory of Computation & Formal Languages and Data Bases. At present, He is engaged in Clustering of Mixed Variety of Data and Attributes with real life application applied by Genetic Algorithm, Pattern Recognition and Artificial Intelligence. Also, he is preparing some interesting algorithms on Swarm Intelligence approaches like PSO, ACO, BCO etc.

Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Ad Hoc Networks

A. Chaudhary[1], V. N. Tiwari[2] and **A. Kumar[3]**

Submitted in September, 2013; Accepted in February, 2014

Abstract – Due to the advancement in wireless technologies, many of new paradigms have opened for communications. Among these technologies, mobile ad hoc networks play a prominent role for providing communication in many areas because of its independent nature of predefined infrastructure. But in terms of security, these networks are more vulnerable than the conventional networks because firewall and gateway based security mechanisms cannot be applied on it. That's why intrusion detection systems are used as keystone in these networks. Many number of intrusion detection systems have been discovered to handle the uncertain activity in mobile ad hoc networks. This paper emphasized on proposed fuzzy based intrusion detection systems in mobile ad hoc networks and presented their effectiveness to identify the intrusions. This paper also examines the drawbacks of fuzzy based intrusion detection systems and discussed the future directions in the field of intrusion detection for mobile ad hoc networks.

Index Terms – Detection Methods, Fuzzy Logic, Intrusion detection system (IDS), Intrusion Detection System Architectures, Mobile Ad Hoc Networks (MANETs), Security issues.

1.0 INTRODUCTION

Mobile ad hoc networks (MANETs) do not have any pre-existing infrastructure or administrative point as like conventional networks. In MANETs, mobile nodes can communicate freely to each other without the need of predefined infrastructure. This effectiveness and flexibility makes these types of networks attractive for many applications such as military operations, rescue operations, neighborhood area networks, education applications and virtual conferences. Mobile nodes play the role of host as well as routers and also support the multihop communication between the nodes. By the help of routing protocols, mobile nodes can send the data packets to each other in mobile ad hoc networks. Some characteristics of MANETs such as communication via wireless links, resource constraints (bandwidth and battery power), cooperativeness between the nodes and dynamic topology make it more vulnerable to attacks [1] [2]. Due to Manet's characteristics, Prevention based techniques such as authentication and encryption are not good solution for ad hoc networks to eliminate security threats because prevention based techniques cannot protect against mobile nodes which contain the private keys. So that Intrusion detection system is an essential part of security for MANETs. It is very effective for detecting the intrusions and usually used to complement for other security mechanism. That's why Intrusion detection system (IDS) is known as the second wall of defense for any survivable network security [3]. There are some groups which works together to enhance the functioning of mobile ad hoc networks (MANETs). IETF constituted the mobile ad hoc networks working group in 1997 [4].The rest of this paper is organized as follows: Section 2 presents the detailed introduction of Intrusion detection system. Section 3 describes the need of fuzzy based IDS on MANETs and Section 4 discusses and analyzes the proposed fuzzy based IDSs in MANETs from the literature. Section 5 discusses the drawbacks of proposed fuzzy based IDS and finally conclusion and direction for future research is outlined in section 6.

[1,3]Dept. of Computer Science & Engineering, Manipal University, Jaipur (India)-302026
[2]Dept. of Electronic & communication, Manipal University, Jaipur (India)-302026
E-Mail: [1]alka.chaudhary0207@gmail.com,
[2]vivekanand.tiwari@jaipur.manipal.edu and
[3]anil.kumar@jaipur.manipal.edu,

2.0 INTRUSION DETECTION SYSTEM

When any set of actions attempt to compromise with the security attributes such as confidentiality, repudiation, availability and integrity of resources then these actions are said to be the intrusions and detection of such intrusions is known as intrusion detection system (IDS) [5]. The basic functionality of IDS depends only on three main modules such as data collection, detection and response modules. The data collection module is responsible for collecting data from various data sources such as system audit data, network traffic data, etc. Detection module is responsible for analysis of collected data. While detecting intrusions if detection module detects any suspicious activity in the network then it initiates response by the response module. There are three main detection techniques presented in the literature such as misuse based, anomaly based and specification based techniques. The first technique, misuse-based detection systems such as IDIOT [6] and STAT [7] detect the intrusions on the behalf of predefined attack signature. The disadvantage of this technique is that it cannot detect new attacks but has low false positive rate so that it is generally used by the commercial purpose based IDSs. Second intrusion detection technique is anomaly-based detection technique e.g. IDES [8]. It detects the intrusion on bases of normal behaviour of the system. Defining the normal behavior of the system is a very challenging task because behavior of system can be changed time to time. This technique can detect the new or unknown attacks but with high false positive rates. The third technique is specification - based intrusion detection

[9]. In this detection method, first specified the set of constraints on a particular protocol or program and then detect the intrusions at run time violation of these specifications. The main problem with this technique is that it takes more time for defining the specification that's why it is a time consuming technique [10]. On the bases of the audit data, Intrusion detection system can be host based and network based. Host based IDS collect the audit data from operating system at a particular host and network based intrusion detection system collects audit data from host as well as trace the network traffic for any type of suspicious activity. Normally there are three basic types of IDS architecture in literature: Stand-alone intrusion detection systems - In this type of intrusion detection system architecture, an IDS run independently on each node in the network; Distributed and Cooperative intrusion detection systems - In this architecture all nodes have IDS agents so that each node can take part in intrusion detection locally and depend on cooperativeness between the nodes it can be made decision globally. This architecture dependent IDS are able to make two types of decision i.e. collaborative and independent. In collaborative decision, all nodes take part actively to make decision but in case of independent decision some particular nodes are responsible for making decision. Hierarchical Intrusion Detection Systems - This type of IDS architecture is an extended form of distributed and cooperative IDS architecture in which whole network is divided into clusters. Each cluster has clusterhead which has more responsibility than the other node members in the cluster [10] [11]. There are many number of IDSs have been proposed in MANETs. We will discuss fuzzy logic based proposed IDSs for MANETs in further sections.

3.0 NEED OF FUZZY BASED INTRUSION DETECTION SYSTEMS

Fuzzy logic is used in intrusion detection since 90's because it is able to deal with uncertainty and complexity which is derived from human reasoning [12]. By the help of fuzzy variables or linguistic terms, intrusion detection features can be viewed easily and decision of normal and abnormal activity in the network are based on its fuzziness nature that can identify the degree of maliciousness of a node instead of yes or no conditions [13] [14]. IF-then-else based fuzzy rules are used to define all situations in the network for identifying the attacks or intrusions. The fuzzy rule based system is known as fuzzy interference system (FIS) that is responsible to take decisions. Many types of fuzzy interference systems are proposed in the literature [15].

4.0 FUZZY BASED INTRUSION DETECTION SYSTEMS IN MANETs

Since, conventional based IDSs cannot be directly applied on MANETs. So due to this reason many authors have presented many IDSs for MANETs. This section is going to describe each category of fuzzy based IDSs which have been proposed in Literature.

1.1 Fuzzy Sets based Agent communication used for tactical MANETs IDS

Domian Walkins [16] proposed stationary intelligent fuzzy agents (SIFA) based IDS for detection of port scanning and distributed DoS attacks in tactical MANETs. Due to the dynamic topology of MANETs it is decided that SIFA resides in each node. For attack recognition, proposed SIFA is dependent on rule based processing system so that reasoning system accomplished with three steps: A knowledge-based, database of derived facts and an interference engine which is used in reasoning logic for processing the knowledge base. This paper used data set for recognition of distributed DoS and port scanning from directly tactical Manet environment. In the large scale Manet's environment, SIFA based IDS could provide overhead.

4. 2 Fuzzy Logic controller based IDS

Sujatha et al. [17] proposed a new fuzzy based response model (FBRM) for the detection of internal attacks in mobile ad hoc network which is depicted in figure 1. In the type of internal attack, they have considered false route request (FRR) attack due to this attack flooding, congestion, DoS attack, exhaustion of resources and exhaustion of bandwidth could happen at nodes in the MANETs. In this scheme Fuzzy logic controller monitors various feature such as route request rate, sequence number, Acknowledgement time and load pattern which can detect FFR attack. The architecture of FBRM is broadly classified into four steps: i) LIDS (local intrusion detection system) log file i.e. for collecting the information based on selected features from each node's local intrusion detection system and also from the neighbors nodes ii) analysis iii) evaluation and iv) response. The overall decision of network state is based on the level of Hacking (LOH) which calculated from sum sequence no., RREQ rate and acknowledge time. LRM (local response module) and GRM (global response module) is responsible for local and global responses.
In global response module, each node sends their response to its neighbor's nodes for global response.

4.3 Biologically Inspired type-2 fuzzy set recognition algorithm based IDS

Andrea and Hooman [18] suggested artificial immune system for detecting misbehaving nodes in Ad-Hoc wireless networks which is based on type-2 Fuzzy set. The purpose of this work is to detect and learn about misbehaviour nodes as well as protect the network without human interference. They assumed that the system is having the different states and any small portion may indicate misbehaviour. This paper used type-2 fuzzy set recognition algorithm for minimizing the uncertainties of some situation in the network where effective network parameter are not well defined for detecting misbehavior nodes, alarm threshold value for selected parameters are not clearly defined, system parameter could be negatively affected by background noise. This paper composed experts knowledge for making the difference between normal and abnormal behavior of selected parameters by the helper T-cells on the bases of person MF (membership function) approach. For reaching the final FOU (Foot print of uncertainty) they used interval type-2 fuzzy map (IT2FM) of each selected parameter

$$IT2FM\,(f_i) = \{(x, [\underline{\mu}_{f_i}(x), \overline{\mu}_{f_i}(x)]), x \in [0, 100]\}$$

Here x percent changes in the parameter f_i is indicated the uncertainty on the behalf of expert knowledge and some indications are used for presenting the changes the parameters such as red region for misbehavior of network parameter f_i, Yellow region for suspicious behavior and white region indicate the normal behavior. Helper T - Cells measure the actual changes of parameter f_i and find the closer region (red, yellow and white) of IT2FM. Once find the final decision then helper T-cells send the signal to Killer T-Cells for particular immune response. Actually, the proposed solution is totally based on the binding process of receptors and antigens. On the other hand, the proposed algorithm could moderate a static artificial immune system because all information of the parameters of the system should be available in advance. So that building the correct type-2 fuzzy map could be inefficient. That' why for future work, they will concentrate on the learning phase of the algorithm.

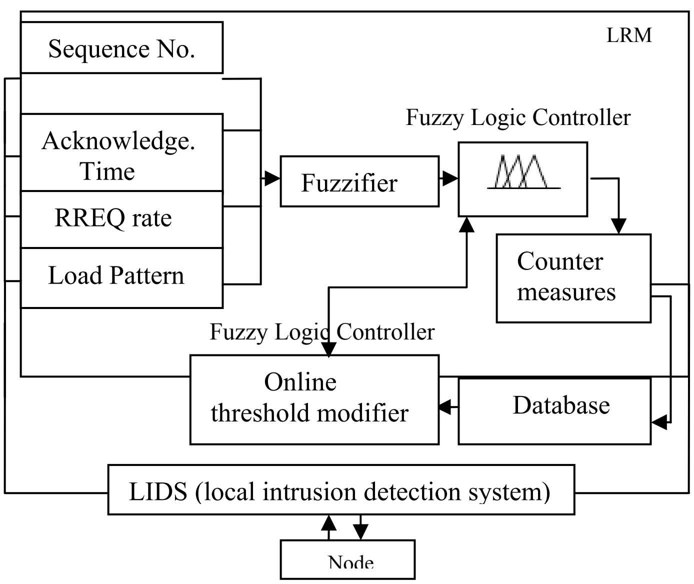


Figure 1: Proposed fuzzy controller based IDS [15]

4.4 Energy based trust solution using Fuzzy logic for MANETs IDS

Vijayan R et al. [19] suggested trust management scheme based on energy utilization using fuzzy logic for detection of selfish nodes in Manets. In the proposed scheme, every node monitors their one hop away neighbours for detection of any kind of malicious behavior with the help of some security components such as supervisor, aggregator, trust calculator and disseminator which is running on each node in the network. In these components, supervisor module is responsible for passively listening to the neighbor's communication with the help of passive acknowledgement (PACK) mechanism to check whether neighbours forwarded the packet or not. Aggregator module calculates the number of packets dropped and based on this each node trust level is determined. Third module trust calculator is calculated with the trust level by using percentage of packet dropped from the previous module. In this module fuzzy logic is used to calculate the trust level where percentage of packet dropped treated as fuzzy input variable. However, fuzzy trust calculator is based on direct trust agent, indirect trust agent and aggregator functionality where aggregator evaluated the total trust values. For total energy measurement at node to another node can be determined as follows:

$$E_{y/x} = P_{n>0}\,(P_{x=y}\,E_{Tack} + P_{x \neq y}\,E_{Rack}) + P_{m>0}\,(P_{x=y}\,E_{Tpck} + P_{x \neq y}\,E_{Rpck})$$

Where $E_{y/x}$ energy spent at node Y to node X, E_{Tack} and E_{Tpck} energy spent at transmit one acknowledgment and one data packet or E_{Tpck} and E_{Rack} energy spent at received one acknowledgment and one data packet. This defined equation and disseminator module is used to get the trust value in the case of mobility of the nodes in the network. They used network simulator NS-2 for carried out the simulation of proposed scheme in the network. At the time of calculation of trust level such factors i.e. link broken, battery exhaustion and replay packet generated are not considered so that it could degrade the accuracy level of proposed scheme.

4.5 Fuzzy Logic based IDS for MANETs

Kulbhushan and Jagpreet [20] proposed a fuzzy logic based IDS which can detect black hole attack on MANETs which is presented in figure 2. They formed the rule for detecting attack based on Mamdani fuzzy model and for drawing the membership function, input parameters such as forward packet ratio and average destination sequence number selected in each time slot. The output of derived rule is dependent on the fidelity level of each node which value is between 0 to 10 and threshold fidelity level chosen 5.5 for analyzing the level of node. If calculated fidelity level of node is less than or equal to fidelity threshold value then node is blackhole otherwise node is not blackhole. Ultimately fidelity level shows the level of node.

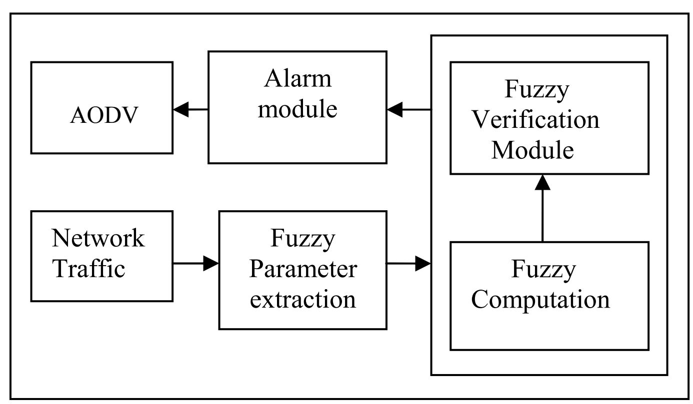


Figure 2: Proposed fuzzy based IDS [18]

This scheme is helpful for detecting blackhole attack but cannot detect new attack. In literature, there are other approaches also available for detecting blackhole attack using fuzzy logic such as M. Wahengbam et al. [21] suggested a fuzzy based IDS for MANETs which is capable to detect packet dropping attack such as Black hole and Gray hole attack. They considered that each node is having IDS and detect malicious activity locally for this purpose and assumed some threshold value for each node. In this proposed approach, each node maintains its packet list with the feature: sequence no., source node, destination node, packet type and expire time. During analysis, they calculated some indications on the bases of degree of symptoms, frequency of occurrence of symptoms and confirmed the presence of attack. Using NS-2 simulator, they have tested their scheme in two ways: when fuzzy logic is used for detection process and when fuzzy logic is not used. On the bases of analysis result, it proved that the fuzzy logic is more capable to find proposed attack accurately. This scheme chosen the threshold value for each node is very confusing job.

4.6 Trust and fuzzy logic based security framework for MANETs

Manoj V. et al. [22] presented a scheme based on certification authority (CA) and fuzzy logic for MANETs. Some central node is authorized by service provider for assigning the keys to source node which is going to request in the network called certification authority nodes and with the help of trust agent, direct and recommended trust values are obtained periodically. Direct and recommended trust values are calculated from direct observation of one hop away neighbors with the help of algorithms. A proposed fuzzy logic based analyzer used to calculate the trust value of a requested node (which is ready to data exchange between source and destination in the network) based on the computed fuzzy table. If requested node is trusted then it would get the certification otherwise not. Fuzzy logic based analyzer has total control on CA nodes. They have tested their approach on Qualnet simulator 5.0 with 6 and 12 no. of nodes. In this approach any one trusted node could be compromised with malicious node due to the communication via wireless link in MANETs.

4.7 Fuzzy based hybrid intrusion detection system for mobile ad hoc networks:

Vydeki et al. [23] used Fuzzy interference system (sugeno type-2) for detection of Black hole attack in Manet and proposed architecture is depicted in

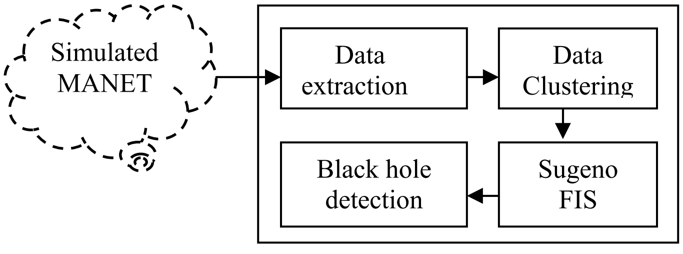


Figure 3: FIS based IDS [21]

They advised that selection of clustering algorithm in the process of FIS based IDSs play an important role so that it compared two well known clustering approaches such as subtractive and Fuzzy c-mean clustering. This proved that the detection rate based subtractive clustering (97%) is more efficient than the fuzzy c-means clustering (91%). This proposed approach only detects the black hole attack.

4.8 Forensic Analysis based on fuzzy Approach for IDS in MANET

Sarah and Nirkhi [24] introduced fuzzy logic based approach for forensic analysis to detect the distributed denial of service attacks (DDoS) in Manets. They suggested use of forensic analysis for intrusion detection because it is able to gather digital evidences from any system which has been compromised. It can reconstruct the compromised system and identify the location of attacker. This paper uses fuzzy Logic approach to forensic analysis based on dynamic source routing (DSR) protocol. Three steps are followed to get the result as a forensic report: first capture the log files then analyzing log files using fuzzy logic and at last presenting the conclusion in terms of forensic report. However, in this paper no simulation and experimental results based on forensic analysis are given.

4.9 Mamdani and Sugeno Fuzzy Inference Systems based IDSs in MANETs:

Alka C. et al. [25] [26] [27] proposed mamdani and sugeno fuzzy inference systems based IDSs for packet dropping attack (PDA) and sleep deprivation attack (SDA) in MANETs. The simulation results are proved that the proposed systems are able to detect the PDA and SDA attacks very efficiently in MANETs.

5.0 DRAWBACKS IN PREVIOUS PROPOSED FUZZY BASED INTRUSION DETECTION SYSTEMS

The proposed fuzzy based IDSs for detection of intrusions in MANETs are not able to cope up all type of attacks. One of few proposed IDSs can cope attacks [18] but it is also having some limitations. We have analyzed that all proposed fuzzy based IDSs are considered very limited features or attributes for data collection which is specific for a particular attack. So that these IDSs are only detect the particular attack in MANETs. In IDS Architecture point of view, due to the complex properties of mobile ad hoc networks are required distributed and cooperated architecture but some of proposed IDSs are concentrated only distributed architecture that's why these IDSs only detect the attacks locally. In case of local detection, each node are only responsible for raise alarm when it detects intrusion locally or not shared it to other nodes in the network for global detection. In terms of detection techniques, as per Table 1 presented that the most of proposed fuzzy based intrusion detection systems use misuse detection techniques and very few fuzzy based IDSs use anomaly and specification based detection techniques. However misused detection technique is responsible for detecting limited attacks i.e. membership function in fuzzy based approaches are defined for only specific attack so that these fuzzy based detection approaches cannot be detect new malicious activity or attacks that's why selection of detection techniques should be anomaly based or hybrid. Table 1 summarizes all fuzzy based IDSs in MANETs.

6.0 CONCLUSION AND FUTURE SCOPE

In this paper, we have analyzed fuzzy based intrusion detection systems which have been proposed in literature for Manets. We have analyzed the working style of proposed fuzzy based IDSs and reached on decision that still we do not have any promising solution for this dynamic environment because most of Proposed fuzzy based IDSs emphasized on very limited features for data collection towards detection of very specific range of attacks. Hence, MANETs are required for more concentration of researchers. It can be a fastest growing area for future research in terms of detection techniques, response mechanism and selection of node features for data collection. In future, we are concentrating to develop a new intrusion detection system that can be used to classify the normal and malicious activities in the network.

REFERENCES

[1]. Y. Li and J. Wei., “Guidelines on selecting intrusion detection methods in MANET”, In Proceedings of the Information Systems Educators Conference, 2004.

[2]. A. Hasti, “Study of Impact of Mobile Ad – Hoc Networking and its Future Applications”, BIJIT – 2012; January - June, 2012; Vol. 4 No. 1; ISSN 0973 – 5658.

[3]. Y. Zhang and W. Lee., “ Intrusion detection in wireless ad hoc networks” , In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom'00), pages 275-283, 2000.

[4]. IETF Mobile Ad-Hoc Networks Working Group (MANET), IETF web-sitewww.ietf.org/dyn/wg/charter/manet-charter.html.

[5]. R. Heady, G. Luger, A. Maccabe, and M. Servilla, “The architecture of a network level intrusion detection system” Technical report, Computer Science Department, University of New Mexico, August 1990.

[6]. S. Kumar and E. H. Spafford, “A software architecture to support misuse intrusion detection” In Proceedings of the 18th national Information Security Conference, pages 194- 204, 1995.

