INTERNATIONAL JOURNAL OF
IMAGE PROCESSING (IJIP)

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

VOLUME 6, ISSUE 1, 2012

EDITED BY
DR. NABEEL TAHIR

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

CSC Publishers, 2012

EDITORIAL PREFACE

The International Journal of Image Processing (IJIP) is an effective medium for interchange of high quality theoretical and applied research in the Image Processing domain from theoretical research to application development. This is the forth issue of volume four of IJIP. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJIP emphasizes on efficient and effective image technologies, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing. IJIP comprehensively cover the system, processing and application aspects of image processing. Some of the important topics are architecture of imaging and vision systems, chemical and spectral sensitization, coding and transmission, generation and display, image processing: coding analysis and recognition, photopolymers, visual inspection etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 6, 2012, IJIP appears in more focused issues. Besides normal publications, IJIP intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

IJIP give an opportunity to scientists, researchers, engineers and vendors from different disciplines of image processing to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in image processing in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to IJIP as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in image processing from around the world are reflected in the IJIP publications.

IJIP editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJIP provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members
International Journal of Image Processing (IJIP)

Dr. E. Sreenivasa Reddy
Vasireddy Venkatadri Institute of Technology
India

Dr. Khalid Mohamed Hosny
Zagazig University
Egypt

Dr. Chin-Feng Lee
Chaoyang University of Technology
Taiwan

Professor Santhosh.P.Mathew
Mahatma Gandhi University
India

Dr Hong (Vicky) Zhao
Univ. of Alberta
Canada

Professor Yongping Zhang
Ningbo University of Technology
China

Assistant Professor Humaira Nisar
University Tunku Abdul Rahman
Malaysia

Dr M.Munir Ahamed Rabbani
Qassim University
India

Dr Yanhui Guo
University of Michigan
United States of America

Associate Professor András Hajdu
University of Debrecen
Hungary

Dr M.Munir Ahamed Rabbani
Qassim University
India

Assistant Professor Ahmed Ayoub
Shaqra University
Egypt

Dr Irwan Prasetya Gunawan
Bakrie University
Indonesia

Assistant Professor Concetto Spampinato
University of Catania
Italy

TABLE OF CONTENTS

Volume 6, Issue 1, February 2012

Pages

Improving Morphology Operation for 2D Hole Filling Algorithm

Mokhtar M. Hasan   *mmwaeli@gmail.com*
Computer Science Department/ Faculty of Science
Banaras Hindu University
Varanasi, 221005, India

Pramod K. Mishra   *pkmisra@gmail.com*
Computer Science Department/ Faculty of Science
Banaras Hindu University
Varanasi, 221005, India

Abstract

Object detection may result with some noises, the correct detecting of such object plays a major role for later recognition steps, the interior noise of the object must be removed, the morphological operations are used successfully for this purpose, these morphological operations are applied for 2D holes filling using dilation operation, in this paper we have enhanced this algorithm to get better and faster version that will reduce the processing time dramatically by using a dynamic marker, we have applied two kind of markers with different structuring elements but same size which is 3x3 and those markers are used according to the structure of the sub-window of the object, the processing time is reduced, and our algorithm reduces this time approximately to one third, the results also enhanced since there are some cases missed by the extant version of morphological operations holes filling algorithm.

Keywords: Morphological Operation, Dilation Morphological, Object Filling, 2D Object Filling, Object Holes, Noise Removal, Scanning 2D Objects.

1. INTRODUCTION

We must start with the formal definition of the hole, cavity, and concavity, the hole has been defined from different views but all referring to the same common ground, we can define the hole as "a background region that is a subset of a object", since the hole region should be a subset of their object that belongs to, the Gonzalez and Woods definition in [1] was verbatim *"background region surrounded by a connected border of foreground pixels"*, the other definition that found in Oxford Dictionary [2] was "*a hollow space in something solid or in the surface of something*", in all of the mentioned definitions; we have to exclude the cup handle from being a hole since it is part of the cup [3] or the area that lies in the middle of a doughnut [3] as well, the 3D definition of the hole can be taken from [3] which is "*break in a surface mesh*".

The cavity and concavity are dedicated for 3D objects since these terms has a more understandable meaning, the Oxford Dictionary [2] for this term verbatim is "*a hole or empty space inside something solid*", so, we have moved to the space view of the hole, the other definition [4] is the cavity is the existence of a hollow [2] in an object, or verbatim from [4] "*a bounded connected component of the background*", while the concavity is considered to be a contour with concave shape [2].

Object scanning may produce some missing portions or missing regions even pixels of the scanned object due to and hence, we address the hole filling algorithms to close up and fill out this missing information.

There are many usages of the holes filling algorithms, the preprocessing steps may add some extra holes [5] to the processed object since the segmentation operation is not always perfect, so,

applying the hole filling will eliminate the original holes and the new added holes as well [5]. Other usage is for Cultural Heritage [6] which has several reasons, the authors in [6] summarize those reasons by the sampling accuracy [6], speed gained from scanning technique [6], and acquiring these object data without touching the object [6] since some objects are very fragile due to their ages.

Other factors that impact upon the producing of the holes are occlusion [5], low reflectance [5] or the original scanned object has some missing portions [5].

Furthermore, in order to produce a best model that fits the original object [7] we need to remove any noises that classified as object region or foreground region.

In this paper we have modify the morphology operation for object filling and we gain a more faster and more enhanced results since there are some missing cases by the original algorithm. This paper is organized as follows: section 2 lists some related work associated before, part 3 gives a review for the meaning of the morphological operations, part 4 shows the other algorithms applied in this field as well as the original version of the algorithm that we have modified, part 5 shows our algorithm, part 6 explains the impact of the selection of the structuring elements that we have made to overcome some shortcoming issues that arose from the original version of the extant algorithm, part 7 shows the experimental results, part 8 shows the speed factor comparisons between the old version and our version, and finally, part 9 gives the conclution.

2. RELATED WORK

Gonzalez and Woods [1] presented two different algorithms for object hole filling, these algorithm use the morphological dilation operation for such purpose, the first algorithm is restricted since an initial pixel within the object hole must be specified, which is more difficult in real time automatic applications that needs the filling to be done completely without human interference.

The other algorithm presented by them was applied by dilation algorithm as well but this time no initial pixel needed to be located. They started with an initial marker which is the complement of the border of the original input image and black pixels elsewhere, and their 3x3 structuring element (SE for shorthand) was all ones, and by applying the algorithm, the border comes and closing the objects as every epoch of the algorithm until there is no more changes in the resulted marker, and by anding logical operation between the complement of the marker and the original image matrix, they got the image holes matrix and they got the filled image by oring logical operation between the last two matrices.

The other approach was applied by us in [7] which depends on finding a connected region around the background pixel to be recognized as hole pixel, any initials or prior assumption have not required, they convolved a diamond filter with every background pixel and a roadmap is created, then they trace this road map to find a connected path which means the background pixel has a connected region surrounds it, if such path existed; then this hole pixel as foreground pixel is considered, otherwise, background pixel is considered .

3. MORPHOLOGY OPERATION: REVIEW

The term of morphology came from the biology branch [1] which means the study of the structure of the animals and plants [1], the same term can applied to image objects for studying the structure of these objects in the image [1], there are two basic and important operations which are the dilation and erosion [1], the Oxford Definition [2] for dilation is "*to become or to make something larger, wider or more open*", and the Oxford Definition [2] for erosion is "*to gradually destroy the surface of something through the action of wind, rain, etc*", and in the image erosion , the object is eroded by the action of filters which is called the SE that has a reference point that dominate the matching with the object window since we have no wind nor rain here.

3.1 Morphology Operation Notations

Let A $\subseteq Z^2$, which is the 2D space of the (x, y), and also let B $\subseteq Z^2$ be the SE which controls the structure of the morphological operations, then for binary images, dilation and erosion can be defined as (1) and (2) respectively:

$$A \oplus B = \{ z \mid z = a . b, \quad a \in A \ \& \ b \in B \} \tag{1}$$

$$A \ominus B = \{ z \mid z = a . b, \quad a \in A \ \& \ b \in B \} \tag{2}$$

Where $\oplus$ and $\ominus$ represent the dilation and erosion operations respectively, and **.** (dot) represents the logical anding operation, Equation (1) says that the output of the dilation operation is set to reference point if there is any hit between object A and SE B, this reference point represents the output value if such hit existed, and for Equation (2); all SE must be included inside the object A to produce the reference point as an output which means fit.

The other term is the reconstruction which may exploit the dilation or erosion to get its process done; we will focus on dilation since it is used in this paper, the formulation for this operation as in (3) which taken from [1]:

$$R_A^D(F) = D_A^{(k)}(F) \tag{3}$$

Where R denotes for reconstruction operation, D for dilation, A is the mask image, F is the marker and k represents the number of repeated times until (4) holds:

$$F^k = F^{(k+1)} \tag{4}$$

We will define no more operations since it is out of the scope of this paper, any more definition can be referred to [1,8].

4. MORPHOLOGY OPERATIONS FOR HOLES FILLING

The formulation of the two morphology filling operation can be formulated as follow:

4.1 Hole-Pixel Initial Algorithm (HPIA)

Consider O = { o_0, o_1, .., o_k | o_i represents the (x, y) pixels of object i, k represents the number of objects in image }, then the initial for this algorithm is a vector P where in (5) and (6) :

$$P = \{I_i \mid i = 1, 2, \dots k\} \tag{5}$$

$$I_i = \{(x_j, y_j) , \quad j \ \square \ \text{number of holes in object i} \} \tag{6}$$

This required having starting points that belongs for each hole in each object of the image, which is very difficult since the real time application for such algorithm cannot go along with this since it requires the human interaction in a very essential step which is all the latter processing depends on. However, the The equation for this algorithm is in (7) taken from [1] which closes the holes of the object after certain number of epochs which means until there is no change in X_k:

$$X_k = (X_{k-1} \oplus B) \cap A^c, \text{ for } k = 1, 2, 3, \dots \tag{7}$$

Where X_0 is the marker with initial points I_i , B is the 3x3 SE with zeros at the corners and ones elsewhere, and A is the original input binary image.