[7]. K. Ilgun, R. A. Kemmerer, and P.A. Porras, “State transition Analysis: A rule- based intrusion detection approach”, IEEE Transactions on software Engineering, Vol. 21 No. 3:181-199, March 1995.

[8]. T.Lunt, A. Tamaru, F. Gilham, R. Jagannathan, P.Neumann, H. Javitz, A. Valdes, and T.Garvey, “A real- time intrusion detection expert system (IDES) – final technical report”, Technical report, Computer Science Laboratory, SRI International, Menlo Park, Clifornia, February, 1992.

[9]. Uppuluri P, Sekar R, “Experiences with Specification-based Intrusion Detection”, In Proc of the 4th Int Symp on Recent Adv in Intrusion Detection , pp. 172-189. 2001.

[10]. S. Sen, J.A. Clark - Guide to Wireless Ad Hoc Networks; In: Chapter 17-Intrusion Detection in Mobile Ad Hoc Networks-Springer, 2008.

[11]. P. Brutch and C. Ko, “Challenges in Intrusion Detection for Wireless Ad-hoc Networks," In Proceedings of 2003 Symposium on Applications and the Internet Workshop, pp. 368-373, January 2003.

[12]. B. Shanmugam and N. B. Idris, “Anomaly Intrusion Detection based on Fuzzy Logic and Data Mining”, In Proceedings of the Postgraduate Annual Research Seminar, Malaysia 2006.

[13]. M. Wahengbam and N. Marchang, “Intrusion detection in manet using fuzzy logic”, 3rd IEEE National Conference on Emerging Trends and Applications in Computer Science (NCETACS), ISBN: 978-1-4577-0749-0, pp. 189 – 192, Shillong, 30-31 March 2012.

[14]. Verma, A. K., R. Anil, and Om Prakash Jain. "Fuzzy Logic Based Revised Defect Rating for Software Lifecycle Performance Prediction Using GMR."Bharati Vidyapeeth’s Institute of Computer Applications and Management, 2009.

[15]. J. S. R. Jang, C. T. Sun and E. Mizutani – Neuro-Fuzzy and Soft Computing - A computational Approach to Learning and Machine Intelligence; First Edition; Prentice Hall of India, 1997.

[16]. Watkins, Damian. "Tactical manet attack detection based on fuzzy sets using agent communication." In 24th Army Science Conference, Orlando, FL, 2005.

[17]. S. Sujatha, P. Vivekanandan, A. Kannan, “Fuzzy logic controller based intrusion handling system for mobile ad hoc networks”, Asian Journal of Information Technology, ISSN: 1682- 3915, pp.175-182, 2008.

[18]. A. Visconti, H. Tahayori, “ A Biologically – Inspired type-2 fuzzy set based algorithm for detecting misbehaving nodes in ad hoc networks” , International Journal for Infonomics, Vol.3, No.2, pp. 270-277, June 2010.

[19]. R. Vijayan, V. Mareeswari and K. Ramakrishna, “Energy based trust solution for detecting selfish nodes in manet using fuzzy logic”, International Journal of research and reviews in computer science , Vo. 2, No. 3, pp. 647-652, June 2011.

[20]. Kulbhushan and Jagpreet Singh, “Fuzzy logic based intrusion detection system against blackhole attack AODV in manet”, IJCA Special issue on “Network Security and Cryptography” Vol. NSC, No. 2 pp. 28-35, December, 2011.

[21]. M. Wahengbam and N. Marchang, “Intrusion detection in manet using fuzzy logic”, 3rd IEEE National Conference on Emerging Trends and Applications in Computer Science (NCETACS), ISBN: 978-1-4577-0749-0, pp. 189 – 192, Shillong, 30-31 March 2012.

[22]. V. Manoj, M. Aaqib, N. Raghavendiran and R. Vijayan “A Novel security framework using trust and fuzzy logic in manet” , International Journal of Distributed and

Parallel Systems , Vol. 3, No. 1, pp. 285-298, January,2012
[23]. D. Vydeki and R.S. Bhuvaneswaran, "Effect of clustering in designing a fuzzy based hybrid intrusion detection system for mobile ad hoc networks", Journal of Computer Science, Vol. 9, No. 4, pp. 521-525, ISSN: 1549 - 3636, 2013.
[24]. S. Ahmed & S.M. Nirkhi, "A Fuzzy approach for forensic analysis of DDoS attack in manet" International Conference on Computer Science and Information Technology, ISBN: 978-93-82208-70-9, Hyderabad, 10th March 2013.
[25]. Chaudhary, A., Kumar, A., & Tiwari, V. N. (2014, February), " A reliable solution against Packet dropping attack due to malicious nodes using fuzzy Logic in MANETs", In Optimization, Reliability, and Information Technology (ICROIT), 2014 International Conference on (pp. 178-181), IEEE.
[26]. Chaudhary, A., Tiwari, V. N., & Kumar, A. (2014, February), "Design an anomaly based fuzzy intrusion detection system for packet dropping attack in mobile ad hoc network", In Advance Computing Conference (IACC), 2014 IEEE International (pp. 256-261), IEEE.
[27]. Chaudhary, A., Tiwari, V. N., & Kumar, A. (2014, February), "Design an Anomaly Based Novel Approach for Detection of Sleep Deprivation Attack in Mobile Ad hoc networks Using Soft Computing", Proceedings of 3rd International Conference on Recent Trends in Engineering & Technology (ICRTET'2014), Elsevier.

IDS	Data Source	IDS Architectures	Detection Techniques	Routing Protocol	Addressed attack type	Decision Making	Response Mechanism	Simulator & Toolbox
IDS using Fuzzy Sets based Agent communication [16]	Collect packet data from data stream	Distributed & cooperative	Misuse based detection	not specified	Distributed denial of service attacks and port scanning attacks	Independent & collaborative	Alarm	SIFA Application
Fuzzy Logic Controller based IDS [17]	LIDS audit log file and neighbors related data	Distributed & cooperative	Misuse based detection	AODV	False route request attack	collaborative	Fuzzy based response model on attacked system	NS-2 and fuzzy logic controller toolbox of MATLAB 6.1
Artificial Immune System based on Type-2 Fuzzy Sets for Manets IDS [18]	Collect sample data of various network parameters	Distributed & cooperative	Partial-Anomaly based detection	not specified	Misbehaving Nodes	collaborative	Active immune based response on attacked system	No detail
Energy based trust solution using fuzzy logic for IDS[19]	network packet level data	Distributed	Anomaly based Detection	DSR	Selfish nodes	Independent	No detail	NS-2

IDS	Data Source	IDS Architectures	Detection Techniques	Routing Protocol	Addressed attack type	Decision Making	Response Mechanism	Simulator & Toolbox
Fuzzy logic based IDS [20]	Network traffic related feature	Distributed	Misuse based detection	AODV	Blackhole Attack	Independent	Alarm	NS-2
IDS using Fuzzy Logic[21]	Packet related feature	Distributed	Misuse based detection	AODV	Blackhole Attack, Gray hole Attack	Independent	Active response	NS-2
Trust and fuzzy logic based IDS[22]	Network packet data	Distributed & cooperative	Cryptographic algorithms and trust based	AODV	Malicious node	collaborative	alarm	Qualnet 5.0
Fuzzy inference system based IDS[23]	Data packets and control packet based features	Distributed	Specification and anomaly based detection	AODV	Blackhole attack	Independent	Active response based on FIS system output	NS-2 and MATLAB Function 'genfis'
IDS using Forensic analysis based on fuzzy logic approach[24]	Data packets and routing packets	Distributed	Misuse based detection	DSR	Distributed denial of service attacks	Independent	not specified	_
Mamdani and Sugeno based IDSs [25][26][27]	Packet based and mobility based data	Distributed Architecture	Misuse based And anomaly based	AODV	Packet dropping attack and sleep deprivation attack	Independent	alarm	Qualnet Simulator 6.1

Table 1: Summarization of All Reviewed Fuzzy Based IDSs

Fuzzy Classification System by Self Generated Membership Function Using Clustering Technique

Shruti S. Jamsandekar[1] and **Ravindra R. Mudholkar**[2]

Submitted in August, 2013; Accepted in February, 2014

Abstract - In this paper we have proposed fuzzy classification system by generating the membership functions using semi supervised learning method. k-means clustering technique is used to form clusters and to obtain membership centers, each cluster boundary values. These boundary values are approximated to vertex values of membership functions for overlapping the membership functions. The work is extended with auto creation of fuzzy inference system for classification to perform the classification on the given input data.

Index Terms – fuzzy classification, clustering, k means, membership function

NOMENCLATURE
FIS – Fuzzy Inference System

1.0 INTRODUCTION

Fuzzy classification is one of the applications of fuzzy logic which is used to deal with classification problems. In development of a fuzzy classification system, the important task is to construct membership functions and to find a set of suitable fuzzy rules in the fuzzy classification system. Fuzzy membership functions and fuzzy rules can be formulated based on expert knowledge approach and other alternative is use data driven approach, these approaches correspond to either manual or automatically through a machine learning process based on training instances respectively, Mostly expert knowledge is used to formulate membership function range and if then rules for inference and it has an advantage as it has link with domain knowledge but it can be very subjective with different experts generating different membership functions and rules for same application. The other approach of generation of fuzzy membership functions based on the input data will convert crisp data into linguistic terms. Most previous attempts in automatic generation of fuzzy membership functions have used machine learning and statistical methods [16, 3, 4, 7, 11]. In this paper, we present a new method for constructing membership functions using k means clustering technique, based on cluster centers and approximation of boundary values of each cluster obtained. The paper has 2 major contributions, first approach data driven membership function generation by clustering algorithm and second contribution is automatic creation of *if then* fuzzy classification system to deal with student semester examination data. The automatic creation of the fuzzy classification system from the membership function generated reduces the hassle of every time creating a new fuzzy Inference system to perform classification on change of data input.The rest of this paper is organized as follows. In Section 2 work in related field is reviewed. In section 3, we briefly review basic concepts of fuzzy sets. The proposed method to construct membership functions based on clustering is explained in section 4. The details and steps involved in automatic creation of fuzzy inference system for classification are in section 5 In Section 6 we discuss the experimental results of the proposed method. The conclusions and future work are discussed in Section7.

2.0 RELATED WORK

In [1] the authors have proposed a fuzzy model for calculating the group maturity rating for different groups within a software organization based on the defect density, residual defect density and the review effectiveness of the historical projects. In [2] authors have discussed fuzzy multi-objective optimization model approach for selecting the optimal COTS software product among alternatives for each module in the development of modular software system they have used fuzzy methodology for the estimation of reliability and cost. In[3] authors discuss an improvement in performance in Fuzzy Rule Based Classification Systems by using interval valued fuzzy set and cooperative tuning of methodology .Oyelade, O. J et.al [15] proposed a method using Kmeans clustering algorithm for analyzing students' results based on cluster analysis. In [7] Sushmita Mitra et.al discussed a new methodology for encoding connection weights of Fuzzy MLP and a technique generating an architecture of the fuzzy MLP in terms of hidden nodes and links.

M.Holena [5] proposed an algorithm to extract rules for fuzzy system using MLP with continuous activation function. In [16] the authors discussed framework for generation of fuzzy membership functions from training data using genetic algorithm. In [17] authors propose a method to generate rules and learning for knowledge base using genetic algorithm.

Roberto R.F. Mendes et.al [4] proposes a system for fuzzy classification by using genetic algorithm to generate fuzzy rules and an evolutionary algorithm to construct membership functions.

[1]*Department of Computer Studies, SIBER, Kolhapur. (MS), India-416004. shrutijams@gmail.com*
[2]*Department of Electronics Shivaji University, Kolhapur, (MS), India-416004.*
E-mail: [1]shrutijams@gmail.com and
[2]*rrm_eln@unishivaji.ac.in*

[6] Proposed a clustering technique by using genetic algorithm which searches the cluster center in the cluster space and optimizing the similarity metrics. Authors in [8] proposed a design for fuzzy classification system based on labeled data using genetic algorithm, they have also reduced the rule set for the system. In [9] authors have discussed different clustering algorithms and used fuzzy logic based clustering in controlling multi compressor system. [10] M.A. Hogo presented an evaluation method of E-learners using Fuzzy C-means and Kernelized Fuzzy C-Means to predict the student profile and group them into different learning categories. Author in [11] have experimented apriori association rule mining using Weka tool to predict the student performance and find the correlation among attributes.

In [12] authors have proposed a new method of constructing fuzzy membership functions using α-cuts of fuzzy sets from training data set. *C Vialardi et.al in* [13] Proposed a recommendation system in higher education using data mining techniques, for student to choose appropriate course enrollment depending upon their academic performance. Hisao Ishibuch et.al [14] proposed an approach for automatic construction of fuzzy classification system using minimize set of rules generated by from genetic algorithm

3.0 INTRODUCTION TO FUZZY LOGIC

A fuzzy set A in a universe of discourse X is defined as the following set pairs [18]

$$A = \{\mu_A(x) : x \in X\} \qquad (1)$$

Where, $\mu_A(x)$: $X \rightarrow [0, 1]$ is a mapping called the membership function of fuzzy set *A* and $\mu_A(x)$ is called the degree of belongingness or membership value or degree of membership of $x \in X$ in the fuzzy set *A*. we write (1) in the following form:

$$A = \{\mu_A(x)/x : x \in X\} \qquad (2)$$

For brevity, however, we often equate fuzzy sets with their membership functions i.e. instead of a fuzzy set *A* characterized by $\mu_A(x)$ we will often say fuzzy sets μ_A [4, 8].

Example**:** Suppose $X = \{6, 2, 0, 4\}$. A fuzzy set of *X* may be given by $A = \{0.2/6, 1/2, 0.8/0, 0.1/4\}$.

Construction of membership function is based on the system design data and choice of the suitable shape. There are many shapes of membership functions. However, the application context dictates the choice of the suitable shape.

For the problem domain addressed in this study, system components have maximum and minimum value that cannot be exceeded. Therefore, any candidate membership function shape should have two extreme bounds with zero and hundred as range values. Triangular and trapezoidal shapes are the simplest MF shapes that meet this requirement.

The membership function of the triangular fuzzy set A can be represented by a triple (b; c; a), where "c" is called the center of the triangular fuzzy set A; "b" is called the left vertex of the triangular fuzzy set A; "a" is called the right vertex of the triangular fuzzy set A.

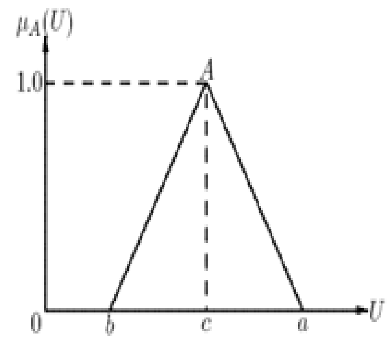


Figure 1: Triangular Membership Function

$$triangular(x; a, b, c) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0 & x \ge c \end{cases} \qquad (3)$$

The membership function of the trapezoidal fuzzy set can be represented by a function of a vector, x, and depends on four scalar parameters a, b, c, and d, as given by

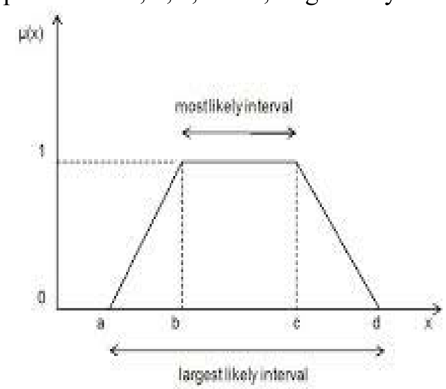


Figure 2: Trapezoidal Membership Function

$$Trapezoidal(x; a, b, c, d) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & b \le x \le c \\ \dfrac{d-x}{d-c} & c \le x \le d \\ 0 & d \le x \end{cases} \qquad (4)$$

4.0 PROPOSED METHOD TO GENERATE MEMBERSHIP FUNCTIONS BASED ON K MEANS CLUSTERING TECHNIQUE

In the proposed method we have created a Fuzzy Inference System for two input variables and one output variable. A combination of two membership functions triangular and trapezoidal is used. The figure of triangular is shown in Figure 1and that of trapezoidal is in Figure 2. For input as well as output variables of the system the extreme membership functions are trapezoidal and membership function in between are triangular. In this paper we have used KMeans clustering algorithm to form clusters and constructed the triangular membership functions. Taking the cluster center as center c of triangular set referring to Figure 1. The extreme trapezoidal fuzzy sets are then approximated .

4.1 Membership Functions for Input Variables Created as Follows

I. Kmeans Clustering algorithm is used to form clusters on student examination data for each subject marks to form 3 cluster. The three cluster center found forms the center of 3 inner triangular fuzzy membership functions

II. The two vertexes of each of the triangular fuzzy membership function referring to Figure 1. b and a are calculated by calculating the maximum and minimum value of each cluster

III. The maximum value obtained for each cluster is increased by 10% which formed vertex b of each of the triangular membership function (Figure 1)

IV. The minimum value obtained for each cluster is decreased by 10% which formed the vertex a of each of the triangular membership function(Figure 1)

V. The trapezoidal membership functions for the input variables are constructed as follows

a) Left extreme trapezoidal membership function according to figure 2 a = 0, b = 0 and c value is calculated by 10 % difference of Minimum of first triangular membership function cluster with itself and d is calculated by 15 % difference of Minimum of first triangular membership function cluster with itself

b) Right extreme trapezoidal membership function according to figure 2 c = 100, d = 100 value and a is calculated as 10% difference of maximum obtained for third triangular membership cluster with itself and b is calculated as 5 % difference of maximum obtained for third triangular membership cluster with itself

4.2 Membership Functions for Output Variables Created as Follows

I. For output variable the inner triangular membership function is calculated as follows

II. Input variable corresponding first triangular membership function centers average forms the center of first triangular membership of output variable and same procedure is repeated for second and third triangular membership function centers

III. The two vertexes of 3 triangular membership for output variable referring to Figure 1

a) a vertex of each of the 3 triangular membership function is Average of minimum of first cluster, second cluster and third cluster for input variable(calculated for input in 4.1) respectively

b) b vertex of each of the 3 triangular membership function is Average of maximum of first cluster, second cluster and third cluster for input variable calculated for input in 4.1) respectively.

IV. The trapezoidal membership functions for the output variables are constructed as follows

a) Left extreme trapezoidal membership function according to figure 2 a, b with zero value and c is 20 and d is 35. Here 35 is taken since the domain data is student examination marks and passing is 35 out of 100

b) Right extreme trapezoidal membership function according to figure 2 c, d is assigned value 100, a is calculated as 10% difference of maximum obtained for third triangular membership cluster with itself and b is calculated as 5 % difference of maximum obtained for third triangular membership cluster with itself

5.0 AUTOMATIC CREATION OF FUZZY INFERENCE SYSTEM (FIS) FOR CLASSIFICATION

Matlab fuzzy tool was used to create the proposed fuzzy system. The fuzzy inference system is mostly constructed manually; instead in this experiment we tried the approach of automatic creation of Fuzzy inference System (FIS). In this paper we generated FIS given the input data for classification. The generation of FIS is enumerated in steps below.

5.1. Create a MATLAB® workspace variable FIS name.

FIS = NEWFIS (name of FIS, type of FIS, andMethod, or Method, implementation method, aggregation method, defuzzication method)

Example:
a= newfis('clustclassFIS','mamdani','min','max', 'min','max','centroid');

5.2. Create input variable and add it to FIS .

Addvar method adds an a variable to an FIS.[19]
The parameters of the function are as follows
First – FIS name
Second – the type of variable input or output
Third – the name of the variable
Fourth – the vector with range values of the variable

a=addvar(a,'input','sub1_marks',[0 100]);

5.3 Create membership function add it to FIS

ADDMF method adds a new membership function to an FIS
a=addmf(a,varType,varIndex,mfName, mfType,mfParams)[19]

The parameters of the function are in this order:
First - a matlab variable name of a FIS structure in the workspace
Second- the type of variable to be add the membership function (input or output)
Third – the index value of the membership function
Fourth- string representing the label of the membership function eg. Low, medium etc.
Fifth - A string representing the type of membership function eg. Triangular, trapezoidal etc.
Sixth- the vector specifying the membership function range values

In this experiment the vector of parameters passed is as per membership range calculated in section 4.1

5.4 Similarly create and add all input variables and output variable with their membership functions to the FIS.

In this experiment the vector of parameters passed to the output variable membership function is as per membership range calculated in section 4.2

5.5 Form the Rule base with rules defined and add it to FIS (to carry out inference mechanism)

a) Create rule list matrix for the FIS

The format of the rule list matrix is as follows. If there are m inputs to a system and *n* outputs, there must be exactly $m + n + 2$ columns to the rule list matrix.

The first *m* columns refer to the inputs of the FIS system. The next *n* columns refer to the outputs of the FIS system. Each column for inputs and outputs contains a number that refers to the index of the membership function for that variable. The $m + n + 1$ column contains the weight that is to be applied to the rule. It can ale value between o and 1.The $m + n + 2$ column contains the fuzzy operator for the rule's antecedent to be used, the value used is 1 if AND fuzzy operator and value is 2 if OR fuzzy operator.[19]

b) addrule(a,ruleList)

addrule method is used to add rule list to the FIS
It has two parameters as follows .
First- is the MATLAB workspace variable FIS name.
Second- is a matrix which represents a given rule.
Example

```
ruleList=[   1   3   2   1   1
             1   4   3   1   1
             5   2   2   1   1];

a = addrule(a,ruleList);
```

If the system a has two inputs and one output, the first rule can be interpreted as:"If Input 1 is MF 1 and Input 2 is MF 1, then Output 1 is MF 1."

6.0 EXPERIMENTAL RESULTS

Comparison of the experimental results of the proposed method with the existing methods
To illustrate the method proposed

i. Accept the student examination data. Student data set for two subject marks in semester exam is shown in , Appendix A
ii. Clustering the student data using kmeans algorithm
a) The cluster center for subject 1 data are center of three triangular membership functions of input variable 1

55.26 39.50 71.66

b) The cluster center for subject 2 data are center of three triangular membership functions of input variable 2
60.95 4 8.47 28.66

iii.. Calculate the left and right vertex points of the 3 triangular membership functions
a) Compute maximum and minimum for subject 1 and subject 2 which is input 1 and input 2 respectively.