4.2 Border Image Initial Algorithm (BIIA)

The algorithm is more efficient for hole filling and the initial value for this algorithm as in (8) which is taken from [1] since it is their algorithm:

$$F(x,y) = \begin{cases} 1 - A(x,y) \text{ ,if } (x,y) \text{ is a border pixel of } A \\ 0 \qquad\qquad\qquad\qquad\qquad otherwise \end{cases}$$

(8)

Where F is the marker, after that, the algorithm starts the following iteration and repeats (9) (taken from [1]) until there is no more change in the marker:

$$H = [R_{A^c}^D (F)]^c$$

(9)

Where H denotes the final image output. This algorithm also suffers from the speed factor since it starts at the border and after that it will go and expand one pixel at a time from border, many epochs required to find H and this depends on the image dimensions, for a rectangular image with 200x200 pixel, it required maximum of 100 epochs; and in case of the input image has no object the algorithm will continue also wasting the time, and in case of filling the letters of image document, we will notice that this algorithm is not working properly since there are some missing issues.

5. OUR SUGGESTED ALGORITHM

We will call our algorithm as Speed Border Image Initial Algorithm (SBIIA) for shorthand. The marker F that used in the previous algorithm had starting points from at the border, if we can increase the number of starting points more than just border points; the dilation operation speed will be increased, we have two ideas in this situation, first idea is the distribution of the random points all over the F, but in this solution; we have to be careful from choosing the random pixel inside the object hole and the distributed points should belong to exterior area of the objects, so, this solution is not working properly as well, the other solution is by scanning the input image row by row starting from the border and filling the corresponding F pixels as a foreground pixels for those are background pixels in A, after that; the job of the dilation will be just to fill the pixels that are between the objects or outside the object in the folded area. Equation (10) shows the initial of marker F.

$$F(x,y) = \begin{cases} 1 & \text{if there is a path from border to } (x,y) \\ 0 & otherwise \end{cases}$$

(10)

By this initialization, the speed of the BIIA is enhanced to be one third of its present time; the performance also enhanced as well since in case of empty image, our algorithm will iterate just one epoch unlike the BIIA that will waste the time. Figure 1 shows the initial marker by using both of BIIA and SBIIA for a given sample.

The gray cells in Figure 1D are also belonging to marker but to show the extra points unlike Figure 1C, by using this marker as initial, the algorithm will be speeded and the processing time will be dropped down significantly. In other words, the BIIA needs many extra epochs until it stops at the object border which is the aim since the hole area is the one that come from the anding operation between complement of the resulted marker and the complement of the original input image, and less epochs for SBIIA since the marker is more closer to the border of the object.
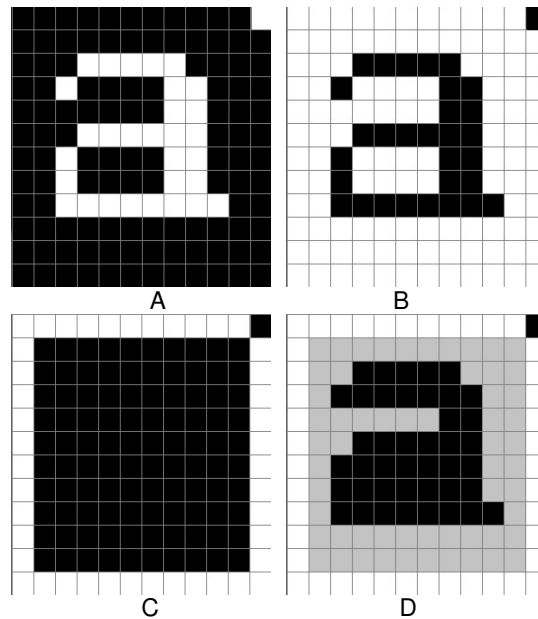
A: the original image, B: the A^c , C: initial marker by BIIA, D: our initial marker by SBIIA.
FIGURE 1: The Initial Marker As Applied By BIIA and SBIIA.

6. STRUCTURING ELEMENT

6.1 SE Operation
Before we penetrate into the selection of the SE, we need to focus some lights on the processing method of the SE, the SE will be matched with the current marker, if there is any hit; the reference point will be produced as an output in the corresponding location of the marker, since we aim to cover all the exterior object area; the input image is first complemented to make the object as background area and background area as a foreground and hence; the searching for the foreground and reconstructing it means the covering for the area that is originally background, after finishing that process, the complement of the marker and the complement of the original input image is anded and the resulting area is the holes which will be ored with the original input image to produce hole-free object as seen by Figure 2.

As seen by Figure 2, the reconstruction operation stops when there will be no change on the produce marker, and this marker is expanded one border pixel at a time, this process continues and stops at a limited number of iteration since the convergence is assured by the expansion process. Furthermore, as noticed by Figure 2 (D-J), this reconstruction operation is obtained by anding the resulted marker at each step with Figure 2B which is the complement of the original input image to ensure this construction operation does need penetrate through the object holes.

The correct selection of the SE plays a crucial role for the successful of the processing operation since it controls the way of expansion or shrinking of the image object, for holes filling, the SE that is used for algorithm in [1] is shown in Figure 3 in which the shaded pixel represents the reference point.
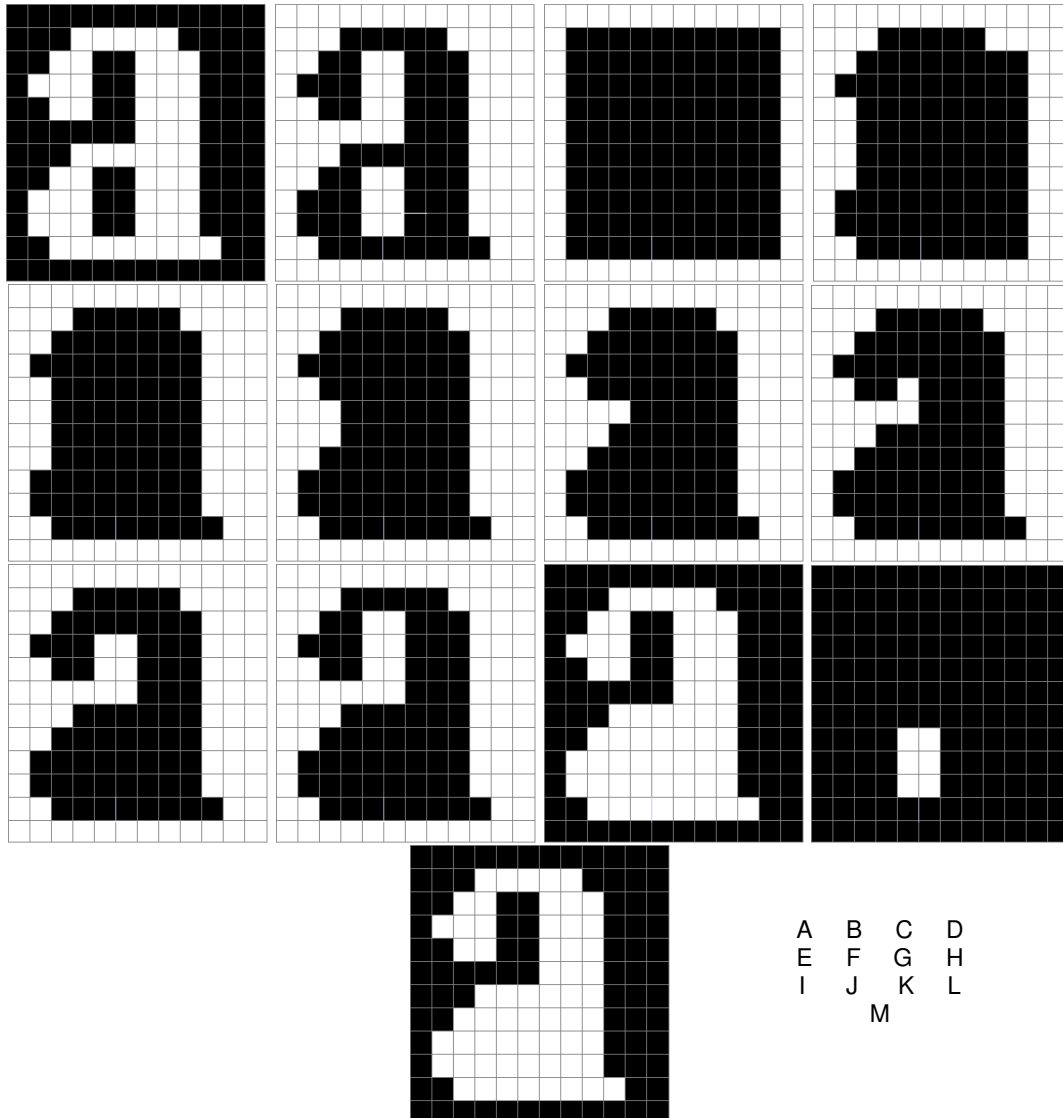
A: the original input image.   B: the complement of A.   C: initial marker as per BIIA.
D: first step of reconstruction operation as per Eq. (9). E: second step of reconstruction.
F: third step of the reconstruction.   G: fourth step of the reconstruction.
H: fifth step of the reconstruction.   I: sixth step of the reconstruction.
J: seventh step of the reconstruction.   K: the complement of the resulted marker in J.
L: the anding operation between K and B.   M: oring operation between L and A.

FIGURE 2: The Application of the Filling Operation.
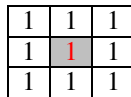
FIGURE 3: The Structuring Element used for Holes Filling.

This SE used here has a disadvantage, which is inability to detect the corner point in case of the edge is thin (one pixel wide), furthermore, this edge is not horizontal nor vertical but diagonal, so, the current version of BIIA continues at this corner edge and expands inside the object which in turn assumes this hole as background since it comes with the trace and the trace should stops at

the object border and surrounds the object interior hole as we mentioned hereinabove, in order to give better understanding for this problem; Figure 4 addresses this issue.

As seen by Figure 4B, the SE hits the corner point at (h) and the reference point will be set to background area and the expansion will go through this area that attached the reference point, which is incorrect since the trace should stops at the border and should not proceed inside the object holes(s), hence, the yellow area as seen by Figure 4A will be considered as a background area and the output image will be the same of the input image without change in this case, so, this is the shortage that infiltrated the performance of the BIIA algorithm.
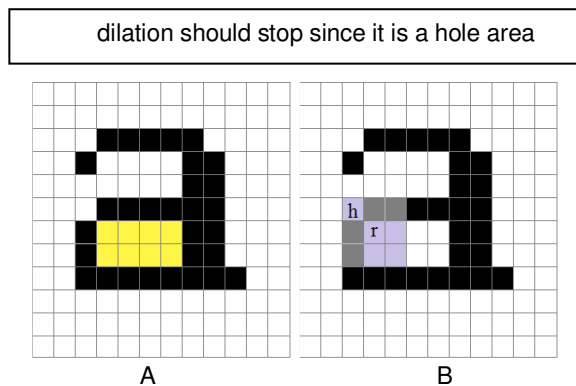
A: complement of 'a', B: same of (A) during process.
FIGURE 4: Addressing the disadvantage of BIIA.