Input 1 Cluster number	**Maximum value**	**Minimum value**
Cluster_ 1	46	29
Cluster_2	63	48
Cluster_3	81	65

Table 1: Max and Min of input 1

Input 1 Cluster number	Maximum value	Maximum value
Cluster_ 1	81	56
Cluster_2	53	43
Cluster_3	36	20

Table 2: Max and Min of input 2

b) Approximate values calculated for overlapping of membership of triangular membership functions of input_1

Left and right vertex of the 3 triangular membership function for input 1.

Membership Function	Right Vertex	Left Vertex
low	39.6	18
medium	58.3	38.7
high	89.1	50.4

Table 3: Left and Right vertex points of input 1

Left and right vertex of the 3 triangular membership function for input 2.

Membership Function	Right Vertex	Left Vertex
low	50.6	26.1
medium	69.3	43.2
high	89.1	58.5

Table 4: Left and Right vertex point of input 2

The following is the membership function generated of input1, input2 variables for subject1 and subject 2 respectively using steps of section 3.1 and section 4.

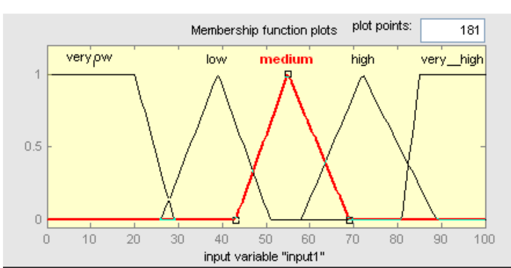


Figure 3: Generated membership function plot for input 1

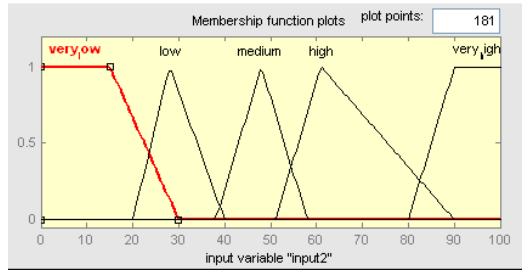


Figure 4: Generated membership function plot for input 2

The following is the membership function generated of output variables of fuzzy inference system for The following is the membership function generated of output variables of fuzzy inference system for classification using steps of section 3.1 and section 4.

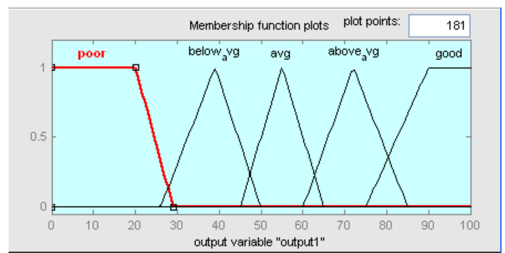


Figure 5: Generated membership function plot for output

c) The rule base for Inference created

ruleList= [1 1 1 1 1
1 2 2 1 1
1 3 2 1 1
1 4 3 1 1
5 2 2 1 1
5 3 3 1 1
2 5 4 1 1
3 1 2 1 1
3 2 3 1 1
3 3 3 1 1
4 2 2 1 1
4 3 3 1 1
4 4 4 1 1
4 5 5 1 1
1 5 3 1 1
2 1 1 1 1
2 2 2 1 1
2 3 2 1 1
2 4 3 1 1
3 4 4 1 1
3 4 4 1 1
4 1 2 1 1
5 1 2 1 1
5 4 4 1 1
5 5 5 1 1]

Snapshot of results obtained by Fuzzy Classification based on Membership function generated by Clustering (FCMFC) .

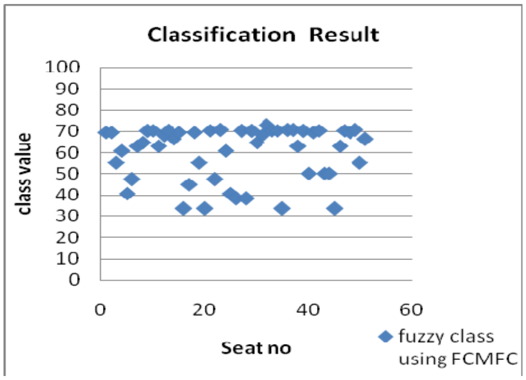


Figure 6: Classification result of Auto generated Fuzzy Inference System

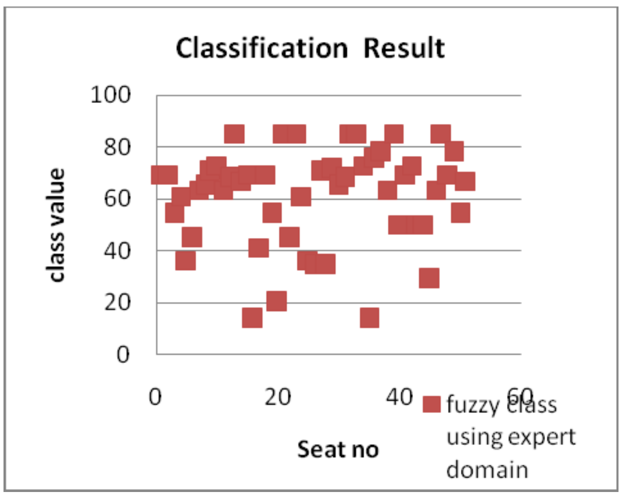


Figure 7: Classification result of Fuzzy Inference System using expert domain knowledge

The detail result in shown in Appendix B

7.0 CONCLUSION

The paper describes an approach for classification of students using fuzzy logic with membership function generated using semi supervised k means clustering and creation of Fuzzy inference system for reasoning automatically given the input data The approach was carried out on student examination data ,after obtaining the results difference is seen between the classes that are created based on the fuzzy membership function created using expert knowledge and proposed Fuzzy Logic classification method . The class value for the FCMFC are in range of 30 to 75 which means classification of students is below average to above average referring to figure 5, comparing it to expert domain method the class vale is in range of 10 an above 80.

REFERENCES

[1]. A. K. Verma, Anil R and Dr. Om Prakash Jain ,Fuzzy Logic Based Revised Defect Rating for Software Lifecycle Performance" Prediction Using GMR,BIJIT - BVICAM's International Journal of Information Technology, Vol. 1 No. 1; January – June, 2009

[2]. P. C. Jha, Shivani Bali and P. K. Kapur Fuzzy Approach for Selecting Optimal COTS Based Software ProductsUnder Consensus Recovery Block Scheme BIJIT - BVICAM's International Journal of Information Technology, January – June, 2011; Vol. 3 No.1

[3]. J. Sanz, et.al, "A Genetic Algorithm for Tuning Fuzzy Rule-Based Classification Systems with Interval- Valued Fuzzy Sets" WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Barcelona, Spain

[4]. Roberto R.F. Mendes et.al " Discovering Fuzzy Classification Rules with Genetic Programming and Co-evolution ", LNAI 2168, pp. 314-325, Springer-Verlag Berlin Heidelberg 2001

[5]. Martin Holena "Extraction Of Fuzzy Logic Rules From Data By Means Of Artificial Neural Networks"Kybernetika — Volume 41, Number 3, Pages 297-314 (2005)

[6]. Ujjwal Maulik and Sanghamitra Bandyopadhyay" Genetic algorithm-based clustering technique", Journal of Pattern Recognition Society ,published by Elsevier Vol 33 pp 1455-1465, (2000)

[7]. Sushmita Mitra, Rajat K. De, and Sankar K. Pal, Fellow, IEEE "Knowledge-Based Fuzzy MLP for Classification and Rule Generation " IEEE Transactions On Neural Networks", Vol. 8, No. 6,pp 1338-1350 November 1997

[8]. Johannes A. Roubos a, Magne Setnes b, and Janos Abonyi, "Learning fuzzy classification rules from labeled data " International Journal of Information Sciences published by Elsevier vol. 150 pp 77–93 (2003)

[9]. Gursewak S. Brar, Yadwinder S. Brar and Yaduvir Singh "An Efficient Data Clustering Algorithm Using Fuzzy Logic For Control Of A Multi Compressor System", IETECH Journal of Electrical Analysis, Vol: 1, No: 2,pp 130-136, 2007

[10]. Mofreh A. Hogo, "Evaluation of E-Learners Behavior using Different Fuzzy Clustering Models: A Comparative Study", (IJSIS)International Journal of Computer Science and Information Security Vol.7 No.2 pp: 131-140 (2010)

[11]. Suchita Borkar and K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining" IJCSMC International Journal of Computer Science and Mobile Computing, ISSN 2320–088X , Vol. 2, Issue. 7, pg.273 – 279 (2013)

[12]. Tzu-Ping Wu and Shyi-Ming Chen, "A new method for constructing membership functions and fuzzy rules from training examples", Systems, Man, and Cybernetics, Part

B: Cybernetics, IEEE Transactions Vol:29 , Issue: 1 pp 25 – 40, Feb 1999
[13]. Cesar Vialardi, Javier Bravo, Leila Shafti, and Alvaro Ortigosa, "Recommendation in Higher Education Using Data Mining Techniques", Proceedings Educational Data Mining EDM 2009, pg 190-199 , July 2009
[14]. Hisao Ishibuch , Ken Nozaki, Naohisa Yamamoto and Hideo Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms", Fuzzy Sets and Systems published by Elsevier vol 65 pp 237-253 (1994)
[15]. Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C," Application of k-Means Clustering algorithm for prediction of Students' Academic Performance", (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 1,pp: 292-295, 2010
[16]. Masoud Makrehchia, Mohamed S. Kamelb, "An information theoretic approach to generating fuzzy hypercubes for if-then classifiers", Journal of Intelligent & Fuzzy Systems vol. no 22 pp: 33–52 (2011)
[17]. Oscar Cordón, Francisco Herrera, and Pedro Villar "Generating the Knowledge Base of a Fuzzy Rule-Based System by the Genetic Learning of the Data Base ", IEEE Transactions On Fuzzy Systems, Vol. 9, No. 4, Pp 667 - 674 ,August 2001
[18]. Shruti S. Jamsandekar and Ravindra R. Mudholkar, "Self generated fuzzy membership function using ANN clustering technique", International Journal of Latest Trends in Engineering and Technology Special issue IDEAS 2013, pp 142-152 (2013)
[19]. MATLAB Fuzzy Logic toolbox 2 user manual

APPENDIX A

seat no	Subject 1 marks	Subject 2 marks
1	58	60
2	58	52
3	51	49
4	53	50
5	46	51
6	49	47
7	54	56
8	55	59
9	59	52
10	61	67
11	54	57
12	57	58
13	70	71
14	56	60
15	58	53
16	29	20
17	48	32
18	58	60
19	51	36
20	34	24
21	73	81
22	49	56
23	76	62
24	53	59
25	46	52
26	45	45
27	59	60
28	45	48
29	60	52
30	55	56
31	57	49
32	81	71
33	72	64
34	61	56
35	30	31
36	63	57
37	65	47
38	54	44
39	70	57
40	50	43
41	58	57
42	61	60
43	50	50
44	50	44
45	41	44
46	54	59
47	73	60
48	58	45
49	65	51
50	51	29
51	56	50

APPENDIX B

seat no	fuzzy class using FCMFC	fuzzy class using expert domain knowledge for membership function
1	69.62347309	69.41016566
2	69.62347309	69.41016566
3	55.18495127	54.1697685
4	61.00549807	61.01374256
5	40.73463047	36.20670391
6	47.38790839	44.875
7	63.05024746	63.32084691
8	64.80762548	65.2206443
9	70.15167421	70.56240231
10	70.38158258	72.85905554
11	63.05024746	63.32084691
12	67.98971257	68.19705015
13	**70.23469398**	**85.08755154**
14	66.41664216	66.78798403
15	69.62347309	69.41016566
16	33.6341888	13.83333333
17	44.98225048	41.21428571
18	69.62347309	69.41016566
19	55.18495127	54.1697685
20	**33.68989073**	**20.1010453**
21	**70.41632333**	**85.1147541**
22	47.38790839	44.875
23	**70.61907508**	**85.1147541**
24	61.00549807	61.01374256
25	40.73463047	36.20670391
26	38.52894557	34.38131313
27	70.15167421	70.56240231
28	38.52894557	34.38131313
29	70.26044891	71.58372067
30	64.80762548	65.2206443
31	67.98971257	68.19705015
32	**72.87273016**	**85.1147541**
33	**70.35233429**	**85.1147541**
34	70.38158258	72.85905554
35	**33.65781565**	**13.83333333**
36	70.66148531	75.40150129
37	70.56933154	78.05734074
38	63.05024746	63.32084691
39	**70.23469398**	**85.08755154**
40	50.23745353	50
41	69.62347309	69.41016566
42	70.38158258	72.85905554
43	50.23745353	50
44	50.23745353	50
45	**33.62465554**	**29.21682243**
46	63.05024746	63.32084691
47	**70.41632333**	**85.1147541**
48	69.62347309	69.41016566
49	70.56933154	78.05734074
50	55.18495127	54.1697685
51	66.41664216	66.78798403

A New Extended Kalman Filtering for Shadow/Fading Power Estimation in Mobile Communications

George P. Pappas[1], **Subramaniam Ganesan**[2] and **Mohamed A. Zohdy**[3]

Submitted in August, 2013; Accepted in June, 2014

Abstract - This paper proposes a New Extended Kalman filter (NEKF) approach to improve local mean power estimation. The method is being validated using a GUI system model and then compared to existing methods, Kalman Filter (KF) with Gaussian and Non-Gaussian noise environments. Our analysis is showing that NEKF is a more accurate method in most situations. NEKF can accurately estimate the parameters and predict states in discrete nonlinear state-space for modeling shadow power.

Index Terms – Extended Kalman Filter; Fading Channel, Handoff, Kalman Filter, local mean, multipath, power estimation, shadowing, state space.

1.0 INTRODUCTION

High performance shadow/fading power estimation methods are very important for use in power control of mobile device and base station handoff. Wireless mobile communications has become an essential part of life, creating the need for research.

Mobile communication performance is affected to a large degree by fading. Fading is defined as the variation in attenuation of a signal over a particular transmission medium.

There are two main causes of fading between a mobile station (MS) and a base station (BS) [1-3]. One is multipath propagation, where the received signal strength fluctuates due to multiple paths, and shadowing (Local Mean), where the transmitted signal is lost through physical phenomena, such as absorption, refraction (Figure 1), scattering and diffraction. Shadowing is caused by obstacles, such as buildings or trees along the path of a signal from the base station (BS) to the mobile station [1-3]. Signal is affected by objects along the path of the signal as it gets reflected thus taking different paths changing the amplitude and phase, resulting in increased or decreased power at the receiver. As the mobile device (Figure 1) is moving relative to the base station the Doppler shift (Figure 2) is causing the received frequency to change in comparison to the emitted frequency. Improving the shadow power estimation and reducing the estimation error more users can be accommodated in the system.

For mobile users, frequently occurring fading dips will cause unnecessary and capacity degrading, retransmissions. To achieve a high throughput over fading channels, adaptive methods for adjustment of (e.g. the modulation alphabet, and the coding complexity) can be used [4-6].

These techniques require accurate shadow power estimation and prediction to combat time-variability.

Weighted sample average estimators of local mean power, are currently used in many wireless communication system providers [4, 30, 31, 34].

Window based estimators work best under the assumptions that the shadow power process is constant over the duration of the averaging window [1]. In reality shadow power varies with time due to fading, which causes deterioration of these estimates as the window size increases beyond a certain value. The window size depends on several variables. One variable is the vehicles velocity v, and sampling period Ts. There are other shadow fading characteristics that affect the optimal window size [23].

The Kalman Filter (KF) algorithm can be used for linear systems. There are continuous and discrete KF methods. KF is an optimal recursive estimator, for stochastic linear dynamic systems are minimized by the Mean Squared Error (MSE) method. Wiener-Kolmogorov filter was the predecessor that Kalman filter [2]. While KF can be applied to linear systems is not a good solution for systems with nonlinearities. NEKF Techniques have been proposed to modify KF to be applied to nonlinear systems. For example, NEKF has been proposed in nonlinear systems estimations by linearizing the estimated variables through deriving Jacobian matrices [2]. However, NEKF may not be a good choice in system with high nonlinearity, or systems that are very difficult to calculate their Jacobian matrices.

This paper has been organized as follows. Section 1 is an introduction. Section 2 explains the new method of Extended Kalman Filter (NEKF) used for determining Shadow power in mobile station. Section 3 explains the non linear state space model theory. Section 3.1 Multipath and Linear Kalman Filter 1st order. Section 4 is the measurements, simulations and results. Section 5 is the conclusion. Section 6 is future work.

Although statistical methods for parameter estimation of linear models in dynamic mobile communication systems have been developed, the estimation of both states and parameters of non linear dynamic systems remains also challenging and is being addressed in this paper.

Electrical and Computer Engineering Department, Oakland University, 2200 N. Squirrel Rd, Rochester, MI 48309USA
Email:[1]*gppappas@oakland.edu,* [2]*ganesan@oakland.edu and* [3]*zohdyma@oakland.edu*

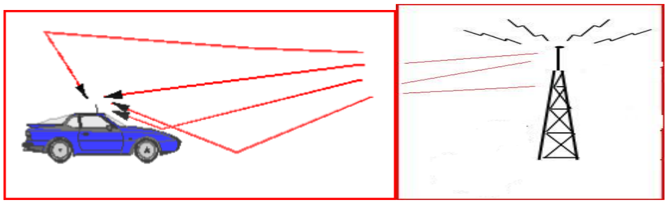


Figure 1: MS Refraction from BS

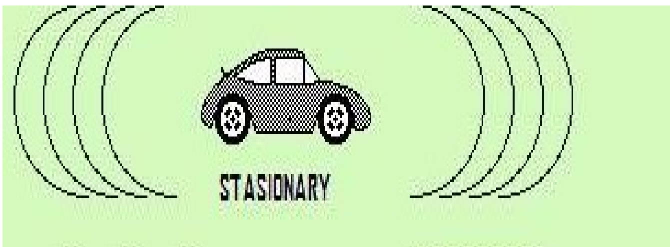


Figure 2: Doppler shift effect

2.0 EXTENDED KALMAN FILTER THEORY

KF is a form of a linear algorithm for the optimal recursive estimation of system state with specific set of output equations. Kalman filter equations can be separated into two parts [6]; the time update equations and the measurement update equation. The Process state is estimated at some time with feedback using measurements that contain noise as it can be shown in equation 4. The time update equations can also be called predictor equations while the measurement equation can be called corrector equations.

K_k: Kalman Gain
R: Environment noise covariance
Q: System noise covariance
P: Error covariance

$$x_{k+1} = f(x_k) + g(x_k, u_k) \tag{1}$$

$$y_k = h(x_k) \tag{2}$$

2.1 Time Update "Predicted"

Project the state ahead

$$\hat{x}_k^- = J_x \hat{x}_{k-1} + J_u u_{k-1} \tag{3}$$

Project the error covariance ahead

$$P_k^- = J_x P_{k-1} J_x^T + Q \tag{4}$$

2.2 Measurement Update "Corrected"

Compute the Kalman gain

$$K_k = P^-_k H_k^T (H_k P^-_k H_k^T + R_k)^{-1} \tag{5}$$

Update estimate with measurement z_k

$$\hat{x}_k = \hat{x}_k^- + K_k \left[z_k - h(\hat{x}_k^-) \right] \tag{6}$$

Update the error covariance

$$P_k = (I - K_k H_k) P^-_k \tag{7}$$

Initial estimates for $\hat{x}_{k-1}$ **and** P_{k-1}

where:
K_k: Kalman Gain
R: Environment noise covariance
Q: System noise covariance
P: Error covariance

The Extended Kalman Filter (NEKF) is the non linear extension of Kalman Filter (KF). NEKF is therefore suitable to take into account the non-linearity of the shadow power system model [7-9]. NEKF is a well known method and standard that has been considered in the theory of nonlinear state estimation [10]. KF and NEKF are known to be recursive data processing algorithms that estimate current mean and covariance. NEKF is reprocessing data at every time step without the need of storing previous measurements. The state distribution along with the mean and the covariance are being propagated analytically using a first order linearization. The predicted state estimation $\hat{x}_k^-$ for a linearized nonlinear process is expressed as follows:

$$\hat{x}_k^- = J_x \hat{x}_{k-1} + J_u u_{k-1} \tag{8}$$

The following expression is representing the error covariance P_k^- of the predicted state estimation:

$$\hat{x}_k^- = J_x \hat{x}_{k-1} + J_u u_{k-1} \tag{9}$$

where Q_{k-1} is the process noise, and $J_f(x^-_{k-1})$, $J_f^T(x^-_{k-1})$ are the Jacobian matrix and its transpose respectively. As it can be seen below J_f is the Jacobian matrix with partial derivative of all the state estimates:

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_n}{\partial y_n} \\ & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \frac{\partial f_n}{\partial y_2} & \cdots & \frac{\partial f_n}{\partial y_n} \end{pmatrix} \tag{10}$$

$$J_y = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \\ & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix} \tag{11}$$

Jacobian J_f and J_y matrices are shown in 10 and 11. K_k the weighting gain is defined by taking into account the measurement error that are due to the process noise as it can be seen in equation (3).The measurement matrix H_k is the Jacobian and R_k is the measurement noise. The gain K_k is directly proportional to H_k and inversely proportional to the measurement noise R_k. Expression (5) shows that the gain decreases to minimize the weight of the noise on the next estimate when the measurement noise increases while other factors are negligible. The predicted state estimation $\hat{x}_k^-$ is corrected by taking the effect of the measurements into account. The actual measurement z_k is compared to the predicted measurement $h(\hat{x}_k^-)$ and scaled by the relevant component of the measurement information, and inversely proportional to the measurement noise R_k. This expression (4) is significant. It shows that the gain decreases to minimize the weight of the noise on the next estimate, when the measurement noise grows and other factors are negligible.