However, our modification resides in the using of two SE's, and their use is managed by a specific threshold after anding with 3x3 image sub-window, Figure 5 shows the SE's used by us.

A: non-threshold SE   B: threshold SE
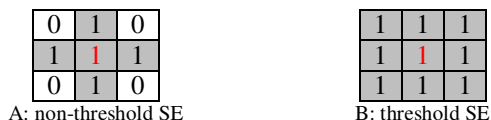
FIGURE 5: Dynamic Structuring Elements Based On Threshold.

6.2 Non-Thresholding SE
Figure 5A is used by us as a general case for dilation operation, Figure 6 shows the application of our SE on Figure4A and we can notice that the problem addressed in Figure4B in not happening here, since, at this corner; there is not infiltrating for the reference point inside the object hole and this problem that addressed by BIIA is recovered here and the area inside letter 'a' will be considered as a hole area and not a background area, since there is no connection between the inner region (yellow region) and outer surrounding one and also since our conditional thresholding SE will not be employed here as we will see in the next section.

6.3 Thresholding SE
The usage of second SE is restricted in case of the summation of overall SE after anding with current sub-window of marker F is equal to a threshold; this threshold is 3, the following Figure 7 illustrates the importance of use such SE in which we can manage the image object having different combinations of corners and non-corners and these corners may be sharp as mentioned in Figure 4 or non-sharp like the left-lower corner at the same latter mentioned figure, normally, we employ the non-thresholding SE in our operation, but, however, if the above stipulation is meet for thresholding SE, this latter SE will be employed to cover the problem of corner points.

dilation stops since it is a hole area

A   B
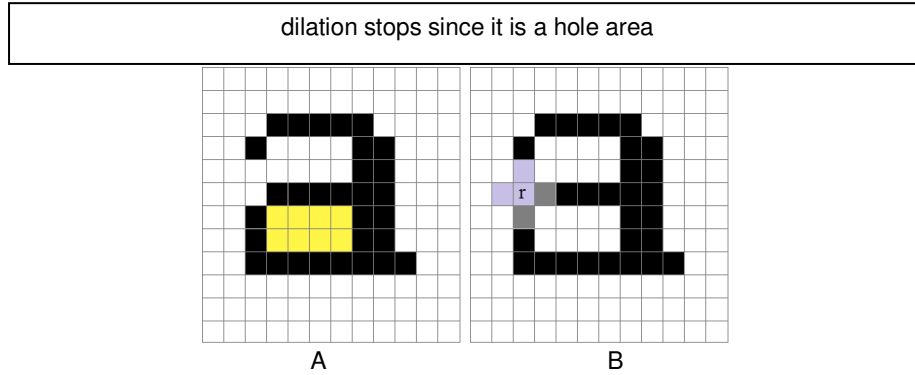
A: complement of 'a', B: farthest hit point which still outside of interior hole
region and there is no connection between inside and outside regions.
FIGURE 6: Our Remedy for Corner Pixel.

dilation should proceed inside since it is a background area

A   B   C
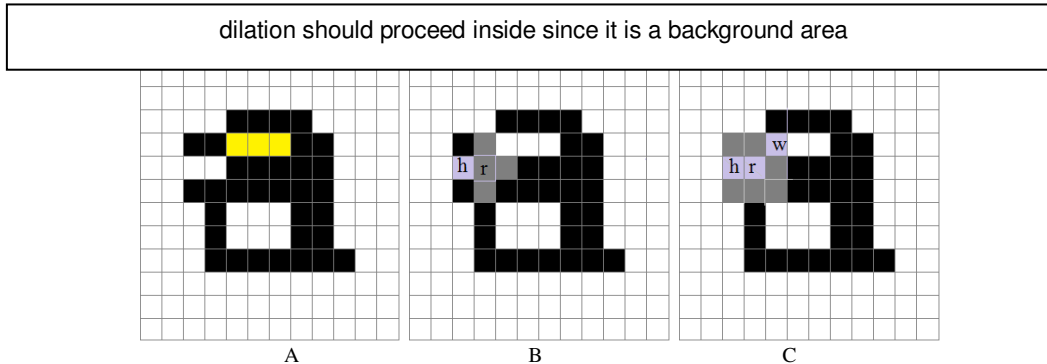
A: complement of letter 'a'. B: during process, non-thresholding SE cannot
penetrate this region.C: we can expand through yellow region starting at point w.
FIGURE 7: Exception Case For Using Of Threshold SE.

As seen by Figure 7B, the yellow area should be attached to a background region, but this region is attached to hole region if we applied the non-thresholding SE since there is no penetration for reference point inside the yellow area, however, Figure7C shows our implementation for the thresholding SE, this thresholding SE is used in case of there is two hit points (including reference point) and one point inside the hole, so, the total is three points out of the 3x3 structure which matches the state in Figure7C, if we compare this issue with Figure 6, we can see that there are three hit points in latter Figure 7 and there are five hit points in Figure 6 (in case of thresholding SE applied), so, the non-thresholding SE will be enough as a general, experiential results for image document will clear this issue.

7. EXPERIMENTAL RESULTS

We have tested our suggested algorithm (SBIIA) with many samples and also with document images; we have gained a good and promising filling and high speed ratio compared to BIIA, and also we have listed a time factor comparison in order to reveal SBIIA's speed. Figure 8 demonstrates the first group of basic sample that will be used as testing objects with their corresponding results after applying BIIA and SBIIA.
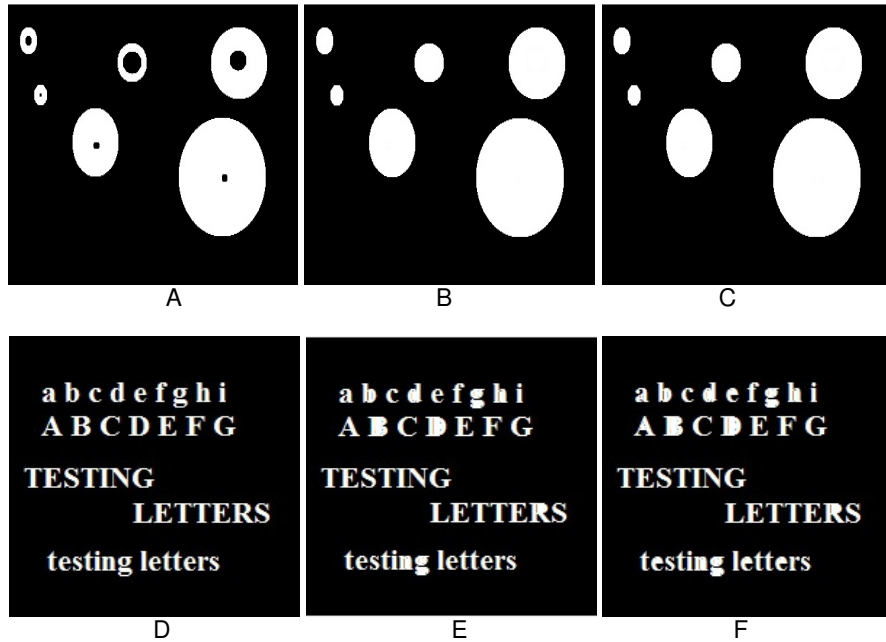
A and D represents the tested samples, B and E is the output of BIIA for each of A and D respectively,
C and F is the output of SBIIA for each of A and D respectively.
FIGURE 8: Testing Output Applied on each of BIIA and SBIIA.

As seen by Figure 8, BIIA is missed 6 letters continuously which are (from top to down, from left to right): a, e, A, e, e, and e, this happened because the problem that we have addressed hereinabove, we made a detailed sketch of letter 'a' and 'e' from above missed letters in order to give a close look for this issue, see Figure 9.
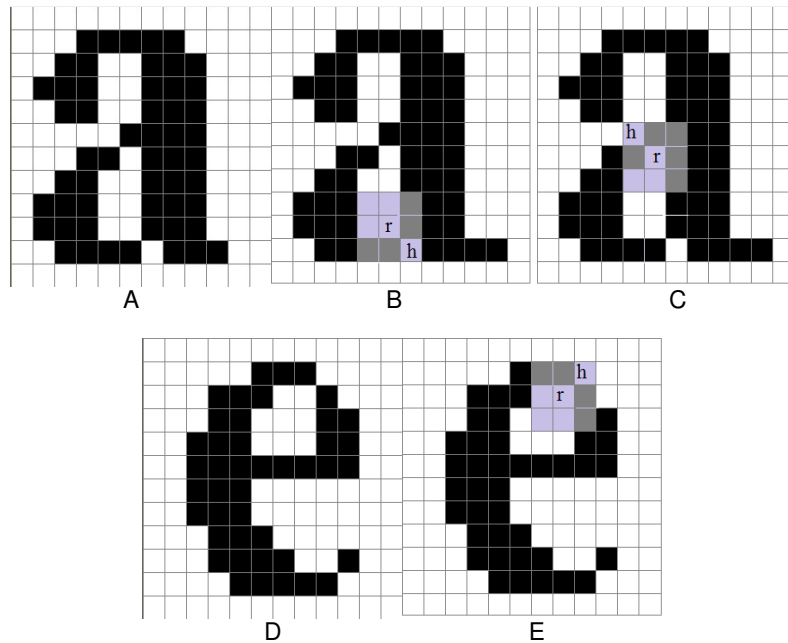
A and D represent the missed letters, B, C and D represent their process.
FIGURE 9: Illustrating the problem of BIIA.