$\hat{x}_k^-$ is the predicted state estimation. It's corrected by taking the effect of the measurements. z_k is the defined as the actual measurement and it is compared to the predicted measurement $h(\hat{x}_k^-)$ and adjusted by scaled K_k weighting gain (4).Equation (5) is expressing the correction phase of the algorithm. The error covariance is updated as shown in equation (6). The parameters of the model were varied in order to test robustness.

3.0 NONLINEAR STATE-SPACE MODEL THEORY

In control theory nonlinear state-space Model (NSSM) is a powerful tool for modeling of an unknown noisy system [10]. Nonlinear dynamical factor analysis (NDFA) scales only quadratic ally with the dimensionality of the observation space, so it is also suitable for modeling systems with fairly high dimensionality [10]. In NSSM, the observations have been generated from the hidden state y_i.

Linearization of the state-space equation by making the first order Taylor expansion around the current estimate $x_{i|i-1}$.

We have a linear state-space model of x_i. The state x_i can be estimated recursively using the solution of a normal linear state-space model.

3.1 Multipath and Non-Linear section Kalman Filter 2nd order

The Non-Linear KF 2nd order model is based on the following equations:

Prediction step:

$$\begin{bmatrix} \hat{S}_{1(n|n-1)} \\ \hat{S}_{2(n|n-2)} \end{bmatrix} = A \begin{bmatrix} \hat{S}_{1(n|n-1)} \\ \hat{S}_{2(n|n-2)} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \hat{S}_{1(n|n-1)} \\ \hat{S}_{2(n|n-1)} \end{bmatrix} \quad (16)$$

$$M_{(n|n-1)} = A[M_{(n|n-1)}]A^T + Q \quad (17)$$

KALMAN GAIN:

$$K_{(n)} = (M_{(n|n-1)} H^T (M_{(n|n-1)}) H^T + \sigma_H{}^2 \quad (18)$$

UPDATE STEPS:

$$\begin{bmatrix} \hat{S}_{1(n|n)} \\ \hat{S}_{2(n|n)} \end{bmatrix} = \begin{bmatrix} \hat{S}_{1(n|n-1)} \\ \hat{S}_{2(n|n-2)} \end{bmatrix} + K \left(P_{(n)} - H \begin{bmatrix} \hat{S}_{1(n|n-1)} \\ \hat{S}_{1(n|n-1)} \end{bmatrix} \right) \quad (19)$$

$$M_{(n|n)} = (I - K_{(n)} H)(M_{(n|n-1)}) \quad (20)$$

$$S^* = H \begin{bmatrix} \hat{S}_{1(n|n)} \\ \hat{S}_{2(n|n)} \end{bmatrix} \quad (21)$$

S^* Is the optimal estimate of the second order shadow process

In order to stabilize continuous and discrete-time systems one has to use time-dependent or discontinuous feedback controls. On the other hand, the criterion of stabilization in the class of R_x^n piecewise-constant feedbacks is established. The piece wise-constant (figure 3) feedback is associated with a piece wise-constant function of the form u – u(x) where x ∈ R.

Piecewise constant is used in the nonlinear form (figure 4) for the coefficient α.

The coefficient α is given by the following equations:

$$S(n) = a_2 S(n-2) + a_1 S(n-1) + \varphi_n \quad (22)$$

$$a_1 = e^{\frac{vT_s}{X_{c1}}} \quad (23)$$

$$a_2 = e^{\frac{vT_s}{X_{c2}}} \quad (24)$$

where a_1 and a_2 are shadow power coefficients. $S(n)$ is $S(nT_s)$, $\varphi_1(n)$ is zero mean white Gaussian noise with variance σ_ϕ^2 that equals to $(1-a^2)*\sigma_s^2$

where the mean $E[S(n)]$ decreases monotonically as n increases.

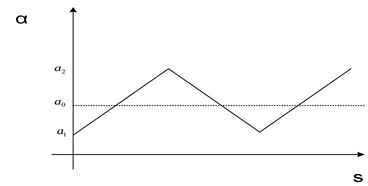


Figure 3: Linear as coefficient constant.

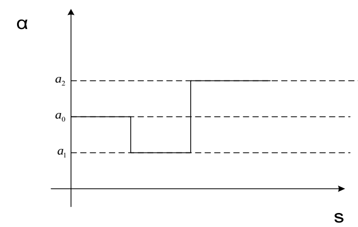


Figure 4: Nonlinear as coefficient piecewise constant.

4.0 GUI MEASUREMENT SIMULATION AND RESULTS

A Graphical User Interface (GUI) was developed to better assist with varying the different parameters. In Figure 5 a flow chart of the GUI identifies the different sections of the code.

Several simulations were executed. After running multiple simulations the results can be shown in Figures 6,7,8, 9.10,11,12,13,14. Clearly the NEKF applied on the incoming signal is performing as expected. The results show that the NEKF are very close to the incoming signal.

It's shown also that NEKF performs satisfactory within the range of -5dB to 5dB. Non-Gaussian noise distributions can be modeled as additive zero-mean Gaussian distributions.

Even though the computational complexity of NEKF is higher than the KF the results are satisfactory. The assumption made when using KF is that the shadow process is driven by white Gaussian noise. It's a window free approach when multipath is white [1].

There are two scenarios that are being taken into account. One scenario is the suburban were the shadow process coefficient can be regarded as constant for a wide range of velocities due to a large X_c with X_c being the correlation distance [1].As an example, in the case X_c= 500m when the velocity v is in the range from 5 to 80km/hr while the sampling period can be chosen as T_s=0.01s,in which case a is between [.9841, 0.9990].

The other scenario is the urban (Figure 12) case with a small X_c and large range in velocity v.

Figure 5 is the flow chart of the graphical interface (GUI)

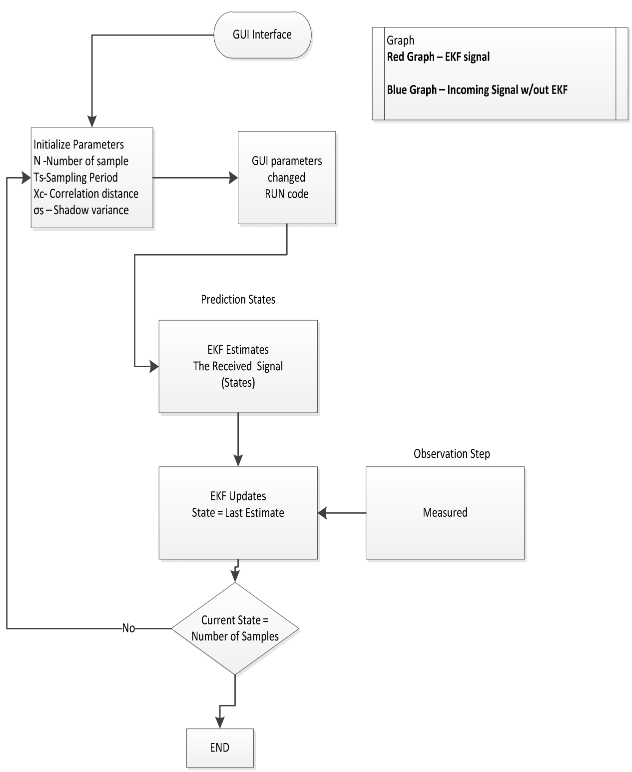


Figure 5: GUI Flow Chart.

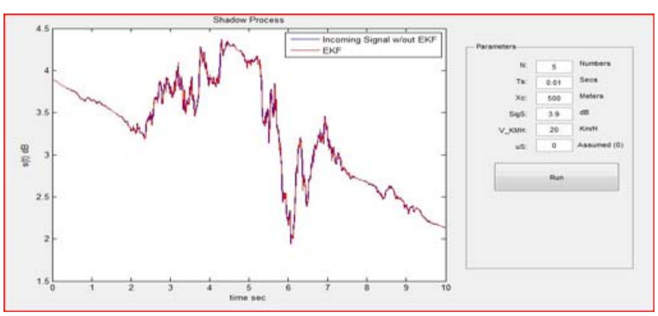


Figure 6: Simulink GUI estimation mean =0 variance = 3.9

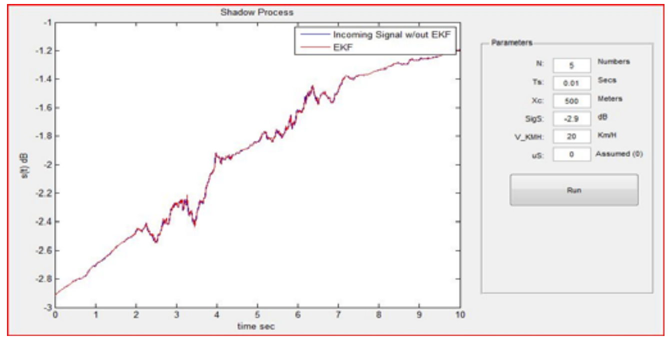


Figure 7: Simulink GUI estimation S(t) = -2.9 db

Simulink GUI was also created to give us the ability of changing the parameters and being able to see the results as in figure 7.

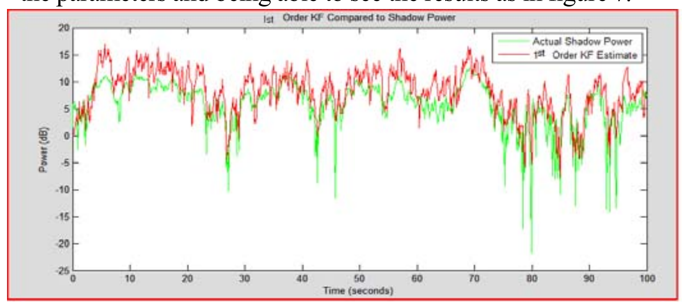


Figure 8: Compare 1st order KF with actual Shadow Power.

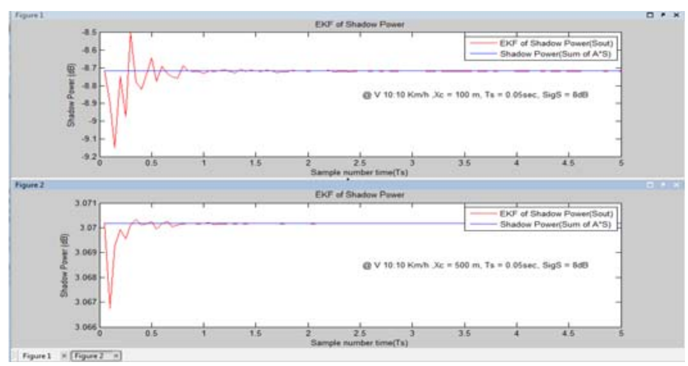


Figure 9: NEKF of Shadow Power at low speeds range [100 500] m.

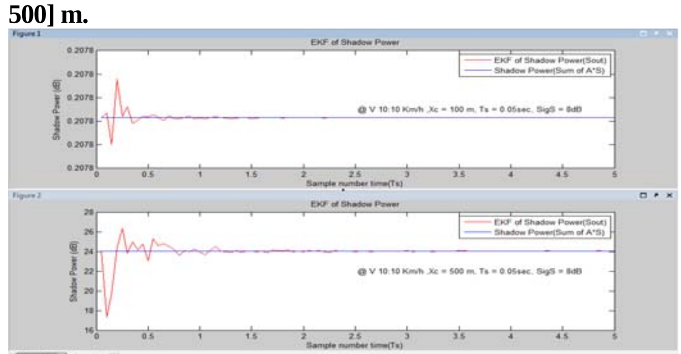


Figure 10:. NEKF ofShadow Power at low speedsrange [100- 500] m.

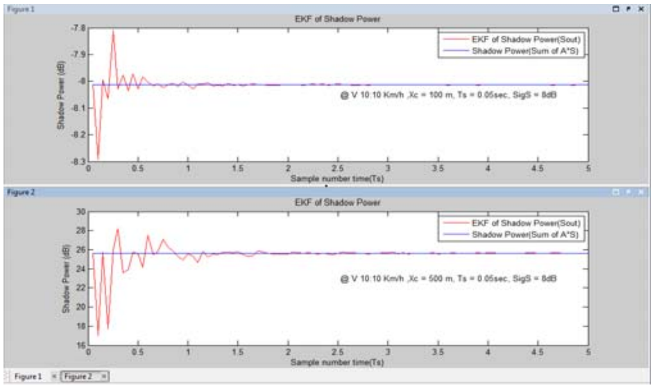


Figure 11: NEKF of Shadow Power at low speedsrange [100- 500] m.

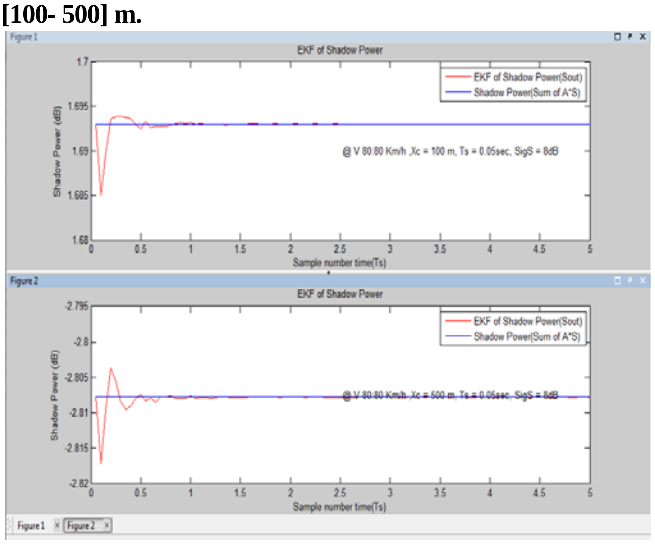


Figure 12: NEKF of Shadow Power at high speedsdistance [100 -500] m.

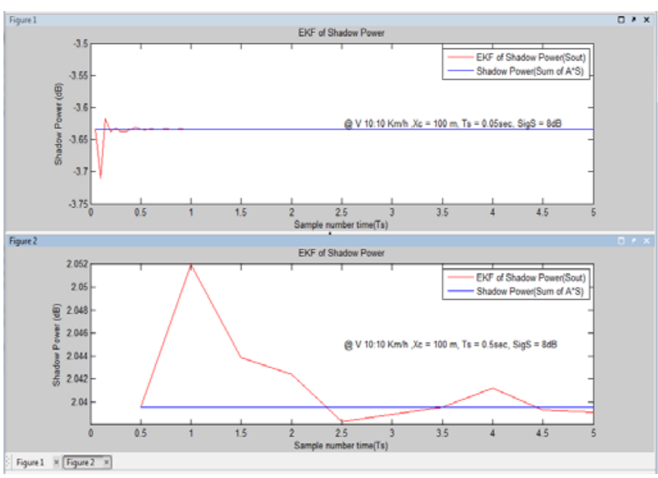


Figure 13: NEKF of Shadow Power at low speeds Time sample[.05 .5] sec.

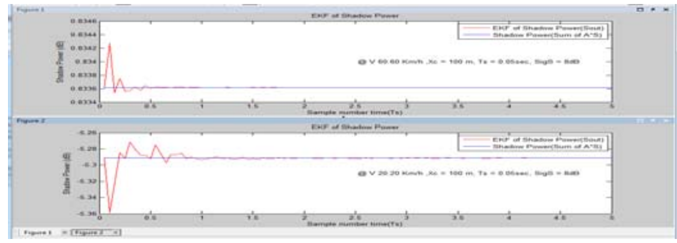


Figure 14: NEKF of Shadow Power Velocity [20 60] Km/h.

5.0 CONCLUSION

In this paper, the NEKF method has been proposed to optimize the performance of the shadow/fading power estimation. Simulation results show that the incoming signal is being tracked in a satisfactory manner. Increasing the shadow power variance has a direct affect in increasing the noise level seen in the shadow power estimate. Having mean =0 and variance = 3.9 then we receive 0 signal on the average (though there is a probability Shadow can be very small according to distribution) increases the noise. The major variables that are considered in this paper are;$\boldsymbol{X_c}$ which is the effective correlation distance, $\boldsymbol{T_s}$ the sampling period, v the vehicle velocity, a the correlation coefficient. In a suburban scenario, the shadow process coefficient which is a factor of the effects of environmental diversity that plays a role in wireless communications is regarded as constant for a wide range of velocities due to the fairly large $\boldsymbol{X_c}$.$\boldsymbol{X_c}$ is the correlation distance with v being vehicle velocity and $\boldsymbol{T_s}$ the sampling period. For example, in the case $\boldsymbol{X_c}$ = 500m when the velocity is in the range from 5 to 80km/hr, the sampling period can be chosen as T_s= 0.01s, in which case a is between [.9841, 0.9990].The results have also shown that this method is more efficient when implemented in both multipath affected signals. NEKF performs significantly better than KF while preserving their structures. Parameters have been changed to simulate conditions of typical urban areas as well as rural ones. The implementation of NEKF in this paper can be used in other wireless communication devices besides the cellular phones [32].

6.0 FUTURE WORK

1. Apply the approach to larger and more complex nonlinear (NL) models and joint state/parameter estimation.
2. Extend the analysis to higher order non-Gaussian channel models.
3. Assess the impact and effects of path loss including multiusers in shadow-fading environment.

REFERENCES

[1]. Tao Jiang, Nicholas D. Sidiropoulos and Georgios B. Giannakis “Kalman Filtering for Power Estimation in Mobile Communication, January 2003.

[2]. M. Ali and M. Zohdy, "Interactive Kalman Filtering for Differential andGaussian Frequency Shift Keying Modulation with Application inBluetooth," Journal of Signal and Information Processing, Vol. 3 No. 1, 2012.

[3]. C.Tepedelenlioğlu,N.D.Sidiropoulos,and G.B. Giannakis,“Medianfilteringforpowerestimationinmobile

communicationsystems,"inProc.3rdIEEESignalProcessingWorkshoponSignalProcessingAdvancesin WirelessCommunications,Taiwan, Mar.20–23, 2001.

[4]. V. S. Rana, "An Innovative Use of Information and Communication Technology (ICT) in Trade Facilitation in India", BIJIT - BVICAM's International Journal of Information Technology, Vol. 4 No. 2, July - December, 2012.

[5]. M.AkarandU.Mitra,"Variationsonoptimaland suboptimalhandoffcontrolforwirelesscommunicationsystems,"IEEEJ.Select.AreasCommunication.,vol.19, June2001.

[6]. Avijit Dutta, "Digital Communication and Knowledge Society India", BVICAM's International Journal of Information Technology, BIJIT, Vol. 4 No. 2, July - December, 2012.

[7]. D.WongandD.C.Cox,"Estimatinglocalmeansignal power levelina Rayleigh fading environment,"IEETrans.Veh.Technol.,vol.48,pp.956–959,May1999.

[8]. A.J.Goldsmith,L.J.Greenstein,andG.J.Foschini,"Errorstatisticsofreal-timepowermeasurementsicellularchannelswithmultipathandshadowing,"IEEETrans.Veh.Technol.,vol.43,pp.439–446,Aug.1994.

[9]. Ashema Hasti, 'Study of Impact of Mobile Ad hoc Networking and its Future Applications', BVICAM's International Journal of Information Technology,BIJIT

[10]. J. Paduart, J. Schoukens, R. "Nonlinear State Space Modeling ofMultivariable Systems", 2005

[11]. M.AkarandU.Mitra,"Variationsonoptimaland suboptimalhandoffcontrolforwirelesscommunicationsystems,"IEEEJ.Select.AreasCommunication., vol.19,June2001.

[12]. G. P. Pollini, "Trends in handover design," IEEE Commun.Mag., vol.34, pp. 82–90, Mar. 1996.

[13]. G. L. Stüber, Principles of Mobile Communication. Norwell, MA:Kluwer, 1996.

[14]. D. Labarre, E. Grivel, Y. Berthoumieu, E. Todini and M. Najim, "Consistent Estimation of Autoregressive Parameters from Noisy Observation Based on Two Interacting Kalman Filters,"Signal Processing, Vol.86, No. 10, 2006.

[15]. R.VijayanandJ.M.Holtzman,"Amodelforanalyzinghandoffalgorithms,"IEEETrans.Veh.Technol.,vol.42,Aug.1993.

[16]. A.J.GoldsmithandS.G.Chua,"Variable-ratevariable owerMQAMforfadingchannels,"IEEE Transactions on Communications.,vol.45,Oct.1997.

[17]. S.WeiandD.L.Goeckel,"Adaptivesignalingbasedonmeasurementswithstatisticaluncertainty,"inProc.Conf.Rec.33rdAsilomarConf.Signals,Systems,andComputers,vol.1,99

[18]. A.Duel-Hallen,S.Hu,andH.Hallen,"Long-range predictionoffadingsignals,"IEEESignalProcessing Mag., vol.17,pp.62–75,May2000.

[19]. L.HanzoandJ.Stefanov,"Thepan-Europendigital cellularmobileradio system nownas GSM,"in MobileRadioCommunications,R.Steel,Ed. New York:IEEEPress,1994.

[20]. Motorola,Inc.,FinalTextforPACSLicensedAir Interface(TAG3)J-STD014,June1995.

[21]. C.Tepedelenlioğlu,N.D.Sidiropoulos,andG.B.Giannakis, "Medianfilteringforpowerestimationin mobilecommunicationsystems,"in Proc.3rdIEEE SignalProcessingWorkshoponSignalProcessing Advancesin WirelessCommunications,Taiwan, Mar. 20–23, 2001,pp.229–231.

[22]. A.Chockalingam,P.Dietrich,L.B.Milstein,andR.R.Rao," Performanceofclosed-looppowercontrolin DS-CDMAcellularsystems,"IEEETrans.Veh. Technol., vol. 47, pp.774–789,Aug.1998.

[23]. D.Giancristofaro,"Correlationmodelforshadow fadinginmo bileradio channels,"Electron.Lett.,vol. 32,no.11,pp.958–959,May1996.

[24]. C.Tepedelenlioğlu,A.Abdi,G.B.Giannakis,andM. Kaveh "Estima-tionofdopplerspreadandsignal strengthin mobilecommunicationswithapplications to handoff andadaptivetransmission,"inWireless CommunicationsandMobileComputing.NewYork: Wiley,Apr./June2001.