As seen by Figure 9, those letters are taken from the same image and has been drew with pixel wide image for better understanding, however, at the corner, there will be an entrance inside the hole and hence, will be considered as a background region since it comes with the trace, and at final these letters will be the same of input without any change since, as we mentioned before, the trace should stops at the border and should not come inside the object. This problem has been solved by our suggested method. In Figure 10 we have demonstrated one more text image example.
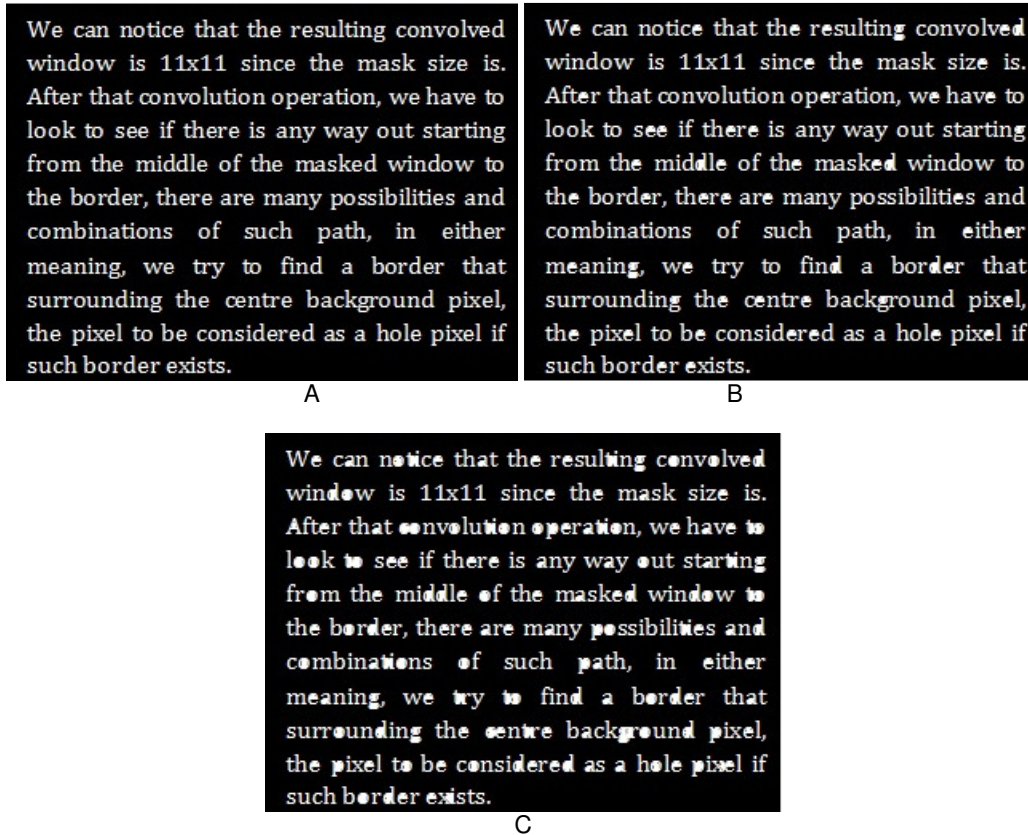
A: original text image, B: BIIA, C:SBIIA.
FIGURE 10: Another Example for Text Image.

Now, by applying our algorithm (SBIIA), Figure 11 shows the remedy of this problem by using of two different SE's as mentioned before.
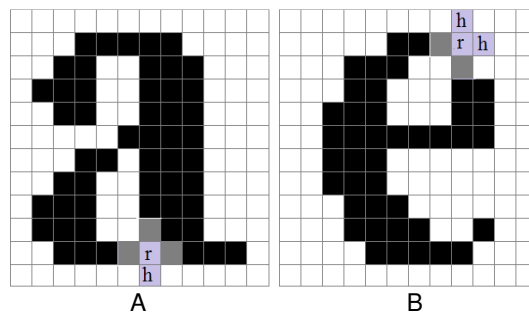
A and B are the output of our algorithm on Figure 9A and Figure 9D respectively.
FIGURE 11: Application of SBIIA on Figure 9 (A and D).

As seen by latter figure, the number of hit points in Figure 11A is 4 points (including the reference point), and 6 points for Figure 11B, so, in both of these cases; the implementation of non-thresholding SE is quite enough, but, however, in other hand, the number of points in Figure 7C is 3 (including the reference point), for that reason we shift to implementation of thresholding SE for discovering and processing this situation and the decision will be applied.

8. PERFORMANCE EVALUATION

Table 1 shows the performance of our algorithm compared with BIIA, we have made a processing time comparison in order to reveal the speed of our algorithm as well as the accuracy as shown before.

Input	Dimension	BIIA (millisecond)	SBIIA (millisecond)	speed ratio %
Figure 7A	50x50	6	3	50
	200x200	484	142	29
	800x800	28528	8790	31
Figure 7D	50x50	13	4	30
	200x200	452	134	29
	800x800	28326	6764	23
Figure 9A	50x50	7	1	14
	200x200	343	89	26
	800x800	27811	5883	21
Blank Image	50x50	7	1	14
	200x200	430	14	3
	800x800	29381	346	1

TABLE 1: Time Factor Performance for BIIA and SBIIA.

As seen by Table 1, the speed up factor has a significant improvement and the overall algorithm has been improved as quality as well as the speed.

9. CONCLUSION

Hole filling algorithm plays a crucial role for object features extraction in order to be applied for recognition stage and the correct features comes from a perfect segmented object, so, segmentation operation may produce such kind of noise along with the existence of the embedded noise in the object, so, by holes filling algorithm we can overcome this problem and obtain a well segmented object, and other applications also for text document and other scanned objects.

We have modified BIIA and we have achieved both better quality and faster version of this algorithm, we have applied a dynamic SE filtering to improve the quantity of the output produced by original version of BIIA, the first SE applied was a general one and applied always which has zero values at the corners and one values elsewhere, the second one is all ones and applied in case of there are two hit points and one missed point out of 3x3 image sub-window which means there is a sharp corner point and special caring should be considered herein, the output of each step is the reference point will be set at the corresponding location in the image object after the above stipulation is been taken into consideration, this technique proved its reliability for filling the letters of the text images as well, the speed factor has been reduced to one third which allows speeding up the overall processing time in which this algorithm is a ring out of a processing chain.

10. REFERENCES

[1]   R. Gonzalez, R. Woods. "Digital Image Processing",Pearson Prentice Hall, First Impression, 2009, pp.665, 682.

[2]   Oxford Dictionary.

[3]   E. Firestone. "An Exploration Of Hole Filling Algorithms", M.Sc. Thesis, Faculty Of California Polytechnic State University, San Luis Obisp, 2008.

[4]   Z. Aktouf, G. Bertrand, and L. Perroton. (2002, March). "A Three Dimensional Holes Closing Algorithm", ELSEVIER Pattern Recognition Letters, Vol. 23 (5), 523–531. Available: 10.1016/S0167-8655(01)00152-0.

[5]   E. Cohen , L. Tekumalla , and E. Cohen. "A Hole-Filling Algorithm for Triangular Meshes", School of Computing, University of Utah, USA, 2004.

[6]   M. Dellepiane, A. Venturi, and R. Scopigno. (2011). "Image guided reconstruction of un-sampled data: a coherent filling for uncomplete Cultural Heritage models", Springer Publication, International Journal of Computer Vision, USA, Vol. 94(1). Available: 10.1007/s11263-010-0382-2.

[7]   M. Hasan, P. Mishra, "Connected Boundary RegionTracing for Object Hole Filling. "International Conference on Advances in Engineering and Technology (ICAET 2011), Bhubaneswar, India, November 2011, pp.67-72.

[8]   M. Droogenbroeck, and H. Talbot. (1996, December 30)."Fast Computationof Morphological Operations with Arbitrary Structuring Elements", Pattern Recognition Letters Journal, USA, Vol. 17(14), 1451-1460. Available: 10.1016/S0167-8655(96)00113-4.

[9]   X. Wu, M. Wang, and B. Han, "An Automatic Hole-Filling Algorithm for Polygon Meshes", Computer-Aided Design and Applications, vol. 5(6):889-899, 2008.

[10] C. Chen, K. Cheng, and H. Liao, "A Sharpness Dependent Approach to 3D Polygon Mesh Hole Filling", Proceedings of EuroGraphics, 2005, pp. 13-16.

[11] A. Kumar, A. Shih, Y. Ito, D. Ross, and B. Soni,"A Hole-filling Algorithm Using Non-uniform Rational B-splines", Department of Mechanical Engineering, University of Alabama at Birmingham, Birmingham, AL, U.S.A., >2008.

[12] K. Oh, S. Yea,and Y. Ho," Hole-Filling Method Using Depth Based In-Painting For View Synthesis in Free Viewpoint Television (FTV) and 3D Video", IEEE 27th conference on Picture Coding Symposium, Chicago, IL, 2009, pp. 1-4.

[13] A. Brunton, S. Wuhrer, C. Shu, P. Bose, and E. Demaine. (2010) "Filling Holes in Triangular Meshes Using Digital Images by Curve Unfolding", International Journal of Shape Modeling, Vol. 16(1-2), 151-171. Available: 10.1142/S0218654310001328.

[14] M. Wei, J. Wu, and M. Pang. (2010). "An Integrated Approach To Filling Holes In Meshes", IEEE International Conference on Artificial Intelligence and Computational Intelligence, 2010, Vol.3, 306 – 310. Available: 10.1109/AICI.2010.302.

Unified Approach with Neural Network for Authentication, Security and Compression of Image: UNICAP

Dattatherya   *dattugujjar28@yahoo.com*
Asst.Prof., Department of TCE
Dyananda Sagar College of Engineering
Bangalore, 500078, India

S. Venkata Chalam   *sv_chalam2003@yahoo.com*
Professor, Department of ECE
CVR College of Engineering
Hyderabad, 501510, India

Manoj Kumar Singh   *mksingh@manuroresearch.com*
Director
Manuro Tech Research
Bangalore, 560097, India

Abstract

The present demands of scientific and social life forced image processing based applications to have a tremendous growth. This growth at the same time has given number of challenges to researcher to meet the desired objectives of either users or from solution perspectives. Among the various challenges, the most dominating areas are: reduction in required memory space for storage or taken transmission time from one location to other, protection of image contents to maintain the privacy and to facilitate the mechanism to identify the malicious modification if there is any, either in storage or in transmission channel. Even though there are number of methods proposed by various researchers and are existed as solutions, questions are remain open in terms of quality, cost and complexity. In this paper we have proposed the concept based on neural network to achieve the quality of compression, protection and authentication all together using the ability of universal approximation by learning, one way property and one to one mapping characteristics correspondingly. With the proposed methods, not only we can authenticate the image but also positions of malicious activity given in the image can be located with high precision. Proposed methods are very efficient in performance as well as carry the features of simplicity and cost effectiveness.

Keywords: Image Compression, Protection, Authentication, Universal Approximation, One-way Property, One to One Mapping, Neural Network.

1. INTRODUCTION

In the past several years there has been an explosive growth in the use of computer networks, whereby digital information such as text, images and other multimedia files are transmitted frequently via the networks. Digital information that is traveling across the networks can potentially be intercepted by someone other than the intended recipient. Due to this digital information such as medical images, intelligence services etc require confidentiality security service. Currently there are several approaches available for protecting digital images; the traditional method of hiding data is to scramble it so that the true message is unknown. Broadly four different approaches for protecting digital images are: compression, digital watermarking, steganography and cryptography. Basically, compression is a process of encoding data to another form by removing all the redundancy that occurs in the data. This encoding technique will change the data into unreadable form as well as reducing the size of the data file. Due to this characteristic, compression is usually employed when transmitting information over the network.