[25]. Dogandžić, Alelcsandar and Zhang, Benhong ,"Dynamic shadow-power estimation for wireless communications", Signal Processing & Analysis, Volume: 53 , Issue: 8,Aug. 2005.

[26]. Harb A.M., Zohdy M.A. Synchronization of two chaotic systems asapplied in communication systems, submitted to Int. Journal of NonlinearDynamics, Aug.,2001

[27]. G. Welch, G. Bishop, "An Introduction to the Kalman Filter,"UNC-CHComputer science Technical Report (2005).

[28]. Nsour, Ahmad R | Zohdy, Mohamed A "One and two dimensionalself-organized learning applied to global positioning system (GPS) data.WSEAS Transactions onInformation Science and Applications.Vol. 3,no.Dec.2006.

[29]. R.E.Kalman,"ANewApproachtoLinearFilteringandPredictionProblems,"ResearchInstituteforAdvancedStudy,Baltimore,1960.

[30]. W.TamandF.C.M.Lau,"Analysisofpowercontrolanditsimperfectionsin CDMAcellularsystems,"IEEETrans.Veh.Technol.,vol.48, Sept.1999.

[31]. W.C.Jakes,MicrowaveMobileCommunications. New York:IEEEPress, 1974.

[32]. Sriram Vaitheswaran, "Virtual Eye For the Blind" in Proceedings of the BVICAM's IETE National Conference; INDIACom-2007, Feb.23-24, 2007.

[33]. J.ProakisandM.Salehi,"DigitalCommunications,"5th Edition,McGrawHillHigherEducation,NewYork, 2008.

[34]. B. V. Ramanamurthy, K. Srinivas Babu and Mohammed Shafruddin, 'Dynamic Data Updates for Mobile Devices by Using 802.11 Wireless Communications', BIJIT - BVICAM's International Journal of Information Technology, Jan-June, 2011 Vol.3 No.1.

Thermal Model and Characteristics of Double Slope Solar Still

Rahul Dev[1] and **Amimul Ahsan**[2]

Submitted in August, 2013; Accepted in February, 2014

Abstract - In this paper, a new thermal model (after assuming temperatures of east and west side glass covers equal to their average temperature) and characteristics of double slope solar still (DSSS) have been developed. Experimental and theoretical results have been compared for the composite climate of New Delhi, India. Theoretical results obtained by previous and new thermal models of DSSS have been found in fair agreements with experimental observations. The characteristics have been developed under quasi-steady state of the solar still by linear and non-linear regression curves between daylight instantaneous gain/loss efficiencies and factor $\left\{\left(T_{wEW}-T_a\right)/\left(I_s(t)_E+I_s(t)_W\right)\right\}$.

Index Terms – Double slope solar still, Characteristic equation, Instantaneous efficiency, Thermal testing.

1.0 INTRODUCTION

Solar still is a solar energy operated water purification system. Researchers have developed several passive and active solar stills over the years [1-4]. Nebbia and Menozzi have mentioned Della Porta's solar still which was designed to extract the essence of herbs [5,6]. Dunkle [7] have developed correlations for various rates of heat and mass transfer of solar still. Similarly, Tsilingiris [8] have also developed temperature dependent correlations for internal heat transfer coefficients between water and glass cover of solar still and it was shown that these are affected adversely above 60 ºC water temperature. Rubio et al. [9-11] have studied asymmetries in temperatures of water and glass cover, and amount of distillate for a double slope solar still (DSSS) by mathematical model (in terms of lumped parameters, and controlled temperatures of glass cover and basin). A correlation of mass transfer for condensing chamber of DSSS at different operating temperature ranges has been developed by Agrawal et al. [12].Omri et al. [13] have examined the natural convection effects in solar stills for governing parameters such as Rayleigh number and tilt angle of glass cover. It has been shown that flow structure is sensitive to cover tilt angle. Dwivedi and Tiwari [14] have reported that DSSS produces daily yield 1.44 kg/m^2 at 'low production cost'. It has also been said that DSSS have energy efficiency 1.35% and energy pay back time (EPBT) 1.6 years. A thermal model for DSSS and carbon credit earned by it, has been studied by Dwivedi and Tiwari [15]. Dwivedi and Tiwari [16] have developed a thermal model for double slope active solar still (DSASS). Cooper [17] has studied solar still's efficiency and suggested maximum efficiency can be as high as 60%. Tamimi [18] has characterized the characteristic curves of solar still between its ideal and the worst conditions. Boukar and Harmim [19] have obtained characteristic curve for vertical solar still and found its daily overall energy efficiency ranges from 7.9 to 21.2% under harsh condition of Sahara desert, Algeria. Tiwari and Noor [20] have presented the concept of an instantaneous thermal efficiency to characterize the solar stills. Dev and Tiwari [21-24] have developed the characteristic equation for single slope passive solar still, hybrid photovoltaic thermal (PVT) active solar still, double slope solar still and inverted absorber solar still at different water depths, inclinations and climatic conditions. Dev and Tiwari [21-24] have also reported that the non-linear characteristic curves are best fit to predict the behavior of solar stills. Dev et al. [23] have given characteristic equations for double slope solar still considering east and west sides of it separately which arose a need for a single characteristic equation for a single system instead of dividing a system into parts. A study of DSSS has been presented by Dwivedi [25]. Recently, a review on solar still have been done including various other water purification technologies e.g. membrane distillation [26]. Singh et al. [27] have experimentally studied the performance of hybrid PVT double slope active solar stills and found its daily average thermal efficiency 17.4% and highest yield 7.54 kg/day for parallel configuration of two flat plate collectors in forced mode of operation in October 2010 for composite climate of Ghaziabad, Uttar Pradesh, India. Dev [28] has studied various passive and active solar stills to develop their modified thermal models and characteristic equations. Dev and Tiwari [29] has recently studied the annual performance of evacuated tubular collector integrated single slope solar still to produce hot water along with distilled water (as a heat recovery from the evacuated tubular collector when water is stored at high temperatures).

Kumar and Agarwala [30] shown the energy computing models for optimally allocating different types of renewable in the distribution system for minimizing energy loss and optimizes the integration of renewable energy resources with technical and financial feasibility. Kamthania and Tiwari [31] shown the way to determine efficiency of semi transparent hybrid photovoltaic thermal double pass air collector for different PV technology and compare it with single pass air collector using artificial neural network (ANN) technique for New Delhi

[1]Mechanical Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad-211004, Uttar Pradesh, India.
[2]University Putra Malaysia, Dept. Civil Engineering, Faculty of Engineering, 43400 UPM Serdang, Selangor, Malaysia.
E-mail:[1]rahuldsurya@yahoo.com

weather station of India. Siddiqui et al. [32] developed efficiency metrics and time with efficiency relationship for the programmers and customers. Kumar et al. [33] describes energy efficient clustering for wireless sensor network with a large number of tiny sensor nodes to be used effectively in various applications with data accumulation. Kanchan and Kamthania [34] calculate the energy payback time for a building integrated semitransparent thermal (BISPVT) system with air duct. The work [30-34] shows broadly the calculation of efficiency with various parameters by software programmes including ANN, Matrices etc.
The objective of this paper is to obtain a new thermal model of DSSS (after assuming temperatures of both east and west side glass covers equal to their average temperature) and validating it by using experimental observations. The thermal model has been extended to obtain the expressions of daylight instantaneous gain and loss efficiencies of DSSS. Following these expressions, the characteristic curves of DSSS for its experimental performance under the composite climate of New Delhi, India, have been obtained and analyzed. The corresponding equations of these characteristic curves have been obtained and are to be used for thermal testing of DSSS for different climatic conditions, design parameters and materials.

2.0 EXPERIMENTAL SETUP AND OBSERVATIONS

A photograph of the experimental setup of DSSS is shown in Fig. 1. The experimental setup has been installed at Solar Energy Park, IIT Delhi, New Delhi, India. The orientation of the solar still has been kept east-west to receive solar radiation for maximum hours of sunshine and for increasing the heat addition into solar still. This affects the amounts of distillate on both sides of the solar still. The solar radiation incidents on both the glass covers and absorbed through the basin liner which heats the water. A small fraction of solar radiation is also absorbed by the glass covers and the water due to their absorptivities. The basin liner transfers the heat to the water through convection effect and also to the surroundings through the basin liner by conductive effect. The evaporated water from the basin comes in contact with the inner surfaces of the glass covers and releases its heat to the glass covers. The vapor again comes into the form of liquid water which trickles down to the troughs and then to the collecting jars placed at the east and west sides of the solar still. The glass takes the heat from the absorbed solar radiation and from the vapor which is released to the atmosphere by the wind through convection effect and radiation effect. In this phenomenon, two phase changes take places, (i) from basin water to vapor leaving impurities on the basin, and (ii) from vapor to water at the inner surface of glass covers. The amount of distillate depends upon the temperature difference between water and glass cover because of that it varies almost unequally at east and west sides of DSSS during daytime. When the sun lies in the east direction then higher temperature difference occur at west side due to low glass temperature which yields higher amount of distillate at this side and vice versa except at the time of noon when both the glass covers have almost the same temperatures [14, 23, 25]. The distillate produced by the solar still has total dissolved solid count in the range of 8-15 ppm because some materials also evaporate within 0-100 ºC and comes out with the distillate. The durations of sunshine hour at New Delhi are: 05:00-19:00 h (in summer);In winter this is 07:00-17:30 h (if no fog appears in the morning) or 9:00-17:30 h (if fog appears in the morning), with timings of solar noon in the respective month.

Figure 1: A photograph of double slope solar still.

DSSS has a box type structure of basin area 2 m^2. It has been made of the glass reinforced plastic (GRP) of thickness 5 mm painted black at inside surfaces to absorb the solar radiation. The heights of the solar still walls are 0.22 m at the east-west ends and 0.48 m at the center. Two simple window glasses of dimensions $1.03 \times 1.06 \times 0.004$ m^3 have been placed over the walls of the solar still at inclination angle 15º. The reasons of selecting this inclination angle are: (i) to receive maximum solar radiation when the sun is in the east or west direction, (ii) to guide condensed water into trough under the effects of forces namely adhesion force (between glass and water), cohesion force (between water molecules), and gravitational force, (iii) the minimum inclination angle can be chosen as $\left\{= latitude\, of\ the\ place\left(\phi_l\right) - 15^o\right\}$ for any solar still [28]. As the latitude of the place is 28°35′ N for New Delhi, the angle of inclination can be equal to 13°35′ which is taken as 15° (considering approximation) for the removing complexity and better understanding of the fabricator. An inlet has been provided through the north wall of the solar still to feed the brackish/saline water (i.e. total dissolved solids in the range of 1200 to 2000 parts per million as per availability in New Delhi). Two troughs have been fixed at inside surfaces of both the east and west walls of the solar still to collect the distilled water to guide the distillate into the collecting jar [14, 23, 25]. Figs. (2-5) show the various measured parameters on hourly basis for Dec'05 and Jun'06. Fig. 2 shows the hourly variations of incident solar radiation on both glass covers of DSSS.

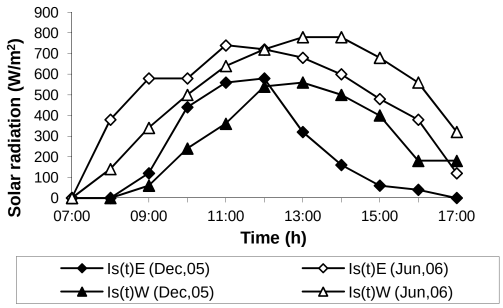


Figure 2: Hourly variations of solar radiation on both the east and west glass covers of DSSS on 9/12/2005 and 13/06/2006 [14, 23, 25].

Fig. 3 shows the hourly variations of water and ambient temperatures. Fig. 4 shows the hourly variations of temperatures of inner surface of both glass covers. Fig. 5 shows the hourly variations of distillate obtained from both sides of DSSS. Similar observations have been recorded for other months in duration from Oct'05 to Sep'06 [14, 15, 23, 25]. In Table 1, 'various design and operational parameters' and 'instruments used for measurements' have been given. Various parameters like temperatures, solar radiation and quantity of distillate on hourly basis have been measured for DSSS [25]. The annual experimental data (from Oct'05 to Sep'06) for DSSS taken at water depth of 0.01 m (i.e. 20 kg in DSSS) for New Delhi, India has been used to carry out the present analysis. The variations of incident total solar radiations (i.e. sum of direct and diffuse) on east and west glass covers for each month from Oct'05 to Sep'06 have been studied. It also shows that during noon hours the magnitudes of incident solar radiations are more or less same for both glass covers.

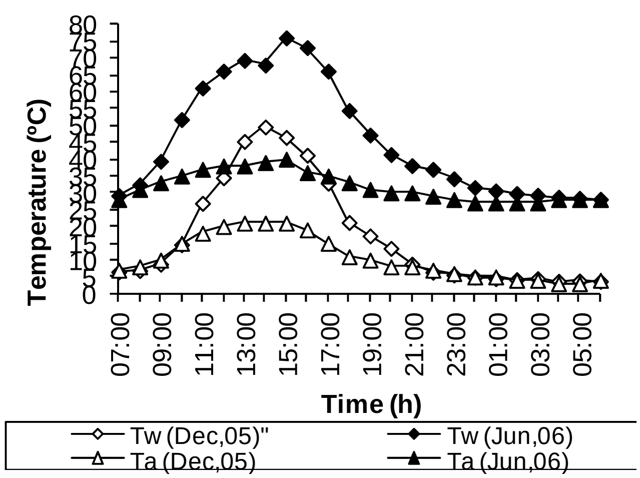


Figure 3: Hourly variations of water and ambient temperatures on 9/12/2005 and 13/06/2006 [14, 23, 25].

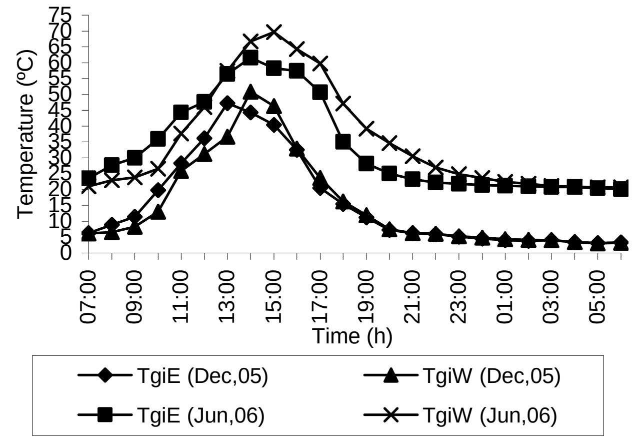


Figure 4: Hourly variations of temperatures of inner surfaces of both glass covers on 9/12/2005 and 13/06/2006 [14, 23, 25].

3.0 THERMAL MODEL

Following are the assumptions made for energy balance equations of different components of DSSS.

i. DSSS is vapor leakage proof and is in quasi steady state.
ii. There is no temperature gradient in the water inside the basin.
iii. Heat capacities of glass and basin material are negligible.
iv. Temperature dependant heat transfer coefficients have been considered.

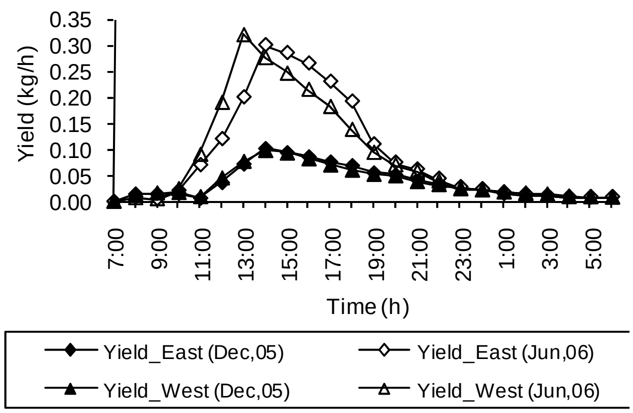


Figure 5: Hourly variations of distillate obtained from both sides of DSSS on 9/12/2005 and 13/06/2006 [14, 23, 25].

(a) Inner surface of east glass cover:

$$\begin{bmatrix}\text{Total energy absorbed}\\ \text{by inner surface}\\ \text{of east glass cover}\end{bmatrix} = \begin{bmatrix}\text{Total energy released}\\ \text{by inner surface}\\ \text{of east glass cover}\end{bmatrix}$$

i.e.

$$\begin{bmatrix}\text{Solar radiation}\\ \text{directly absorbed}\\ \text{glass cover}\end{bmatrix}+\begin{bmatrix}\text{Heat energy released}\\ \text{from vapor to inner}\\ \text{surface of glass}\\ \text{cover}\end{bmatrix}-\begin{bmatrix}\text{Heat energy}\\ \text{exchange}\\ \text{between both}\\ \text{glass covers}\end{bmatrix}$$

$$=\begin{bmatrix}\text{Energy released}\\ \text{by inner surface}\\ \text{of east glass cover}\\ \text{to its outer surface}\end{bmatrix}$$

$$\alpha'_g . I_s(t)_E . A_{gE} + h_{1wE} . A_b . (T_w - T_{giE}) -$$
$$U_{EW} . A_{gE} . (T_{giE} - T_{giW}) = h_{kg} . A_{gE} . (T_{giE} - T_{goE}) \quad (1)$$

(b) Outer surface of east glass cover:

$$\begin{bmatrix}\text{Total energy absorbed}\\ \text{by outer surface}\\ \text{of east glass cover}\end{bmatrix}=\begin{bmatrix}\text{Total energy released}\\ \text{by outer surface}\\ \text{of east glass cover}\\ \text{to ambient}\end{bmatrix}$$

$$h_{kg} . A_{gE} . (T_{giE} - T_{goE}) = h_o . A_{gE} . (T_{goE} - T_a) \quad (2)$$

(c) Inner surface of west glass cover:

$$\begin{bmatrix}\text{Energy absorbed}\\ \text{by inner surface}\\ \text{of west glass cover}\end{bmatrix}=\begin{bmatrix}\text{Energy released}\\ \text{by inner surface}\\ \text{of west glass cover}\end{bmatrix}$$

i.e.

$$\begin{bmatrix}\text{Solar radiation}\\ \text{directly absorbed}\\ \text{glass cover}\end{bmatrix}+\begin{bmatrix}\text{Heat energy released}\\ \text{from vapor to inner}\\ \text{surface of glass}\\ \text{cover}\end{bmatrix}-\begin{bmatrix}\text{Heat energy}\\ \text{exchange}\\ \text{between both}\\ \text{glass covers}\end{bmatrix}$$

$$=\begin{bmatrix}\text{Total energy released}\\ \text{by inner surface}\\ \text{of west glass cover}\\ \text{to its outer surface}\end{bmatrix}$$

$$\alpha'_g . I_s(t)_W . A_{gW} + h_{1wW} . A_b . (T_w - T_{giW}) -$$
$$U_{EW} . A_{gW} . (T_{giW} - T_{giE}) = h_{kg} . A_{gW} . (T_{giW} - T_{goW}) \quad (3)$$

(d) Outer surface of west glass cover:

$$\begin{bmatrix}\text{Total energy absorbed}\\ \text{by outer surface}\\ \text{of west glass cover}\end{bmatrix}=\begin{bmatrix}\text{Total energy released}\\ \text{by outer surface}\\ \text{of west glass cover}\\ \text{to ambient}\end{bmatrix}$$

$$h_{kg} . A_{gW} . (T_{giW} - T_{goW}) = h_o . A_{gW} . (T_{goW} - T_a) \quad (4)$$

(e) Basin liner:

$$\begin{bmatrix}\text{Total Energy}\\ \text{absorbed}\\ \text{by basin}\end{bmatrix}=\begin{bmatrix}\text{Energy}\\ \text{released}\\ \text{by basin}\\ \text{to water}\end{bmatrix}+\begin{bmatrix}\text{Energy released}\\ \text{by basin}\\ \text{to ambient}\end{bmatrix}$$

$$\alpha'_b . \{I_s(t)_E + I_s(t)_W\} . A_b = h_{bw} . A_b . (T_b - T_w) +$$
$$h_{ba} . A_b . (T_b - T_a) \quad (5)$$

(f) Water mass:

$$\begin{bmatrix}\text{Total Energy absorbed}\\ \text{by water i.e. from basin}\\ \text{and directly absorption of}\\ \text{solar radiation}\end{bmatrix}=\begin{bmatrix}\text{Energy}\\ \text{stored in}\\ \text{water}\end{bmatrix}+\begin{bmatrix}\text{Energy released}\\ \text{by water to inner}\\ \text{surfaces of east and}\\ \text{west glass covers}\end{bmatrix}$$

$$+\begin{bmatrix}\text{Energy released}\\ \text{by water to ambient}\\ \text{through side walls}\end{bmatrix}$$

$$\alpha'_w . \{I_s(t)_E + I_s(t)_W\} . A_b + h_{bw} . A_b . (T_b - T_w) =$$
$$= (MC)_w . \frac{dT_w}{dt} + h_{1wE} . A_b . (T_w - T_{giE})$$
$$+ h_{1wW} . A_b . (T_w - T_{giW}) + h_s . A_s . (T_w - T_a) \quad (6)$$

Solving Eqs. (2) and (4), the following expressions have been obtained, temperatures of outer surface of east glass cover, of outer surface of west glass cover, inner surface of east glass cover, inner surface of west glass cover, basin:

$$T_{goE} = \frac{h_{kg} . T_{giE} + h_o . T_a}{h_{kg} + h_o} \quad (6a)$$

$$T_{goW} = \frac{h_{kg} . T_{giW} + h_o . T_a}{h_{kg} + h_o} \quad (6b)$$

$$T_{giE} = \frac{A_1 + A_2 . T_w}{p} \quad (6c)$$

$$T_{giW} = \frac{B_1 + B_2 . T_w}{p} \quad (6d)$$

$$T_b = \frac{\alpha'_b . \{I_s(t)_E + I_s(t)_W\} + h_{bw} . T_w + h_{ba} . T_a}{h_{bw} + h_{ba}} \quad (6e)$$

Eq. (6) has been rearranged as follows,

$$\frac{dT_w}{dt} + aT_w = f(t) \quad (6f)$$

The following assumptions have also been made for solving Eq. (6f),

i) Time interval dt ($0<t<dt$).

ii) Function $f(t)=\overline{f(t)}$ and a are constants for small interval dt.

iii) Initial values of water and condensing cover temperatures have been used to determine the value of internal heat transfer coefficients.

Eq. (6f) has been solved to get following expression of water temperature T_w,

$$T_W = \frac{\overline{f(t)}}{a}.\left(1-e^{-at}\right)+T_{WO}.e^{-at} \quad (6g)$$

Expressions of A_1, A_2, B_1, B_2, $f(t)$ and a have been given in Appendix.

The rate of evaporative heat transfer for east and west side of the solar still can be written as,

$$\dot{q}_{ewE} = h_{ewE}\left(T_w - T_{giE}\right) \quad (6h)$$

$$\dot{q}_{ewW} = h_{ewW}\left(T_w - T_{giW}\right) \quad (6i)$$

Substituting values of T_{giE} and T_{giW} in Eqs. (6h) and (6i) respectively, $\dot{q}_{ewE} = h_{ewE}\left\{T_w - \left(\left(A_1 + A_2 T_w\right)/p\right)\right\}$ (6j)

or

$$\dot{q}_{ewE} = \frac{h_{ewE}}{p}\{-\alpha_g^{'}.\left(I_s(t)_E\, h_{goW} + I_s(t)_E\, h_{1wW} + I_s(t)_E U_{EW} + I_s(t)_W\, U_{EW}\right) + \left(h_{goE}\, h_{goW} + h_{goE}\, h_{1wW} + h_{goE}\, U_{EW} + h_{goW}\, U_{EW}\right).\left(T_w - T_a\right)\} \quad (6k)$$

and

$$\dot{q}_{ewW} = h_{ewW}\left\{T_W - \left(\left(B_1 + B_2 T_W\right)/p\right)\right\} \quad (6l)$$

or

$$\dot{q}_{ewW} = \frac{h_{ewW}}{p}\left\{\begin{array}{l} -\alpha_g^{'}.\left(I_S(t)_E U_{EW} + I_S(t)_W\, h_{goE} + I_S(t)_W\, h_{1wE} + I_S(t)_W U_{EW}\right) + \\ \left(h_{goE} U_{EW} + h_{goE}\, h_{goW} + h_{goW} h_{1wE} + +h_{goW}\, U_{EW}\right).\left(T_w - T_a\right) \end{array}\right\} \quad (6m)$$

Eqs. (6k) and (6m) can be added to get the rate of total evaporative heat transfer by making some additional assumptions for a different approach of thermal modeling of DSSS. These assumptions are as follows:

v) Areas of both glass covers are 1 m^2.

vi) Temperatures of both glass covers are equal to their average temperature (T_{giEW}) i.e. $T_{giEW} = \left(T_{giE} + T_{giW}\right)/2$ (6n)

Then the temperature difference between water and glass cover temperatures become $\left(T_W - T_{giEW}\right)$. Due to which various internal heat transfers of both sides from water to glass covers become equal. Convective, radiative and evaporative heat transfer coefficients from water to east and west side glass covers can be written as:

$h_{cwE} = h_{cwW}$; $h_{rwE} = h_{rwW}$; $h_{ewE} = h_{ewW}$ which results $h_{1wE} = h_{1wW} = h_{1wEW}$ (6o)

vii) the total (radiative and convective) heat transfer coefficients from both glass covers to ambient air become equal as these terms depends upon wind velocity (Eqs. 1-4) i.e. $h_o = 5.7 + 3.8v = h_{oEW}$ (6p)

And, therefore, the heat transfer coefficients from inner surfaces of glass covers to ambient air also become equal and can be represented by following expressions, $h_{goE} = h_{goW} = h_{goEW}$ (6q)

where, $h_{goE} = h_{goW} = \dfrac{h_{kg}.h_o}{h_{kg} + h_o}$,

From Eqs. (1-6), Eqs. (6n - 6q) and assumptions similar to solve Eq. (6f), the following expression of water temperature can be obtained as,

$$T_{wEW} = \frac{\overline{f(t)_{EW}}}{a_{EW}}\left[1-e^{-a\Delta t}\right] + T_{w0}\, e^{-a\Delta t} \quad (7)$$

where, $a_{EW} = \dfrac{1}{\left(MC\right)_w}\left[\dfrac{2h_{bw}h_{ba}}{h_{bw}+h_{ba}} + \dfrac{h_{1wEW}}{p_{EW}}\left\{2h_{goEW}^2 + 4h_{goEW}U_{EW} + 2h_{goEW}h_{1wEW}\right\}\right]$;

$$p_{EW} = \left(h_{goEW}^2 + h_{1wEW}^2 + 2h_{goEW}U_{EW} + 2h_{1wEW}h_{goEW} + 2h_{1wEW}U_{EW}\right)$$

$$f(t)_{EW} = \frac{1}{(MC)_w}\left[\left(\alpha'_w + \frac{\alpha'_b h_{bw}}{h_{bw}+h_{ba}}\right)\left(I_s(t)_E + I_s(t)_w\right) + \frac{h_{1wEW}}{p_{EW}}\begin{Bmatrix}\left(\alpha'_g h_{1wEW} + h_{goEW} + 2U_{EW}\right)\left(I_s(t)_E + I_s(t)_w\right) \\ +T_a\left(2h_{goEW}h_{1wEW} + 2h_{goEW}^2 + 4h_{goEW}U_{EW}\right)\end{Bmatrix} + \left(\frac{2h_{bw}h_{ba}}{h_{bw}+h_{ba}}\right)T_a\right]$$

Adding Eqs. (6k) and (6m), the expression of combined rate of evaporative heat transfer from water to both glass covers can be written as

$$\dot{q}_{ewEW} = \frac{h_{ewEW}}{p_{EW}}\left(I_s(t)_E + I_s(t)_W\right)\left\{-\alpha'_g\cdot\left(2U_{EW} + h_{goEW} + h_{1wEW}\right) + \left(2h_{goEW}^2 + 4h_{goEW}U_{EW} + 2h_{goEW}h_{1wEW}\right)\left(T_{wEW} - T_a\right)/\left(I_s(t)_E + I_s(t)_W\right)\right\} \tag{8}$$

Further, the following linear expression of daylight instantaneous gain efficiency of DSSS $\left(\eta_i = \frac{\text{Heat output in the form of distilled water per unit time}}{\text{Total heat input in the system per unit time}} = \text{denoted as } y_1 \text{ also}\right)$ has been obtained by using Eq. (7).

$$\eta_{iEW} = \frac{\dot{q}_{ewEW}}{\left(I_s(t)_E.A_{gE} + I_s(t)_W.A_{gW}\right)} = (\alpha\tau)_{eff1} + (UA)_{eff1}\cdot\left\{\left(T_{wEW} - T_a\right)/\left(I_s(t)_E + I_s(t)_W\right)\right\} \tag{9}$$

where, $(\alpha\tau)_{eff1} = -\alpha'_g\cdot\left(h_{goEW} + 2U_{EW} + h_{1wEW}\right)\cdot\frac{h_{ewEW}}{p_{EW}}$; $(UA)_{eff1} = 2\left(h_{goEW}^2 + 2h_{goEW}U_{EW} + h_{goEW}h_{1wEW}\right)\cdot\frac{h_{ewEW}}{p_{EW}}$;

Similarly, an expression for the daylight instantaneous loss efficiency (i.e. efficiency due to heat storage effect of water

$\left(\eta_i = \frac{\text{Heat stored in the water per unit time}}{\text{Total heat input in the system per unit time}} = \text{denoted as } y_2 \text{ also}\right)$) can be given as,

$$\eta_{iLEW} = \frac{(MC)_w \times \left(T_{wEW} - T_{wo}\right)}{\left(I_s(t)_E.A_{gE} + I_s(t)_W.A_{gW}\right)} = (\alpha\tau)_{eff2} - (UA)_{eff2}\cdot\frac{\left(T_{wEW} - T_a\right)}{\left(I_s(t)_E + I_s(t)_W\right)} \tag{10}$$

Eq. (10) is also a linear characteristic equation.

where, $$(\alpha\tau)_{eff2} = \frac{(MC)_w\left(1-e^{-a\Delta t}\right)\left\{\alpha'_w + \frac{\alpha'_b h_{bw}}{h_{bw}+h_{ba}} + \frac{h_{1wEW}}{p_{EW}}\left(\alpha'_g h_{1wEW} + h_{goEW} + 2U_{EW}\right)\right\}}{\left\{\frac{2h_{bw}h_{ba}}{h_{bw}+h_{ba}} + \frac{h_{1wEW}}{p_{EW}}\left(2h_{goEW}^2 + 4h_{goEW}U_{EW} + 2h_{goEW}h_{1wEW}\right)\right\}}$$

and $(UA)_{eff2} = (MC)_w\left(1 - e^{-a\Delta t}\right)$ are heat gain and heat loss terms respectively.

Location	Solar Energy Park, I.I.T. Delhi	Quantity of glass	2
Specification of the location	28°35′ N, 77°12′ E, altitude 216 m from mean sea level	Inclination angle	15°
Climate	Composite	Color of walls inside	Black
Orientation	East - West	No of inlet to saline water	1
Body material	Glass reinforced plastic (GRP)	No of outlet connected with trough at ends	2
Basin area	2 m^2	Water depth	0.01 m
Thickness of GRP	0.005 m	Latent heat of vaporization (L)	2390×10^3 J/kg
Height at ends	0.22 m	Specific heat of water (C_w)	4200 J/kgK
Height at centre	0.48 m	Mass of water (M)	20 kg
Simple window glass	$1.03 \times 1.06 \times 0.004$ m^3	Wind velocity (v)	4.5 m/s
K_b	0.0351 W/mK	α'_w	0.5
K_g	0.78 W/mK	ε_g	0.9
α'_b	0.7	ε_w	0.9
α'_g	0.2	ε_{eff}	0.82
U_{bw}	300 W/m K		
Calibrated Thermocouples	copper-constantan (T type) measuring range -200 to 350 °C and sensitivity of 43 μV/°C		
Digital temperature indicator	resolution 0.1°C, range -20 to 450 °C		
Solarimeter	resolution 20 W/m^2, range 0-1200 W/m^2		
Calibrated mercury thermometer	resolution 1°C, range 0-120 °C		
Measuring jar	resolution 10 ml, range 0-100 ml		

Table 1: Design and operational parameters of DSPSS, and description of measuring instruments.

4.0 RESULTS AND DISCUSSION

Fig. 6 shows the comparison of previous (Eq. (6g)) and new thermal model (Eq. (7)) with experimental observations of DSSS for the month of June (date: 13/06/06). A close agreement have been found for both thermal models (previous thermal model and new thermal model with RMS errors 10.08% and 10.12% respectively with experimental results) for a given design and operational parameters (Table 1) over 24 h with experimental values. The RMS errors are at higher sides due to several assumptions and heat losses but with similar variation as experimental values have. Figs. 7 and 8 show the hourly variations of experimental and theoretical values of various temperatures, and amount of distillate respectively obtained in duration 7:00 h to 17:00 h for DSSS at water depth 0.01 m for the month of June'06. A fair agreement in temperature of water with root mean percentage square (RMS) error 6.8% (coefficient of correlation r=0.9671) has been found. Similarly, RMS error for total amount of yield obtained experimentally and theoretically from both east and west sides of DSSS has been found 34.7% with r=0.9753, as shown in Fig. 8.

Further for plotting the characteristic curves based on Eqs. (9) and (10), average values of various temperatures, solar radiation and yield for any time (e.g. 10:00 h) over a year considering all months have been calculated. Variable nature of climatic parameters (which are non-linear in nature) like solar intensity, ambient temperature and wind speed give results un-realistic values of instantaneous gain/loss efficiencies such as $\eta_i, \eta_{iL} > 1$ and $\eta_i, \eta_{iL} < 0$ especially for low solar intensities, low water temperature in the morning and cooling of water takes place in the evening to next day's morning. Hence, the quasi steady state for DSSS has been taken during the mid-day sunshine hours only i.e. 10:00 h to 14:00 h [21-24]. It is a high solar intensity period in which the 'average value of solar radiation' can be seen nearly equal to the 'hourly variation of solar radiation of the period'.

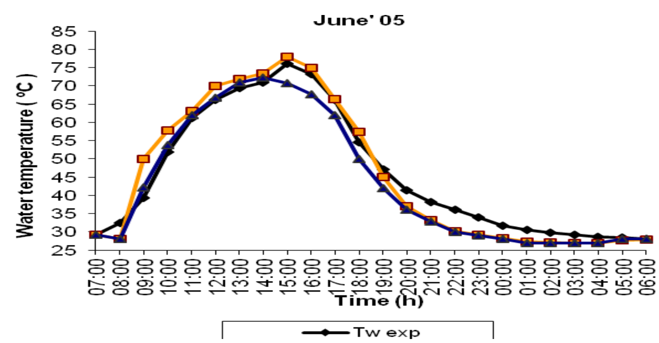


Figure 6: Comparison of old and new thermal model with experimental observations of DSSS for the month of June (date: 13/06/06).

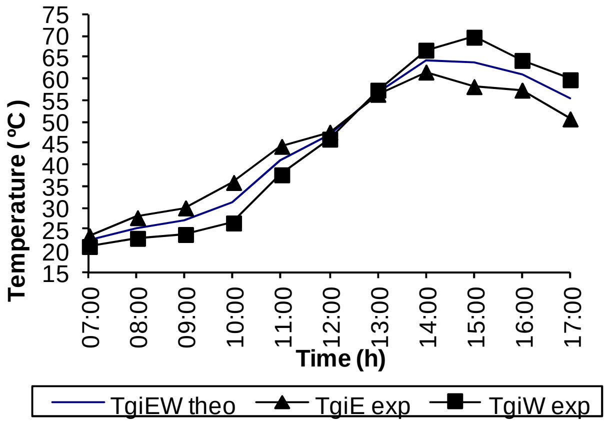


Figure 7: Variation of theoretical and experimental values of temperatures of inner surfaces of both glass covers for the month of June (date: 13/06/06).

Figs. 9 and 10 show linear and non-linear characteristics with their equations for DSSS for the month of Dec'05. The valid range of $x=\left\{\left(T_{wEW}-T_a\right)/\left(I_s(t)_E+I_s(t)_W\right)\right\}$ including threshold and maximum values has been found $0.0069 \le x \le 0.0645$ from the Figs. 9 and 10. Similarly, linear and non-linear characteristics with their equations for DSSS have been obtained for the month of Jun'06. The valid range of x including threshold and maximum values has been found $0.0137 \le x \le 0.0229$. Following are the instantaneous gain and loss efficiency equations obtained for the month of June'06.

for instantaneous gain efficiency: y_{1exp}=18.348x-0.2785 for (R^2=0.983, RMSe=45.4, x_{thr}=0.0152)

y_{1exp}=1226.2x^2-28.355x+0.1582 for (R^2=0.9989, RMSe=1.6, x_{thr}=0.0137)

for instantaneous loss efficiency: y_{1exp}=-16.338x-0.3933 for (R^2=0.9175, RMSe=30.4, x_{thr}=0.0241)

y_{1exp}=-2331.2x^2-72.45x-0.4369 for (R^2=0.9848, RMSe=15.8, x_{thr}=0.0229)

The linear and non-linear characteristic curves based on an annual experimental performance of DSSS has been analyzed. The corresponding equations of these characteristic curves are given as follows, for instantaneous gain efficiency: y_{1exp}=19.621x-0.1458 for (R^2=0.9856, RMSe=25.6, x_{thr}=0.0074) & y_{1exp}=18.26x^2+19.048x-0.1418 for (R^2=0.9857, RMSe=24.1, x_{thr}=0.0074)

for instantaneous loss efficiency: y_{1exp}=-12.478x+0.3392 for (R^2=0.8591, RMSe=15.5, x_{thr}=0.0272) &

y_{1exp}=1013.3x^2-44.278x+0.5617 for (R^2=0.9925, RMSe=5.4, x_{thr}=0.0218)

The valid range of x for linear characteristic curves (including both instantaneous gain and loss efficiency curves) is found to be $0.0074 \le x \le 0.0272$ ºC/W-m^2. Similarly, for non-linear characteristic curves the valid range of x is found to be $0.0074 \le x \le 0.0218$ ºC/W-m^2.

Considering, all linear and non-linear characteristic curves of DSSS, the actual valid range of x is established as $0.0069 \le x \le 0.0229$ ºC/W-m^2. It is seen that non-linear characteristic curves have lower values of RMS errors with higher values of coefficient of determination (R^2) to their corresponding linear characteristic curves. Both 'RMS error' and 'coefficient of determination' are subjected to various parameters (climatic, operational and design) and procedure of measurement (i.e. time interval between two consecutive measurements) in case of solar stills. It can be concluded that to increase the accuracies of characteristic equations, it is required to reduce the heat loss, vapor loss by using appropriate materials for insulation and packing respectively, and also the time interval between two consecutive observations. On the basis of these observations, non-linear characteristic curves (due to higher accuracy) are found better analytical tools for thermal testing of DSSS for different materials, operational and climatic parameters. In other words, it is because solar stills can be made of fiber reinforced plastic/glass reinforced plastic, Galvanized iron sheet with proper insulation, wood, concrete etc. All these materials have different thermal conductivities and material thickness

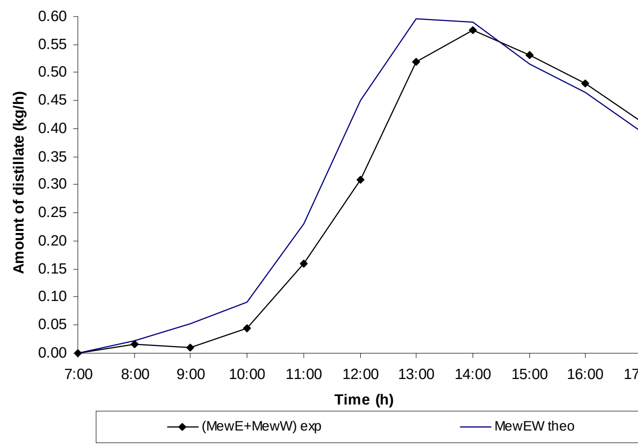


Figure 8: Theoretical and experimental variations of total amount of distillate collected from east and west sides of DSSS for the month of June (date: 13/06/06).

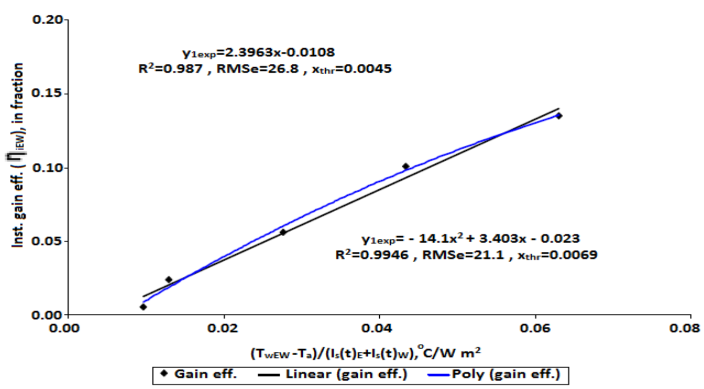


Figure 9: Linear and non-linear gain characteristic curves for DSSS for the month of Dec'05.

Hence, on the basis of use of different materials different characteristics of DSSS can also be obtained and compared for better material and its thickness with optimized cost of fabrication (the cost of fabrication is dependent upon material).

Similarly, operational parameters such as water depth, salinity would also give different characteristics of DSSS which can be compared to get optimized operational parameters. The climatic parameters depend upon the latitude of the place. Hence, for different locations different characteristics can be obtained. The established results for characteristic equations of DSSS is similar to previously established results for single slope solar still [21], hybrid (PV-T) active solar still [22], double slope solar still [23], and inverted absorber solar still [24].

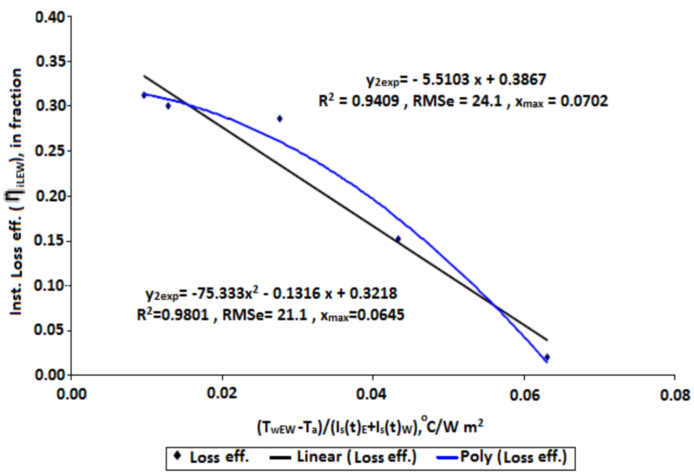


Figure 10: Linear and non-linear loss characteristic curves for DSSS for the month of Dec'05.

CONCLUSIONS

On the basis of above analysis, the new thermal model of DSSS (developed after considering temperatures of east and west side glass covers equal to their average temperature) has been found in fair agreement with experimental observations. Further, the characteristic equations based on the thermal model and experimental data have been obtained. Non-linear characteristic curves have been found more accurate for DSSS in comparison to its linear characteristic curves.