For text file, retrieving back the data that is the process of decompression can be done successfully without any loss of information. However, this is not the case for digital images because most conventional image compression schemes such as Cosine Transforms, Wavelets, and Fractals inevitably produce image distortion or loss of resolution after decompression. These image distortions may include: blurring; visible tile boundaries, introduced image artifacts, and jagged or blurry motion. Further increase in compression will result in worse distortions and image quality can be unacceptable. In security perspective this is not tolerable because the true message and its integrity are lost. Besides this, there are several issues that need to be concerned. Firstly, compression technique employs pattern library for encoding. This means for any group to compress or decompress the information, they must have the pattern library. This raises the issue of distribution. Providing that only authorized groups have the pattern library, then only it can be said that the secrecy of the information is maintained. Another issue is that almost all compression algorithms do not integrate password or key in the process of compression or decompression, this will reduce the security strength of the system. Digital watermarking or also known as digital fingerprinting is another technique that is used for digital image protection. This technique inserts pattern of bits known as signature into a digital image, audio or video file. The signature identifies the image's copyright information such as profile information, or an identification number and it is integrated within digital files as noise or random information that already exists in the file, thereby making the detection and removal of the watermark difficult. Even though digital watermarking technique is meant for copyright protection, it can be extended for hiding digital images instead of signature. Steganography employs the same idea as digital watermarking. Classical steganography concerns with ways of embedding a secret message in a cover or host message such as a video, audio recording or computer code. The embedding is typically parameterized by a key; whereby without knowledge of this key it is difficult for any third party to detect or remove the embedded material. Both digital watermarking and steganography techniques do not randomize the information but instead it hides the digital image under a host image. The main drawback of these two techniques is that it requires another image whereby the size of the host image must be big enough to accommodate all the bits values of the protected digital image. Another technique for securing digital data is cryptography. Unlike steganography, cryptography does not hide the message but instead scrambles the message through an encryption process to produce an unreadable cipher text. The cipher text needs to undergo a process called decryption in order to retrieve back the original message. Likewise as in steganography, these two processes are done based on a specific key value. As an alternative technique for multimedia data especially image, it has been suggested by several researchers to use chaos encryption. In most of the system, the encryption algorithm manipulates the pixels of an image instead of manipulating the bits of the image. Chaotic maps and cryptographic algorithms have some similarities namely sensitivity to a change in initial conditions and parameters, random-like behavior and unstable periodic orbits with long periods. The desired diffusion and confusion properties required in a cryptography algorithm are achieved through the repeated processing. On the contrary, iterations of a chaotic map spread the initial region over the entire phase space. An important limitation of cryptography is that encryption transformations are defined on finite sets.

2. RELATED WORK

In article [1], authors proposed an extension to the block-based image encryption algorithm (BBIE) scheme that works in combination with Blowfish encryption algorithm. Whereas BBIE is meant for 256-color bitmap images, the proposed technique also handles RGB color images and, for the cases studied, improves the security of digital images. In this technique, enhanced block based image encryption technique (EBBIE) the digital image is decomposed into blocks, then two consecutive operations - rotating each 3D true color image block by 90 degree followed by flipping row-wise down - are performed to complicated the relationship between original and processed image. These rendered blocks are then scrambled to form a transformed confused image followed by Blowfish cryptosystem that finally encrypts the image with secret key. [2] Presented a method of chaotic image encryption called the "Triple-Key" method. In this method, it is required to enter an 80-bit session key in addition to the initial parameter key and the control

parameter key. Each of the keys forms just one part of the lock that needs to be opened to obtain the original image. The position of bits in the 80-bit key determines the scrambling of individual pixels in the encrypted image. Low correlation coefficient between adjacent pixels in the encrypted image has achieved, which implies higher security and lower probability of security breach through brute force attacks or statistical analysis. Authors in the paper [3] propose a novel image encryption method based on changing the pixel positions as well as pixel values to confuse the relationship between the cipher image and the plain-image. From the results, it is observed that the proposed technique significantly reduces the correlation among the pixels by shuffling the image matrix using a random vector. Moreover, the scheme has less computational complexity, good security and satisfies the property of confusion and diffusion. Transforming method of original image into the encrypted one using randomly generated vectors has presented in[4]. The original image is decrypted by applying least square approximation techniques on the encrypted image and the randomly generated vectors. In the [5] encryption scheme is proposed based on permutation of the pixels of the image and replacement of the pixel values. The permutation is done by scan patterns generated by the SCAN methodology. The pixel values are replaced using a progressive cellular automata (CA) substitution. The proposed image encryption method satisfies the properties of confusion and diffusion due to the CA substitution. Liao's chaotic neural system, cat map and general cat map are introduced and analyzed respectively in [6].Then, a image encryption scheme by employing the Liao's chaotic system to generate parameters of general cat map is designed. In [7] security of image encryption based on two dimensional chaotic maps has presented. Chaotic maps are often used in encrypting images. However, the encryption has periodicity, no diffusion, and at the same time, the real keys space of encryption is fewer than the theoretical keys space, which consequently results in potential security problems. They have given several ways to solve the problems including adding diffusion mechanism, changing the design of keys and developing a composite encryption system. Authors in [8] presented the concept use of Polynomial Chaotic Maps (PCMs) to provide the security of image. Chaotic maps have good properties such as ergodicity, sensitivity to initial conditions and control parameters, etc. Due to these features, they are good candidate for information encryption. In [9] machine learning based concept has applied by authors for image security. One of the areas of new applications is to design cryptographic systems by using chaotic neural network due to the fact that chaotic systems have several appealing features for information security applications. Encryption algorithm is proposed in [10], which encrypts the plaintext based on alternant of the stream cipher and block cipher. A pseudo-random number is used to control which encryption mode is chosen. Using this algorithm, multiple kinds of files (such as TXT, DOC, WMA, and JPEG) are encrypted and decrypted, and the security of the proposed cryptosystem is analyzed. In [11] a camera signature is extracted from a JPEG image consisting of information about quantization tables, Huffman codes, thumbnails, and exchangeable image file format (EXIF). They have shown that this signature is highly distinct across 1.3 million images spanning 773 different cameras and cell phones. Specifically, 62% of images have a signature that is unique to a single camera, 80% of images have a signature that is shared by three or fewer cameras, and 99% of images have a signature that is unique to a single manufacturer. The signature of Adobe Photoshop is also shown to be unique relative to all 773 cameras. These signatures are simple to extract and offer an efficient method to establish the authenticity of a digital image. Paper [12] proposed an algorithm for image authentication and verification. The algorithm is based on public key. The owner of image extracts the information as watermark information from the image which he wants to transmit to others. The owner embeds the watermark information into the image with public key. Anyone can judge the valid of the watermarked image, and locate the place which has been tampered. In [13] the proposed scheme dynamically generates the watermark using messy models. And, it is embedded inside the image by expanding intra plane difference between any two color planes of images. It is known as intra-plane difference expanding. The proposed scheme is very sensitive to the jittering, geometrical and various filtering attacks. A scheme for JPEG grey scale images is proposed based on a data embedding method that utilizes a secret key and a secret mapping vector in the frequency domain has given on [14]. An encrypted feature vector extracted from the frequency domain is embedded redundantly and invisibly in the marked image. On the receiver side, the feature vector from the received image is derived again and compared against the extracted

watermark to verify the integrity and authenticity.[15] has given a semi-fragile watermarking scheme for color image authentication is proposed based on spatiotemporal chaos and SVD (singular value decomposition). Wavelet transform is applied to watermarking. In contrast to conventional approaches where the watermark is embedded directly on the wavelet coefficients, we embed the watermark onto the SVs (singular values) of the blocks within wavelet sub- band. In order to enhance the security, spatiotemporal chaos is employed to select the embedding positions for each watermark bit as well as for watermark encryption. In [16] this paper proposes an image authentication scheme which detects illegal modifications for image vector quantization (VQ). In the proposed scheme, the index table is divided into non-overlapping index blocks. The authentication data is generated by using the pseudo random sequence. Proposed scheme can adaptively determine both the size of the authentication data and the number of the indices in each index block. Then the selected indices are used to embed the secret data to generate the embedded image.

3. PROPOSED SOLUTION

To capture the quality of compression which also has the protection of itself and a mechanism to define the malicious activity involved with the transmitted or stored image is a difficult task. Even though solution are available for each purpose but problem and limitations are appear at various level of performances like quality, speed etc. along with implementation cost and complexity.

To overcome the issues with existing methods and to improve the performance with respect to three most important requirement compression, security and authentication, when image especially transfer through internet ,rather than applying conventional methods ,intelligent method based on artificial neural network taken as solution platform. There are number of qualities available in ANN which can apply for image processing among them universal approximation, one to one mapping and one-way properties selected to achieve the objectives as shown in Figure1.description of these properties given below.
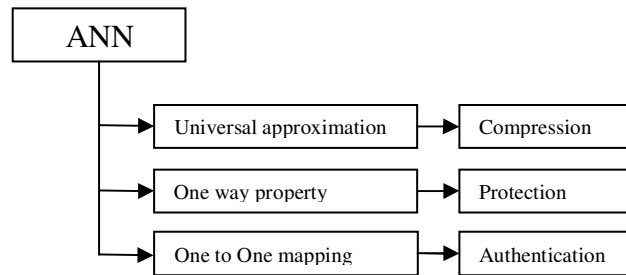
FIGURE 1: ANN properties and its proposed application

(a)Universal approximation
Let F be any Borel measurable or continuous function from $k \subset \Re^n$ on $(0, 1)^m$ and let Φ be any strictly increasing continuous squashing function. Then, for any $\varepsilon > 0$ there exists a multilayer feed forward network N with the quashing function in the output layer and with only one hidden layer such that
$$\| N(x)-F(x)\| < \varepsilon , \quad \forall x \in K \qquad \text{----------------} \quad (1)$$

(b) One to one mapping

If there are interaction of two parameters happen under such environment and circumstances so that resultant outcome could be a unique value, then this unique value can be defined as the one to one mapping between these parameters. Mathematically this can be stated as:
$$\varphi(x_i, \ y) \neq \varphi(x_j, \ y) \ ; \forall j \ if \ j \neq i \ ; \qquad \text{----------------} \quad (2)$$

Where x_i is external stimulus and having an establish relationship with environment φ contains parameter y and x_j is a new test input in the same environment. This principle is also valid if stimulus is same for different parameter available in environment and this can express as

$$φ(x, y_i) \neq φ(x, y_j) \; ; \; \forall \; j \; if \; j \neq i \; ; \qquad \text{----------------} \quad (3)$$

This mapping characteristic can be utilized for authentication and recognition purpose in various applications especially in the field of image recognition where authentication and recognition process cascaded with automatic action as response of recognition.

(C)One-way Property

As it appears by the name a system contains one-way property allow to compute the output from the input easily while very difficult to compute the input from the output. There are two very clear reasons why neural network having one way property
(a) Number of neurons having nonlinear characteristics are involved and interconnected to produce the output.
(b) After learning all previous changes in iterations have lost permanently i.e. no trace available how it has been up to output. Situation will become worse if weights are not available in fact it is impossible to find the input if weights are not available.
         Taking a simple neuron model for example, the input P is composed of n elements [$p_1, p_2, p_3, \ldots p_n$] while the output is a unique element C. It is defined as:

$$C = f(\; \sum_{j=1}^{n} \; w_j p_j \; + b \;) \qquad \text{------------------} \quad (4)$$

As can be seen, it is easy to compute C from P[$p_1, p_2, p_3, \ldots p_n$], while difficult to compute P from C. The difficulty is equal to solve a singular equation. Thus, it is a one-way process from the input P to the output C.

4. IMPLEMENTATION DESCRIPTION

4.1 Preprocessing of Image
Preprocessing is a step which makes the raw data suitable for proper processing. Without which either it is not possible to complete the processing or it may happen with error. From neural network perspectives two different stages required for preprocessing (a) spatial block formation (b) normalization. In spatial block formation image is divided into number of block, each block have a size of m*n pixels, generally m = n. Depends upon the number of pixels in each block input layer neuron number decided. As a rule of thumb about size of block is it should not suppose to too large otherwise it will carry more information which will make difficult to extract local information or should not suppose to be too small otherwise block will not carry proper information about its neighbors. Any moderate size will give a better chance to capture the correlation of pixels in a small local region. Normalization will transfer the pixels value in the range of [0 1] so that it can directly taken as an input for neural network.

4.2 Universal Approximation as a Compression Mechanism
Purpose of compression is to reduce the memory requirement to represent the same image. Fundamentally this is achieved by redundancy available in image. Problem with this quality is redundancy varies from one image to another and in result with same process different compression ratio appears. There are number of application where we required fixed compression ration irrespective of quality of images. To achieve this a feed forward architecture can be taken as solution where hidden layer neuron number are less compare to input layer .For this structure compression ratio directly appear as ratio of number of selected neurons in input layer and hidden layer. Output layer neuron number is same as input layer. With respect to an image for each block training has given to get the learning of correlation available in each block in other word neural network get the approximation knowledge of relationship available in each block. Because there is a large data set available for learning in result after completion of training neural network having capability to define the approximate characteristics available in any type of test image on block basis. Design of compression mechanism has shown in

Figure3-Figure5.Trained output layer weights of neural network called here decompression key which can be passed to receiver only once physically or through secret channel.

4.3   One-way Properties for Security of Compressed Image

As it was stated earlier neural network having characteristics of one-way property. After learning compressed data for each block taken from hidden layer neurons are completely secure. Even though compressed data in channel are integer number but inside the neural network it is transform in to real number which increase level of security further .not only that until the proper weights for decompression are not available it is impossible to decompressed the compressed information. This gives one extra step of security where there are less number of user in group and situation forced them not to revel the others information for example in the case of intelligence services. Even all users are using the same concept of solution but each one will communicate the image with other one only with a proper set of weights which are not available for others.

4.4   One to One Mapping Property for Authentication

Purpose of learning in neural network is to map the input information corresponding to desired output. This property applied here to give the authentication of image with very high precision. With respect to compressed data set one architecture of feed forward network having input layer neuron equal to hidden layer neurons in the compression system and $\lfloor 30\% \rfloor$ of that taken as hidden layer neuron with single neuron in output layer is created. To make the authentication sensitive target of learning given as 0.5.once learning completed for each compressed block generated output is taken as authentication code for that block. Trained weights are taken as authentication key for that particular image and its size is very small even decrease further with higher compression ratio. This authentication key passed to receiver through the secret channel where as authentication code passed through public channel. Detail working process of develop system is shown in Figure2 and algorithmic construction has shown below

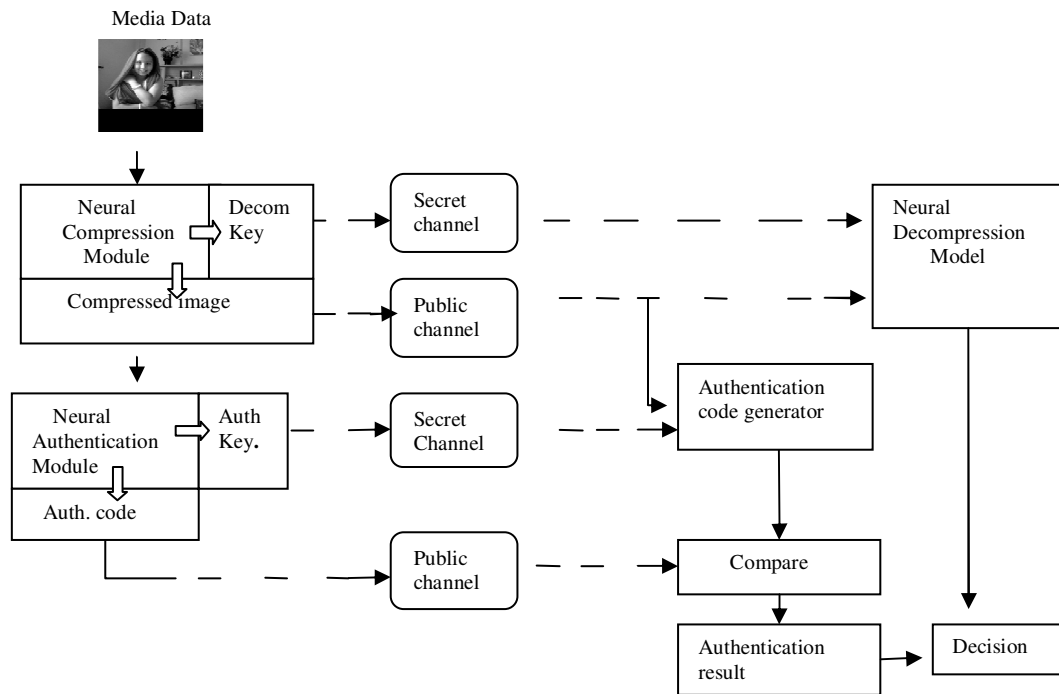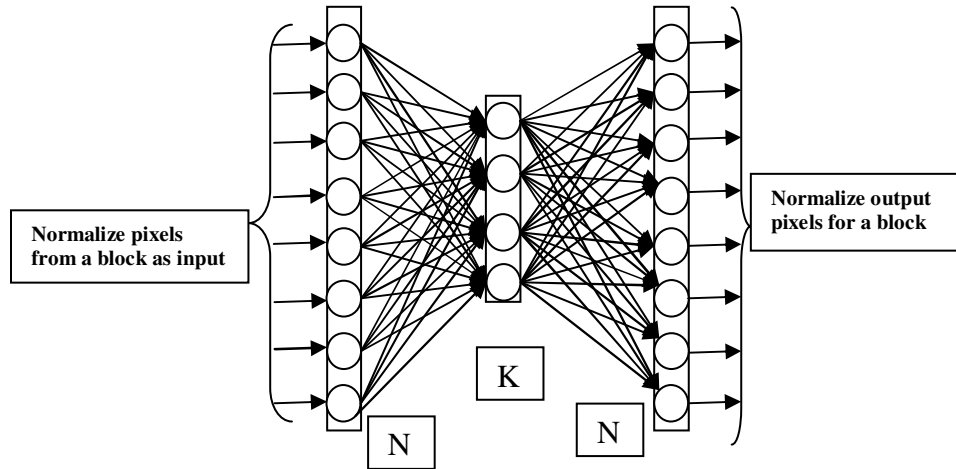**FIGURE 2:** Architecture of the proposed multimedia content UNICAP scheme.

FIGURE 3: Compression architecture of ANN in learning
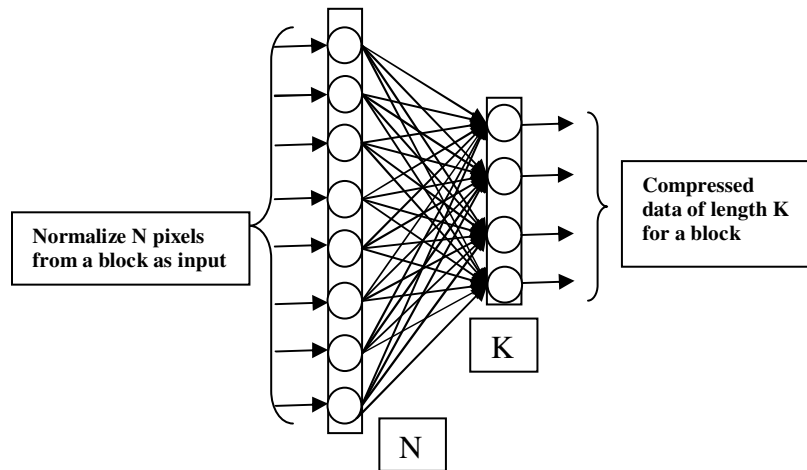
FIGURE 4: Architecture for the compression module
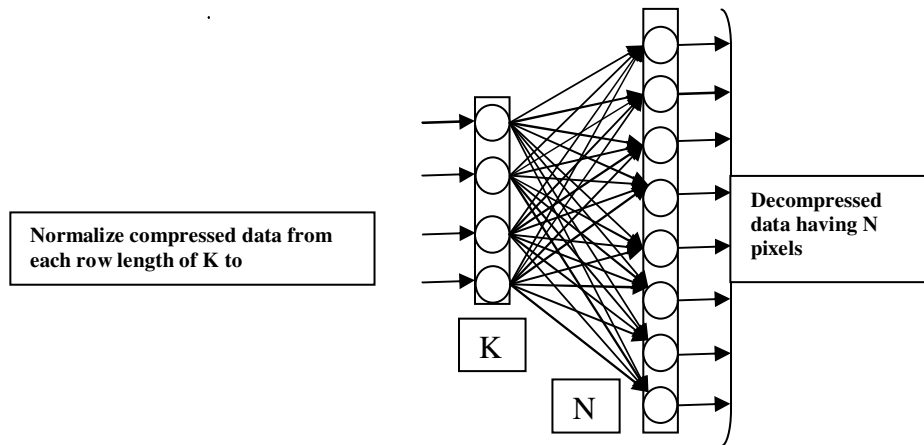
FIGURE 5: Architecture for Decompression module

5. EXPERIMENTS & RESULTS

To get the output with proposed solution for training purpose a gray scale image having size of 'Lena' having 512*512 pixels has taken. Size of block has defined as 8*8 pixels and in result there are 4096 blocks are available. A fully interconnected multi layer feed forward architecture having size [64 64/CR 64] has taken (CR represents compression ratio taken at present 4:1).Bias has also applied for hidden and output layer neuron. Transfer function in hidden and output neurons has taken as unimodel sigmoid function. Back propagation learning algorithm has applied for learning with learning rate equal to 0.1 and momentum rate equal to 0.5 for 500 iterations. Initialization of weights taken as random number derived from uniform distribution in the range of [-1 1].authentication architecture also taken as feed forward architecture and back propagation learning applied for 50 iterations. Performance generated in the experiments for various test images have shown below.