NOMENCLATURE

A_b	Area of the basin (m^2)
A_{gE}	Area of the east side glass cover (m^2)
A_{gW}	Area of the west side glass cover (m^2)
C	specific heat of water (J/Kg °C)
dt	Small time interval (s)
$\frac{dT_w}{dt}$	Change in water temperature in small time dt (°C/s)
h_{ba}	Conductive heat transfer coefficient from basin to ambient (W/m^2 °C)
h_{bw}	Convective heat transfer coefficient from basin to water (W/m^2 °C)
h_{goE}	Combined convective and radiative heat transfer coefficient from inner surface of east glass cover to ambient (W/m^2 °C)
h_{goEW}	Combined convective and radiative heat transfer coefficient from inner surfaces of glass covers to ambient (W/m^2 °C)
h_{goW}	Combined convective and radiative heat transfer coefficient from inner surface of west glass cover to ambient (W/m^2 °C)
h_{kg}	Conductive heat transfer coefficient of glass cover (W/m^2 °C)
h_s	Heat transfer coefficient of side wall (W/m^2 °C)
h_o	total convective and radiative heat transfer coefficient from east/west side glass cover to ambient (W/m^2 °C)
h_{oEW}	total convective and radiative heat transfer coefficient from either east or west side glass cover to ambient (W/m^2 °C)
$I_s(t)_W$	solar intensity incident on the west side glass cover (W/m^2)
K_g	thermal conductivity of glass covers (W/m K)
L_g	thickness of glass covers (m)
L	latent heat of vaporization (J/kg)
M	mass of water in the basin of solar still (kg)
P_w	partial saturated vapor pressure at water temperature (N/m^2)
P_{gi}	partial saturated vapor pressure at inner glass temperature (N/m^2)
$\dot{q}_{ewEW}$	Total evaporative heat gain (W)
T_a	ambient air temperature (°C)
T_b	basin temperature (°C)
T_{giE}	inner glass covers temperature on east side (°C)
T_{giW}	inner glass covers temperature on west side (°C)
T_{giEW}	Average temperature of inner surfaces of glass covers of DSSS (°C)
T_{goE}	outer glass covers temperature on east side (°C)
T_{goW}	outer glass covers temperature on west side (°C)
T_w	water temperature (°C)
T_{wo}	water temperature at time t=0 (°C)
T_{wEW}	Water temperature after assumption of average glass cover temperature (°C)
U_{EW}	radiative heat transfer coefficient between east and west glass cover (W/m^2 °C)
v	wind velocity (m/s)
α'_b	Solar flux absorption factor for basin liner
α'_g	Solar flux absorption factor for glass
α'_w	Solar flux absorption factor for water
α_{eff}	Effective absorptivity of the whole solar still assembly
ε_{eff}	Effective emissivity
ε_g	Emissivity of glass cover
ε_w	Emissivity of water
σ	Stephan-Boltzman constant ($W/m^2 K^4$)
η_{iEW}	Instantaneous gain efficiency (y_1)
η_{iLEW}	Instantaneous loss efficiency (y_2)

h_{1wE} total internal heat transfer coefficient from water to glass cover on east side (W/m^2 ºC)

h_{1wW} total internal heat transfer coefficient from water to glass cover on west side (W/m^2 ºC)

h_{cwE} internal convective heat transfer coefficient from water to glass cover on east side (W/m^2 ºC)

h_{cwEW} internal convective heat transfer coefficient from water to both glass covers (W/m^2 ºC)

h_{cwW} internal convective heat transfer coefficient from water to glass cover on west side (W/m^2 ºC)

h_{ewE} internal evaporative heat transfer coefficient from water to glass cover on east side (W/m^2 ºC)

h_{ewEW} internal evaporative heat transfer coefficient from water to both glass covers (W/m^2 ºC)

h_{ewW} internal evaporative heat transfer coefficient from water to glass cover on west side (W/m^2 ºC)

h_{rwE} internal radiative heat transfer coefficient from water to glass cover on east side (W/m^2 ºC)

h_{rwEW} internal radiative heat transfer coefficient from water to both glass covers (W/m^2 ºC)

h_{rwW} internal radiative heat transfer coefficient from water to glass cover on west side (W/m^2 ºC)

$I_s(t)_E$ solar intensity incident on the east side glass cover (W/m^2)

$I_s(t)_{EW}$ Total solar intensity on both glass covers ($I_s(t)_{EW} = I_s(t)_E + I_s(t)_W$, W/m^2)

REFERENCES

[1]. G.N. Tiwari, A.K. Tiwari, Solar distillation practice for water desalination systems. Anamaya Publishers, New Delhi, India, 2007.

[2]. A. Kaushal, Varun, Solar stills: A review, Renew. Sust. Ener. Rev. 14 (2010) 446-453.

[3]. G.N. Tiwari, H.N. Singh, R. Tripathi, Present status of solar distillation, Sol. Ener. 75 (2003) 367–73.

[4]. M.A.S. Malik, G.N. Tiwari, A. Kumar, M.S. Sodha, Solar distillation, Pergamon Press, Oxford, UK 1982.

[5]. G. Nebbia, G. Menozzi, Acque Dolce Dal Mare, 11 Inchiesta Inter. Milano. 1966; 129-172.

[6]. G.B. Della Porta, Magiae naturalis libri XX. Napoli. 1589.

[7]. R.V. Dunkle, Solar water distillation: The roof type still and multiple effect diffusion solar still, Int. Dev. Heat Trans., ASME proceedings, (part 5): 895-902, 1961.

[8]. P.T. Tsilingiris, Modeling heat and mass transport phenomenoa at higher temperatures in solar distillation systems-The Chilton-Colburn analogy, Sol. Ener. 84 (2010) 308-317.

[9]. E. Rubio, J.L. Fernandez, M.A. Porta-Gandara, Modeling thermal asymmetries in double slope solar stills, Renew. Ener. 29 (2004) 895-906.

[10]. E. Rubio, M.A. Porta, J.L. Fernanadez, Cavity geometry influence on mass flow rate for single and double slope solar stills, Appl. Ther. Engg. 20 (2000) 1105-1111.

[11]. E. Rubio-Cerda, M.A.P. Gandara, J.L.F. Zayas, Thermal performance of the condensing covers in as triangular solar still, Renew. Ener. 27 (2002) 301-308.

[12]. S. Agrawal, G.N. Tiwari, Thermal modeling of a double condensing chamber solar still: an experimental validation, Ener. Conver. Manage. 40 (1999) 97-114.

[13]. A. Omri, J. Orfi, S.B. Nasrallah, Natrual convection effects in solar stills, Desalination 183 (2005) 173-178.

[14]. V.K. Dwivedi, G.N. Tiwari, Annual energy and exergy analysis of single and double slope passive solar stills, Appl. Sci. Res. 3(3) (2008) 225-241.

[15]. V.K. Dwivedi, G.N. Tiwari, Thermal modeling and carbon credit earned of a double slope passive solar still, Desalination and Water Treatment 13 (2010) 400-410.

[16]. V.K. Dwivedi, G.N. Tiwari, Experimental validation of thermal model of a double slope active solar still under natural circulation mode, Desalination 250 (2010) 49-55.

[17]. P.I. Cooper, Maximum efficiency of a single effect solar still, Sol. Ener. 15 (1973) 205-214.

[18]. A. Tamimi, Performance of a solar still with reflectors and black dye, Solar & Wind Technology 4 (1987) 443-446.

[19]. M. Boukar, A. Hramim, Performance evaluation of a one-sided vertical solar still tested in the desert of Algeria, Desalination 183 (2005) 113-126.

[20]. G.N. Tiwari, M.A. Noor, Characterisation of solar stills, Sol. Ener. 18 (1996) 147-171.

[21]. R. Dev, G.N. Tiwari, Characteristic equation of single slope passive solar still, Desalination 245 (2009) 246-265.

[22]. R. Dev, G.N. Tiwari, Characteristic equation of hybrid (PV-T) active solar still, Desalination 254 (2010) 126-137.

[23]. R. Dev, H.N. Singh, G.N. Tiwari, Charateristic equation of double slope passive solar still, Desalination 267 (2011) 261-266.

[24]. R. Dev, G.N. Tiwari, Characteristic equation of the inverted absorber solar still, Desalination 269 (2010) 67-77.

[25]. V.K. Dwivedi, Performance study of various designs of solar still, PhD Thesis, Centre for Energy Studies, IIT Delhi, New Delhi, India (2008) 67-91.

[26]. R. Dev, G.N. Tiwari, Chap-6: Solar Distillation, In: C. Ray, R. Jain editors, Drinking Water Treatment: Focusing on Appropriate Technology and Sustainability, Springer, New York 1st ed. (2011) 159-210.

[27]. G. Singh, S. Kumar, G.N. Tiwari, Design, fabrication and performance evaluation of a hybrid photovoltaic thermal (PVT) double slope active solar still, Desalination 277 (2011) 399-406.

[28]. R. Dev, Thermal modeling and characteristic equations for passive and active solar stills, PhD Thesis, Centre for Energy Studies, IIT Delhi, New Delhi, India (2012) 23-24.

[29]. R. Dev, G.N. Tiwari, Annual performance of evacuated tubular collector integrated single slope solar still, Desalination and Water Treatment, 41 (2012) 204-223.

[30]. R. Kumar, A. Agarwala, Survey of Energy Computing in the Smart Grid Domain, BIJIT - BVICAM's International Journal of Information Technology, 5 (1) (2013) 525-530.

[31]. D. Kamthania, G. N. Tiwari, Determination of Efficiency of Hybrid Photovoltaic Thermal Air Collectors using Artificial Neural Network Approach for Different PV Technology, BIJIT, 4 (1) (2012) 397-404.

[32]. T. Siddiqui, M. A. Wani, N. A. Khan, Efficiency Metrics, BIJIT, 3 (2) (2011) 377-381.

[33]. D. Kumar, T. C. Aseri, R. B. Patel, EECHDA: Energy Efficient Clustering Hierarchy and Data Accumulation For Sensor Networks, BIJIT, 2(1) (2010) 150-157.

[34]. K. Mudgil, D. Kamthania, Energy Payback Time Calculation for a Building Integrated Semitransparent Thermal (BISPVT) System with Air Duct, BIJIT, 5 (2) (2013) 593-596.

APPENDIX [7, 14, 25]

solar flux absorption factor for glass $\alpha'_g = (1-R_g)\alpha_g$; for water $\alpha'_w = (1-\alpha_g)(1-R_g)(1-R_w)\alpha_w$;

for basin liner $\alpha'_b = \alpha_b(1-\alpha_g)(1-R_g)(1-R_w)(1-\alpha_w)$; and $\varepsilon_{eff} = \left[\left(1/\varepsilon_g\right)+\left(1/\varepsilon_w\right)-1\right]^{-1}$; $h_o = 5.7 + 3.8v$; $h_{kg} = k_g/l_g$; $h_b = k_b/l_b$;

$h_{rwE} = \varepsilon_{eff}.\sigma\left[\left(T_w+273\right)^2+\left(T_{giE}+273\right)^2\right]\left[T_w+T_{giE}+546\right]$; $h_{rwW} = \varepsilon_{eff}.\sigma\left[\left(T_w+273\right)^2+\left(T_{giW}+273\right)^2\right]\left[T_w+T_{giW}+546\right]$;

$h_{goE} = \left[(1/h_{kg})+(1/h_o)\right]^{-1} = h_{goW}$; $h_{ba} = \left[(1/h_b)+(1/h_o)\right]^{-1}$; $h_{cwE} = 0.884\left[\left(T_w-T_{giE}\right)+\left(\left(P_w-P_{giE}\right)(T_w+273)/268.9\times10^3-P_w\right)\right]^{1/3}$;

$P_w = \exp\left[25.317-(5144/(T_w+273.15))\right]$; $h_{cwW} = 0.884\left[\left(T_w-T_{giW}\right)+\dfrac{\left(P_w-P_{giW}\right)\left(T_w+273\right)}{268.9\times10^3-P_w}\right]^{1/3}$; $h_{ewE} = 16.273\times10^{-3}\times h_{cwE}\times\dfrac{\left(P_w-P_{giE}\right)}{\left(T_w-T_{giE}\right)}$;

$h_{ewW} = 16.273\times10^{-3}\times h_{cwW}\times\dfrac{\left(P_w-P_{giW}\right)}{\left(T_w-T_{giW}\right)}$; $P_{giE} = \exp\left[25.317-(5144/(T_{giE}+273.15))\right]$; $P_{giW} = \exp\left[25.317-(5144/(T_{giW}+273.15))\right]$;

$h_{1wE} = h_{cwE}+h_{rwE}+h_{ewE}$; $h_{1wW} = h_{cwW}+h_{rwW}+h_{ewW}$; $U_1 = h_{1wE}A_b + U_{EW}A_{gE} + h_{go}A_{gE}$; $U_2 = h_{1wW}A_b + U_{EW}A_{gW} + h_{go}A_{gW}$

$R_1 = \alpha'_g I_s\left(t\right)_E A_{gE} + h_{go}A_{gE}T_a + h_{1wE}A_bT_w$; $R_2 = \alpha'_g I_s\left(t\right)_W A_{gW} + h_{go}A_{gW}T_a + h_{1wW}A_bT_w$; $A_1 = R_1.U_2 + R_2.U_{EW}.A_{gE}$;

$B_1 = R_1.U_{EW}.A_{gW} + R_2.U_1$; $A_2 = h_{1wE}.A_b.U_2 + h_{1wW}.A_b.U_{EW}.A_{gE}$; $B_2 = h_{1wE}.A_b.U_{EW}.A_{gW} + h_{1wW}.A_b.U_1$; $p = U_1.U_2 - U_{EW}^2.A_{gE}.A_{gW}$

$U_{EW} = 0.034\sigma\left[\left(T_{giE}+273\right)^2+\left(T_{giW}+273\right)^2\right]\left[T_{giE}+T_{giW}+546\right]$; $\left(T_w-T_{giE}\right) = \left((p-A_2).T_w-A_1\right)/p$; $\left(T_w-T_{giW}\right) = \left((p-B_2).T_w-B_1\right)/p$

$\left(T_b-T_w\right) = \left(\alpha'_b.\left\{I_s(t)_E+I_s(t)_W\right\}+h_{ba}.T_a-h_{ba}.T_w\right)/\left(h_{bw}+h_{ba}\right)$; $a = \left\{\dfrac{h_{bw}.h_{ba}.A_b}{h_{bw}+h_{ba}}+\dfrac{h_{1wE}.(p-A_2).A_b+h_{1wW}.(p-B_2).A_b}{p}+h_{sw}.A_{sw}\right\}/(MC)_w$

$f(t) = \left[\left\{\dfrac{\alpha'_b.h_{bw}}{h_{bw}+h_{ba}}\right\}.A_b\left(I_s(t)_E+I_s(t)_W\right)+\left\{\dfrac{h_{bw}.h_{ba}}{h_{bw}+h_{ba}}.A_b+h_{sw}.A_{sw}\right\}.T_a+\left\{\dfrac{(h_{1wE}.A_1+h_{1wW}.B_1).A_b}{p}\right\}\right]/(MC)_w$;

After assuming➔ $T_{giE} = T_{giW} = \left(T_{giE}+T_{giW}\right)/2 = T_{giEW}$; $h_{rwEW} = \varepsilon_{eff}.\sigma\left[\left(T_w+273\right)^2+\left(T_{giEW}+273\right)^2\right]\left[T_w+T_{giEW}+546\right]$;

$h_{cwEW} = 0.884\left[\left(T_w-T_{giEW}\right)+\left(\left(P_w-P_{giEW}\right)\left(T_w+273\right)/(268.9\times10^3-P_w)\right)\right]^{1/3}$; $h_{ewEW} = 16.273\times10^{-3}\times h_{cwEW}\times\left(P_w-P_{giEW}\right)/\left(T_w-T_{giEW}\right)$;

$P_{giEW} = \exp\left[25.317-(5144/(T_{giEW}+273.15))\right]$; $P_w = \exp\left[25.317-(5144/(T_w+273.15))\right]$; $h_{1wEW} = h_{rwEW}+h_{cwEW}+h_{ewEW}$

Video Denoising based on Stationary Wavelet Transform and Center Weighted Median Filter

Soundarya K

Submitted in May, 2013, Accepted in February, 2014

Abstract - Noise removal using wavelet has the characteristic of preserving signal uniqueness even if noise is going to be minimized. Images are getting corrupted by impulse noise during image acquisition and transmission. A new median filter termed as the center weighted median filter (CWMF) in the wavelet coefficient domain combined with stationary wavelet transform (SWT) is proposed for video denoising. This filter iteratively smoothes the noisy wavelet coefficients variances preserving the edge information contained in the large magnitude wavelet coefficients. This Paper deals with uncompressed video of .avi format. The proposed algorithm works well for suppressing Gaussian noise with noise density from 10 to 70% while preserving image details. Simulation results show that higher peak signal to noise ratio can be obtained as compared to other recent image denoising methods.

Index Terms – Denoising, SWT, CWMF, PSNR and MSE

1.0 INTRODUCTION

Many scientific datasets are contaminated with noise, either because of the data acquisition process, or because of naturally occurring phenomena. Preprocessing is the first step in analyzing such datasets [3]. There are several different approaches to denoise images. The main problem faced during diagnosis is the noise introduced due to the consequence of the coherent nature of the image capture. In image processing applications, linear filters tend to blur the edges and do not remove Gaussian and mixed Gaussian impulse noise effectively [4]. Gaussian noise is an additive noise, which degrades image quality that originates from many microscopic diffused reflections leads to discriminate fine detail of the images in diagnostic examinations. Noise represents unwanted information which destroys image quality. The noise usually corrupts images by replacing some of the pixels of the original image with new pixels having luminance values near or equal to the minimum or maximum of the allowable dynamic luminance range. In Gaussian noise each and every pixel of the image gets affected. Image denoising is the method of removing unwanted noise from the corrupted images. Several filters are used for denoising the images. The main goal of the filtering process is to remove noise while preserving edges and image detail information. In general image denoising provides the quality images by increasing the PSNR values.

Denoising is an important task in image processing and analysis and it plays a significant role in modern applications in different fields including medical imaging and preprocessing for computer vision. Image denoising is a technique which removes out noise which is added in the original image. Image quality may get disturbed while capturing, processing and storing the image. Noise is nothing but the real world signals. In images, noise suppression is a particularly delicate task. In this task, noise reduction and the preservation of actual image features are the main focusing parts. Noise filtering can be used as replacing every noisy pixel in the image with a new value depending on the neighbouring region. Filters that are used for the purpose of denoising are broadly divided into two types, linear and nonlinear filter. Linear filters tend to blur the edges and other image details. Also these filters perform poorly on images corrupted by non- Gaussian type of noises. Hence for the removal of impulsive noise, nonlinear filters are used. Image denoising finds applications in fields such as astronomy where the resolution limitations are severe, in medical imaging where the physical requirements for high quality imaging are needed for analyzing images of unique events, and in forensic science where potentially useful photographic evidence is sometimes of extremely bad quality. Conventional median filtering approaches apply the median operation to each pixel unconditionally, that is, without considering whether it is uncorrupted or corrupted. As a result, even the uncorrupted pixels are filtered and this causes image quality degradation [6]. An intuitive solution to overcome this problem is to implement noise detection mechanism prior to filtering; hence, only those pixels identified as "corrupted" would undergo the filtering process, while those identified as "uncorrupted" would remain intact. A fusion technique is proposed to find the best possible solution, so that after denoising PSNR and MSE values of the image are optimal. The proposed method is based on SWT and center weighted median filtering, which exploits the potential features of the combination of both wavelet and center weighted median. This paper only deals with uncompressed video of .avi format. This paper is organized as follows. Section 2 discusses the stationary wavelet transform for denoising Gaussian noise. Section 3 describes the center weighted median filter for denoising noise. The proposed methodology is explained in Section 4. Experimental results

Department of EEE, PSG College of Technology, Coimbatore, India
E-mail: soundhi10@gmail.com

are given in Section 5. Finally Conclusion and reference are discussed in Section 6.

2.0 STATIONARY WAVELET TRANSFORM

A tool for the analysis of transient, non-stationary or time varying phenomena that has energy concentrated in time is a wavelet. The SWT can be obtained by modifying the basic DWT algorithm [1], [2]-[3]. The DWT does not preserve translation invariance due to sub-sampling operations in the pyramidal algorithm. The SWT has been introduced because it preserves the property that a translation of the original signal does not necessarily imply a translation of the corresponding wavelet coefficients. To halve the bandwidth from one level to another level, the SWT utilizes recursively dilated filters instead of subsampling. The stationary wavelet transform (SWT) was introduced in 1996 to make the wavelet decomposition time invariant [5]. This improves the power of wavelet in signal de-noising. This paper exploits the benefits of stationary wavelet transform in suppressing noise at high frequencies and median filter to suppress noise in low frequency bands. The proposed algorithm is divided into two steps. After taking SWT to the noisy image, soft thresholding method is applied to the details subbands; then a transformed image is generated from approximation subband only while the other subbands are made equal to zero, applying inverse SWT to the generated 2-D array, then applying the weighted median filter, to remove the residual noise in the low frequency band. After that the approximation band is returned by applying SWT to the denoised signal, the resulted approximation subband is grouped with the thresholded subbands, applying inverse SWT to obtain the denoised image.

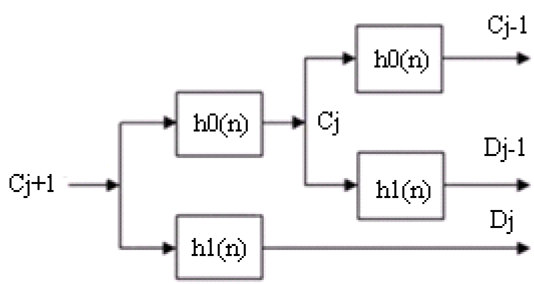


Fig. 1 Wavelet Decomposition

The SWT is an inherent redundant scheme, as each set of coefficients contains the same number of samples as the input. So for a decomposition of N levels, there is a redundancy of 2N. At each level, when the high-pass and low-pass filters are applied to the data, the two new sequences have the same length as the original sequences. To do this, the original data is not decimated. However, the filters at each level are modified by padding them out with zeros. However, for stationary or redundant transform, instead of downsampling, an upsampling procedure is carried out before performing filter convolution at each scale. The Discrete Wavelet Transform (DWT) is the simplest way to implement MRA. It necessitates decimation by a factor 2N, where N stands for the level of decomposition, of the transformed signal at each stage of the decomposition. As a result, DWT is not translation invariant which leads to block artifacts and aliasing during the fusion process between the wavelet coefficients. For this reason, we use the Stationary Wavelet Transform (SWT). For the SWT scheme the output signals at each stage are redundant because there is no signal down- sampling; insertion of zeros between taps of the filters are used instead of decimation.