(a)   (b)
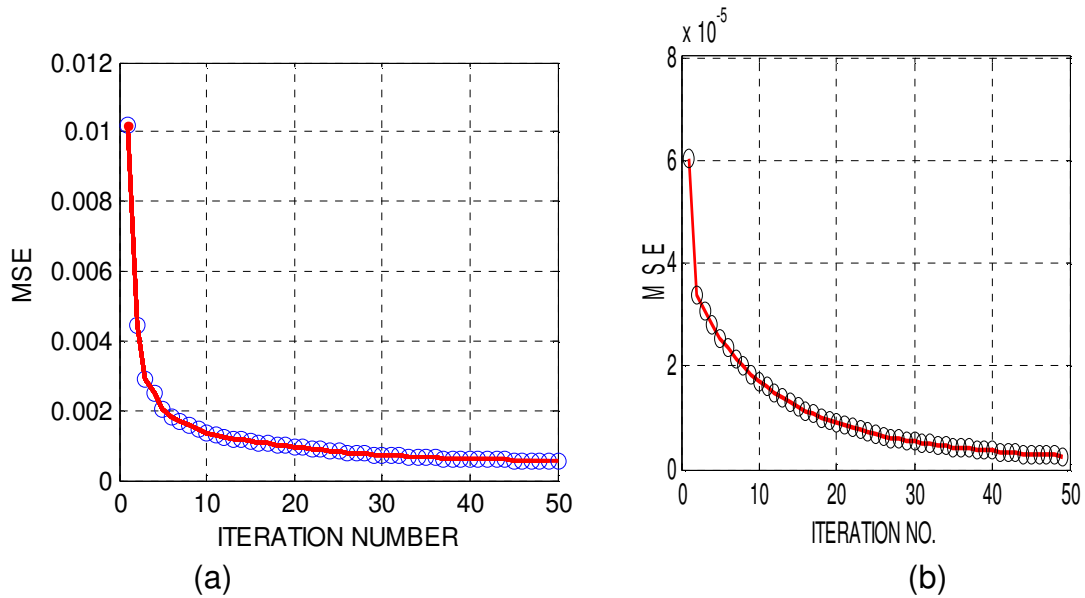
FIGURE 6: UNICAP performance with training image

(a)   (b)   (c)

(d)   (e)   (f)
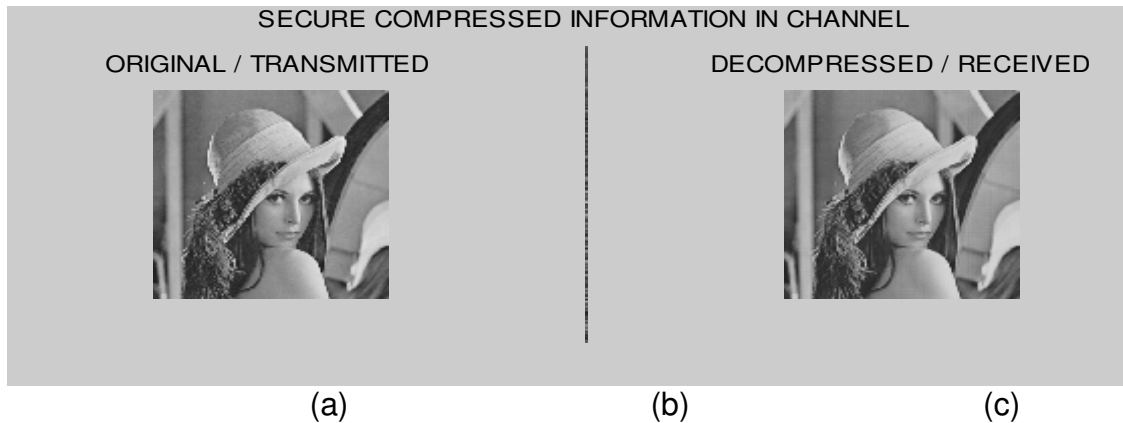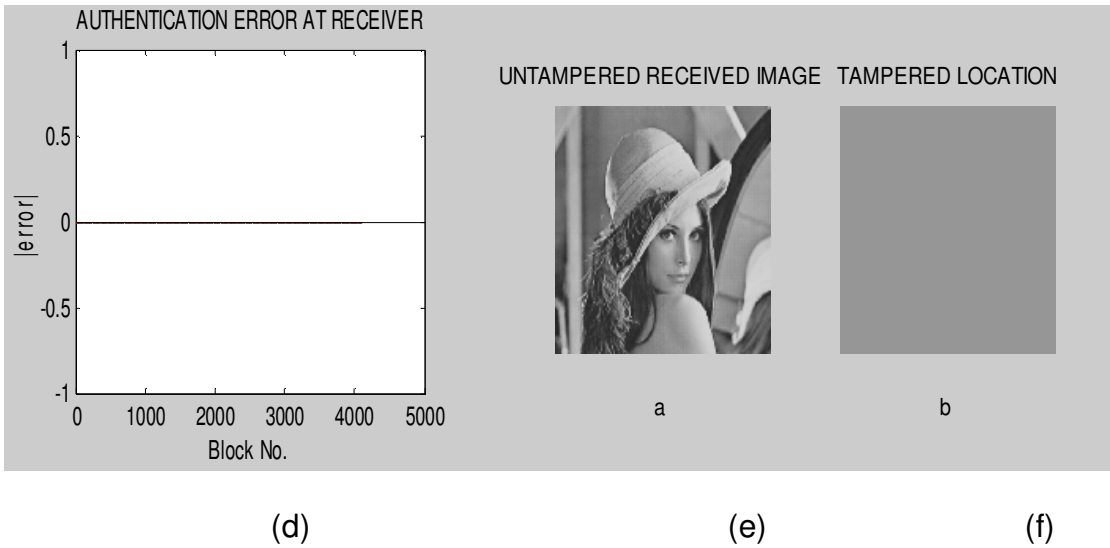
FIGURE 7: UNICAP performance with training image

Tampered case 1:

(a)   (b)   (c)

(d)   (e)   (f)
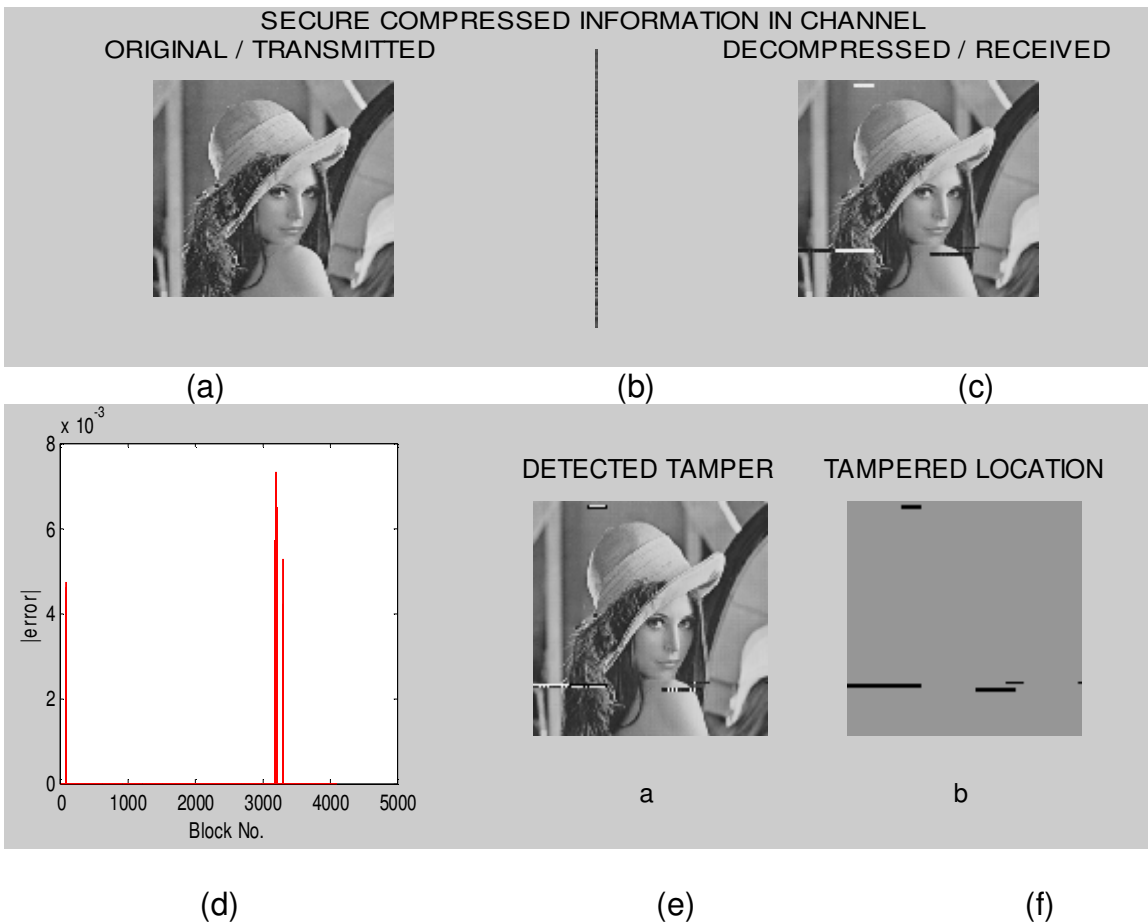
FIGURE 8: UNICAP performance with trained image (bmp format) [a: f]

Tampered case 2:

SECURE COMPRESSED INFORMATION IN CHANNEL
ORIGINAL / TRANSMITTED   DECOMPRESSED / RECEIVED

(a)   (b)   (c)