3.0 CENTER WEIGHTED MEDIAN FILTER

The standard median filter is a simple rank selection filter that attempts to remove impulse noise by changing the luminance value of the center pixel of the filtering window with the median of the luminance values of the pixels contained within the window [7]. Although the median filter is simple and provides a reasonable noise removal performance, it removes thin lines and blurs image details even at low noise densities. Median filter replace every pixel of the image by the median value of its neighborhood. The filter performs well for noise densities less than 50% above which the noise present in the neighborhood is more than the information and hence the filter's performance deteriorates [8]. Fig. 2 describes the function of Median Filter.

3	3	4
4	87	4
4	5	5

Sorted Pixel values: 3, 3, 4, 4, 4, 5, 5, 87
Median Filtered value: 4

Fig. 2 Median Filter

3.1 Median Filter

It provides better removal of impulse noise from corrupted images. It replaces the center value with the median of all the pixels in the window. According to window size each pixels in the window is taken and sort the pixel then find the median of the each window. Then this median value is replaced by center value.

3.2 Weighted Median Filter

The basic idea is to give weight to the each pixel. Every pixel is given a weight. This weight is multiply with pixel. According to this weight the pixels are sort into ascending order, and then find the median value from the sorted list. This value is replaced with center value.

3.3 Center Weighted Median Filter

In CWM center pixel of window is considered as test pixel. Check that the center pixel is less than minimum value available in window and center pixel is greater than maximum value available in window then center pixel is considered as corrupted pixel [10]. This corrupted pixel is replaced by calculated value. Weight is given to the center pixel then sorts all element of window in ascending order and calculate median of elements. The weighted median filter and the center weighted median filter are modified median filters that give more weight to the appropriate pixels of the filtering window. These filters have been proposed to avoid the inherent drawbacks of the standard median filter by controlling the tradeoff between the noise suppression and detail preservation. When one give more weight to the central value of the window a special case of weighted median filters called the Center Weighted Median filter will be produced, and thus it is easier to design and implement the general weighted median filters [9]. Obviously, a center weighted median filter with a larger central weight performs better in detail preservation than with a smaller central weight. The central weight should be carefully selected depending on the characteristics of the input image and its noise. The advantage of center weighted median filter is to reduce noise and to preserve fine details.

4.0 PROPOSED SYSTEM

In the hybrid work, two techniques namely, wavelet thresholding and center weighted median filters are combined to form a hybrid denoising model. The proposed method performs median filtering on the wavelet coefficients to denoise an image degraded by Gaussian Noise. These techniques are used to suppress the Gaussian noise. This Paper only deals with uncompressed video of .avi format. Fig.3 describes the process of noise removal.

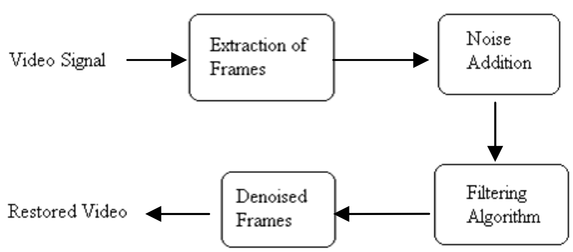


Figure 3: Noise removal process

Wavelets have made quite a splash in the field of image processing. Proposed model is the newly designed hybridized one as shown in fig.4. In this model, the image is denoised first with wavelet decomposition into four sub-bands using haar wavelet filters. It is used for suppressing the Gaussian noise. Resultant coefficients are used for image reconstruction with IWT. The results obtained after inverse transform are then used to reconstruct the image. In the last level, center weighted median filter is used to remove noise present in the image during transformation. The final denoised image is obtained. Wavelets work for decomposing signals into hierarchy of increasing resolutions. The advantage of wavelet denoising is possible to remove the noise with little loss of details. The wavelet mode denoises only the Gaussian type of noise. So when multiple noise present in the image it will remove only Gaussian the remaining noise are unremoved. So for removing the remaining noise and to preserve the fine details CWM filter is applied. The advantage of center weighted median filter is that it can denoise the large window size.

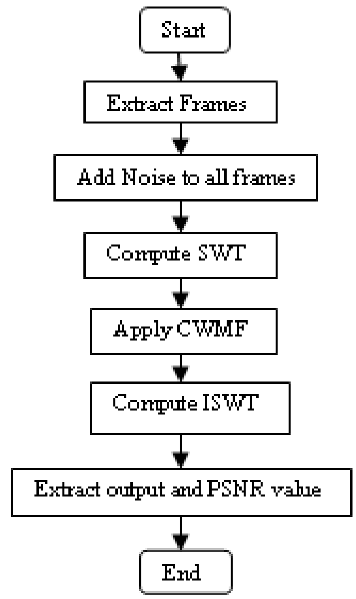


Figure.4: Flow chart of proposed system

The proposed algorithm is described in this section; it is applied for each (Red, Green, and Blue) band separately. It consists of the following steps:

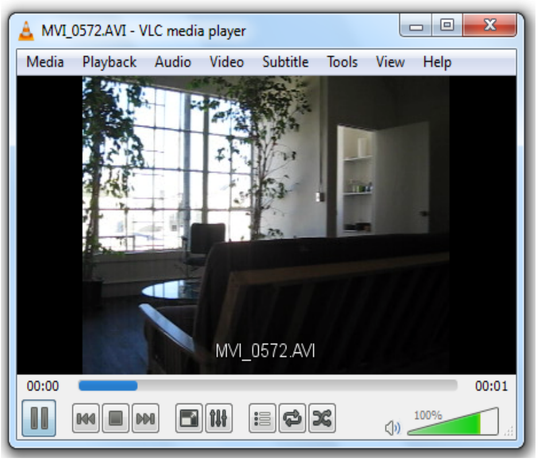
1- Read the noisy color image
2- Apply SWT (one level of decomposition)
3- Reconstruct a new image from LL1 subband only while making other subbands equal to zero, by applying inverse stationary wavelet transform.
4- Apply center weighted median filter to the reconstructed image in spatial domain.
5- Reapplying SWT to the resultant image of step 4.
6-Apply inverse stationary wavelet transform, to get the denoised image.

5.0 RESULTS AND DISCUSSIONS

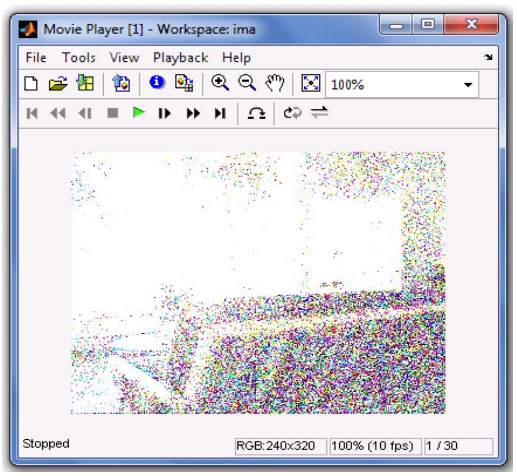
The video used for testing this algorithm is acquired from “http://www.frequency.com/video”. The performance of the proposed algorithm is tested for various levels of noise corruption and compared with the Existing algorithms. The algorithm is implemented in MATLAB 7.2 on a PC equipped with 2.4 GHz CPU and 2 GB RAM memory for the evaluation of computation time of algorithms. The Quantitative performance of the proposed algorithm is evaluated frame by frame based on Peak signal to noise ratio and Mean Square Error which is given in (1) and (2) respectively.

$$PSNR = \log_{10} 10 \frac{255^2}{MSE} \quad (1)$$

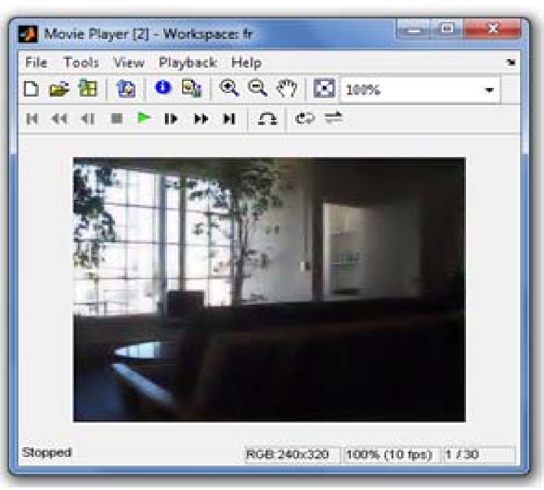
$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(R(i,j) - X(i,j)\right)^2 \qquad (2)$$



(a)



(b)



(c)

Figure 5 (a) Input Video (b) Video corrupted by random noise with 30 Percent noise density and (c) Restored Video

Noise Density (%)	Existing System		Proposed System PSNR (db)
	DTCWT PSNR (db)	DDDTCWT PSNR (db)	
10	34.3432	32.3623	43.0659
20	32.9458	31.5489	40.7863
30	29.6963	30.4888	38.9385
50	28.5542	28.8194	35.5476
70	28.1860	28.2885	32.4698

Table 1: PSNR for different noise densities

In addition to the visual quality, the performance of the developed algorithm and other standard algorithms are quantitatively measured by the following parameters such as peak signal to noise ratio (PSNR) and Mean square error (MSE). The proposed algorithm shows excellent noise suppression and edge preservation capabilities. The proposed algorithm effectively removes the random noise and preserves edges for low, medium and high detail images. Fig. 5 describes the graph plot between PSNR value and noise density for different algorithms [13].

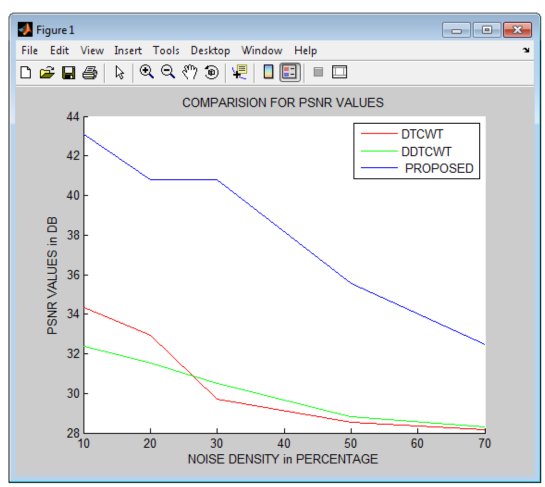


Figure 6: Graph between Existing and Proposed system

The PSNR value of the proposed algorithm is compared against the existing methods by varying the noise density from 10% to 70% and is shown in Table I. From the table, it is clear that the proposed system gives better PSNR values irrespective of the nature of the input image. This algorithm takes 1628 s for computing. Time complexity is higher when compared to other algorithms. But PSNR value obtained is higher.

6.0 CONCLUSIONS

A new technique based on the hybridization of stationary wavelet transform and center weighted median filters for denoising a noisy video. Results prove that utilization of center weighted median filters in combination with stationary wavelet transform deteriorates the performance. By using this technique, the noise is getting reduced and also we can get the originality of the image. The proposed filter gives suitable results on the basis of PSNR and MSE. Experimental results indicate that the proposed method performs significantly better than many other existing techniques. The proposed method is simple and easy to implement. As noise increases the proposed methodology works superior than other algorithms.

REFERENCES

[1]. Abdullah Al Jumah, “Denoising of an Image using Discrete Stationary Wavelet Transform and Various Thresholding Techniques”, Journal of Signal and Information Processing, Vol. 4, pp:33-41, 2013.

[2]. Iman M.G. Alwan, “Color Image Denoising using Stationary Wavelet Transform and Adaptive Wiener Filter”, Al-Khwarizmi Engineering Journal, Vol. 8, No. 1,pp:18 -26, 2012.

[3]. S. Ruikar, Andria and D. D. Doye, “Image Denoising using Wavelet Transform,” International Conference on Mechanical and Electrical Technology, Singapore City, 10-12, pp. 509-515, September 2010,

[4]. Mamta Juneja, Rajni Mohana, “An Improved Adaptive Median Filtering Method for Impulse Noise Detection”, International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.

[5]. Rosebell John and V. Karunakaran, “A Survey on Denoising Methods”, IOSR Journal of Engineering, Volume 2, Issue 10, pp: 85-89, October 2012.

[6]. J Umamaheswari and Dr.G.Radhamani, “Hybrid Denoising Method for Removal of Mixed Noise in Medical Images”, International Journal of Advanced Computer Science and Applications, Vol. 3, No. 5, 2012.

[7]. R. Marudhachalam and Gnanambal Ilango, “Fuzzy Center Weighted Hybrid Filtering Techniques for Denoising of Medical Images”, International Journal of Fuzzy Mathematics and Systems, ISSN 2248-9940 Vol. 2, No. 4, pp. 383-390, 2012.

[8]. S.M. Mahbubur Rahman, Md. Kamrul Hasan, “Wavelet-Domain Iterative Center Weighted Median Filter for Image Denoising “, ELSEVIER, Signal Processing 83, 1001 – 1012, 2003.

[9]. Behrooz Ghandeharian, Hadi Sadoghi Yazdi and Faranak Homayouni, “Modified Adaptive Center Weighted Median Filter for suppressing impulsive Noise in Images”, International Journal of Research and Reviews in Applied Sciences ISSN: 2076-734X, EISSN: 2076-7366 Vol. 1, Issue 3, and December 2009.

[10]. Aparna Madhavrao Harale and J.S.Chitode, “Analyze performance of Median Filter and Center Weighted Median Filter for Efficient Removal of Impulse Noise Using ARM”, International Journal of Engineering and Advanced Technology, ISSN: 2249 – 8958, Vol.-2, Issue-4, April 2013.

[11]. B.Chinnarao and M.Madhavilatha, “Improved Image Denoising Algorithm using Dual Tree Complex Wavelet Transform”, International Journal of Computer Applications (0975–8887) Vol. 44–No20, April2012.

[12]. RC Patil, “Image Retrieval using 2D Dual Tree Discrete Wavelet Transform”, International Journal of Computer Applications (0975– 8887), Vol. 14 ,No.6, February2011.

[13]. Mr. R. K. Sarawale, Dr. Mrs. S.R. Chougule, ”Image Denoising using Dual-Tree Complex DWT and Double-Density Dual-Tree Complex DWT”, International Journal of Advanced Research in Computer Engineering & Technology, Vol. 2, Issue 6, June 2013.

[14]. S. K. Muttoo, Sushil Kumar, "Robust Source Coding Steganographic Technique Using Wavelet Transforms", BIJIT - BVICAM's International Journal of Information Technology, Vol.1 No.2, July - December, 2009.

[15]. S. K. Muttoo, Sushil Kumar, "Data Hiding in JPEG Images", BIJIT - BVICAM's International Journal of Information Technology, Vol.1, No. 1, Jan 2009.

[16]. Sushil Kumar and S. K. Muttoo, "A Reversible Image Steganographic Algorithm Based on Slantlet Transform", BIJIT - BVICAM's International Journal of Information Technology, Vol.5 No.1, January – June, 2013.

BIJIT - BVICAM's International Journal of Information Technology

(A Half Yearly Publication; ISSN 0973 - 5658)

Subscription Rates (Revised w.e.f. January, 2012)

Category	1 Year		3 Years	
	India	**Abroad**	**India**	**Abroad**
Companies	Rs. 1000	US $ 45	Rs. 2500	US $ 120
Institution	Rs. 800	US $ 40	Rs. 1600	US $ 100
Individuals	Rs. 600	US $ 30	Rs. 1200	US $ 075
Students	Rs. 250	US $ 25	Rs. 750	US $ 050
Single Copy	Rs. 500	US $ 25	-	-

Subscription Order Form

Please find attached herewith Demand Draft No.________________ dated ___________________

For Rs.______________ drawn on ____Bank

in favor of **Director, "Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi"** for a period of 01 Year / 03 Years

Subscription Details

Name and Designation ____

Organization ____

Mailing Address ____

____ PIN/ZIP _____________________________

Phone (with STD/ISD Code)______________________________FAX___________________

E-Mail (in Capital Letters)____

Date: **Signature**

Place: (with official seal)

Filled in Subscription Order Form along with the required Demand Draft should be sent to the following address:-

Prof. M. N. Hoda
Editor-in- Chief, BIJIT
Director, Bharati Vidyapeeth's
Institute of Computer Applications & Management (BVICAM)
A-4, Paschim Vihar, Rohtak Road, New Delhi-110063 (INDIA).
Tel.: +91 – 11 – 25275055 Fax: +91 – 11 – 25255056 E-Mail: bijit@bvicam.ac.in
Visit us at: www.bvicam.ac.in/bijit

Announcement and Call for Papers

INDIACom-2015

9th INDIACom; 2015 2nd International Conference on
Computing for Sustainable Global Development
(11th-13th March, 2015)
IEEE Conference Record Number # 35071

BHARATI VIDYAPEETH PUNE

Bharati Vidyapeeth's
Institute of Computer Applications & Management (BVICAM)
A-4, Paschim Vihar, Rohtak Road, New Delhi-63 (INDIA)

Technically Sponsored by

IEEE Delhi Section

Supported by

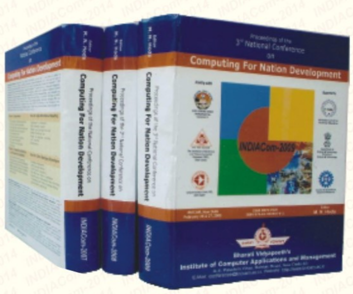
GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY

The Institution of Electronics and Telecommunication Engineers (IETE), Delhi Centre

IET The Institution of Engineering and Technology Delhi Local Networks

CSI Region-I & CSI Divisions- I, II, III, IV & V

ISTE, Delhi Section



(Copies of the proceedings of past *INDIAComs*)

Correspondence

All correspondences related to the conference must be sent to the address:-

Prof. M. N. Hoda
General Chair, *INDIACom - 2015*
Director, BVICAM, A-4, Paschim Vihar, New Delhi -63 (INDIA)
Tel.: 91-11-25275055, TeleFax:91-11-25255056, 09212022066 (Mobile)
E-Mails: conference@bvicam.ac.in, indiacom2015@gmail.com
visit us at: http://www.bvicam.ac.in/indiacom

Paper Submission Link : http://www.bvicam.ac.in/indiacom/ SubmitPaper.asp.

INDIACom-2015 is aimed to invite original research papers in the field of, primarily, Computer Science and Information Technology and, generally, all interdisciplinary streams of Engineering Sciences, having central focus on sustainable computing applications, which may be of use in enhancing the quality of life and contribute effectively to realize the nations' vision of sustainable inclusive development using Computing. it is an amalgamation of four different international conferences, organized as parallel tracks. These are listed below:-

Track #1: International Conference on Sustainable Computing (ICSC-2015)
Track #2: International Conference on High Performance Computing (ICHPC-2015)
Track #3: International Conference on High Speed Networking & Information Security (ICHNIS-2015)
Track #4: International Conference on Software Engineering & Emerging Technologies (ICSEET-2015)

INDIACom-2015 will be held at **Bharati Vidyapeeth, New Delhi (INDIA)**. The conference will provide a platform for technical exchanges within the research community and will encompass regular paper presentation sessions, special sessions, invited talks, key note addresses, panel discussions and poster exhibitions. In addition, the participants will be treated to a series of cultural activities, receptions and networking to establish new connections and foster everlasting friendship among fellow counterparts.

Full length original and unpublished research papers based on theoretical or experimental contributions related to the following topics, but not limited to, are solicited for presentation and publication in the conference:-

- Algorithms and Computational Mathematics
- Green Technologies and Energy Efficient Systems
- IT for Education, Health & Development
- IT for Environmental Sustainability
- IT for Sustainable Agriculture Development
- IT for Water Resources Management
- IT for Consumers' Right
- IT for Crisis Prevention & Recovery
- IT for Disaster Management and Remote Sensing
- IT for other day to day problems
- E-Governance
- Knowledge Management
- E-Commerce, ERP, CRM & Knowledge Mining
- Technology for Convergence
- Distributed and Cloud Computing
- Parallel, Multi-core and Grid Computing
- Reconfigurable Architectures
- Changing Software Architectural Paradigms
- Programming Practices & Coding Standards
- Software Inspection, Verification & Validation
- Software Sizing and Estimation Techniques
- Agile Technologies
- Artificial Intelligence and Neural Networks
- Computer Vision, Graphics, and Image Processing
- Modelling and Simulation
- Embedded Systems and Robotics
- Human Computer Interaction
- Databases
- Data Mining and Business Intelligence
- Big Data Analytics
- Operating Systems
- Data Communication, Computer Networks and Information Security
- Wireless Networking
- Network Monitoring Tools
- Next Generation Internet
- Mobile Computing
- Entertainment Technologies
- Multimedia Computing
- Information and Collaboration Systems
- Fuzzy and Soft Computing
- Bioinformatics
- Medical Informatics
- Education Informatics
- Computational Finance
- Research Methods for Computing
- Case Studies & Applications

Paper Submission

Authors from across different parts of the world are invited to submit their original papers online at http://www.bvicam.ac.in/indiacom/ SubmitPaper.asp. Only electronic submissions will be considered. Papers submitted through E-mail, as attachment, will not be considered.

Review Process, Publication and Indexing

All the submitted papers shall be doubled blind reviewed, by 03 experts, on the basis of their technical suitability, scope of work, plagiarism, originality, novelty, clarity, completeness, relevance, significance and research contribution. The shortlisted papers will be accepted for presentation and publication in the conference proceedings, having **ISSN 0973–7529** and **ISBN 978-93-80544-14-4** serials. Conference proceedings will also be available in soft copy. All accepted papers, which will be presented in the conference, will be submitted for publication and indexing to **IEEE Xplore.**

Important Dates

Submission of Full Length Paper	**10th November, 2014**	**Paper Acceptance Notification**	**12th January, 2015**
Submission of Camera Ready Copy (CRC) of the Paper	**22nd January, 2015**	**Registration Deadline** (for inclusion of Paper in Proceedings)	**22nd January, 2015**