DETECTED TAMPER   TAMPERED LOCATION

a   b

(d)   (e)   (f)
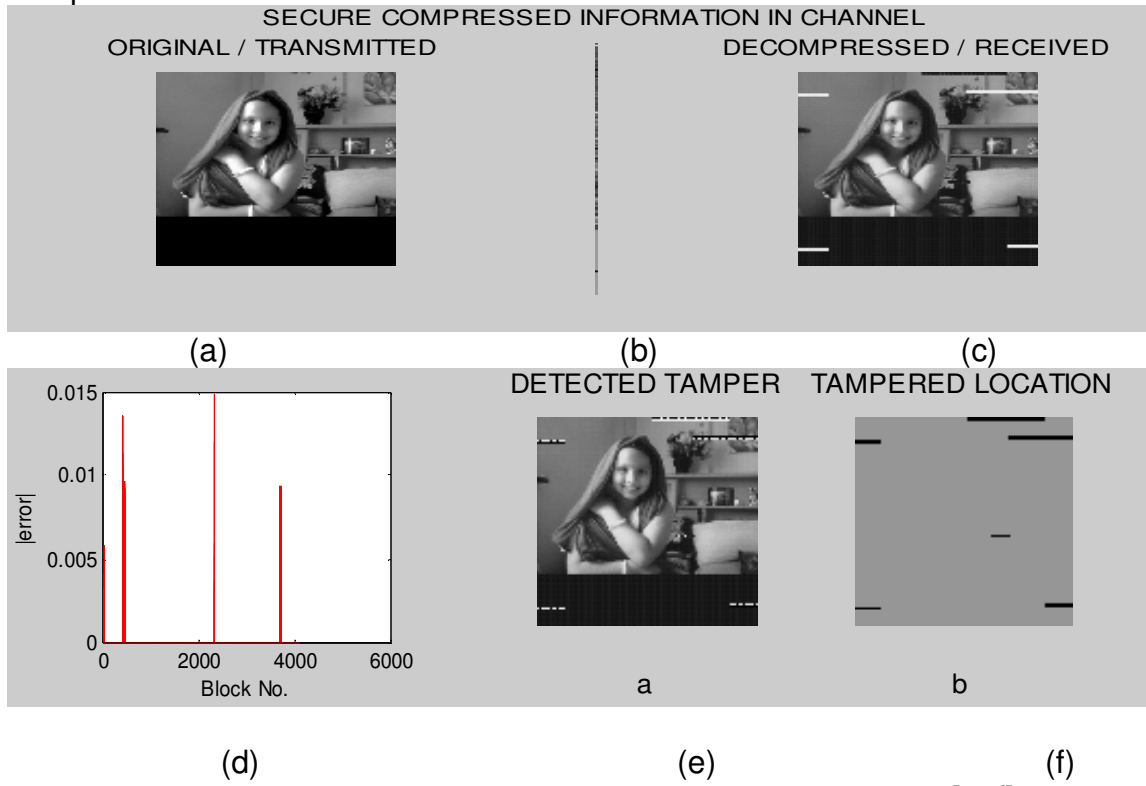
FIGURE 9: UNICAP performance with Test image (jpg format) [a: f]

Tampered case 3:

SECURE COMPRESSED INFORMATION IN CHANNEL
ORIGINAL / TRANSMITTED   DECOMPRESSED / RECEIVED

(a)   (b)   (c)

DETECTED TAMPER   TAMPERED LOCATION

a   b

(d)   (e)   (f)

FIGURE 10: UNICAP performance with Test image (Tiff format) [a:f]

Tampered case 4**:**

SECURE COMPRESSED INFORMATION IN CHANNEL

ORIGINAL / TRANSMITTED   DECOMPRESSED / RECEIVED

(a)   (b)   (c)

(d)   (e)   (f)

FIGURE 11: UNICAP performance with Test image (bmp format) [a:f]

Performance for various images in different format with various tempered condition have shown in Figure.6 to Figure.11.Learning performance for compression and for authentication code generation have shown in Fig.6 (a) &(b).It is clear from curve that in both case proper learning happen with l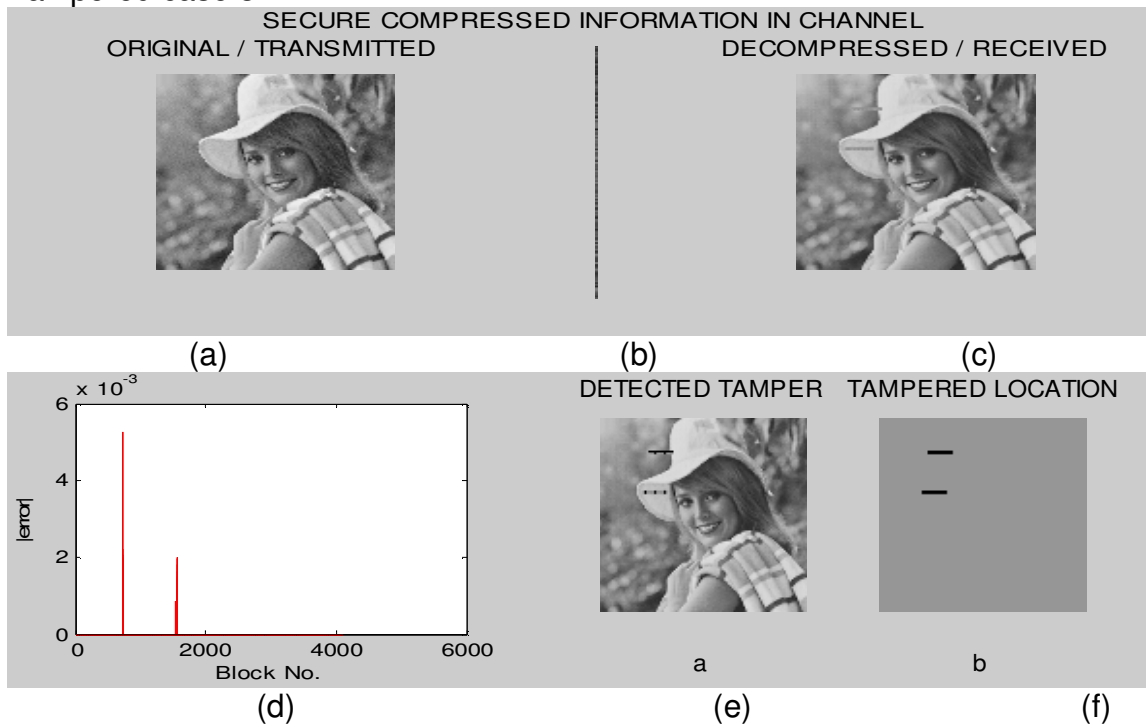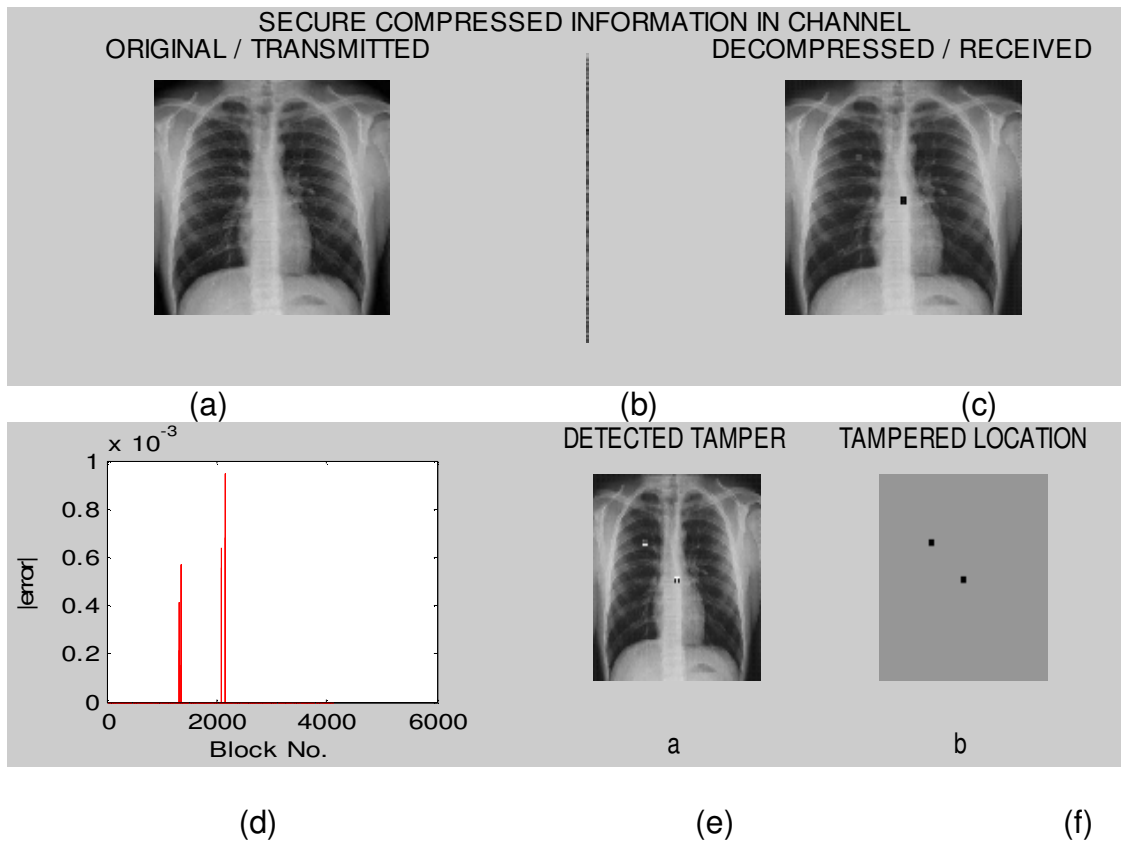ess number of operations. Performances for different images have shown in Figure7 to Figure11 in subsection of (a) to (f).Definition of information in each subsection has given below:
(a): Original image which has to be transfer from one location to other.
(b): Compressed data display of original image with compression ratio equal to 4.it is appear always as a vertical line hence it does not revel any information and protection available for original image.
(c): Received image after decompression of compressed data, obtained PSNR for each image without tampering is 32.50, 25.6, and 31 correspondingly.
(d): Absolute error in authentication code for each block, higher value indicate level of tampering is high, low value indicate low level tampering where as zero value represents unaltered.
(e): Image with identified tamper location available to clearly see the affect of tampering. Intentionally each block border of tampering made black or white to visualize clearly.
(f): Clearly indicate the position and location of tampering so that user can take the right decision with respect of tampered image.

7. ADVANTAGES OF PROPOSED UNICAP METHOD

(i) A good compression quality can achieve with simple process. Compression ratio is remain same irrespective of quality of image i.e not depends upon redundancy available in the image. This supports the requirement where fixed bandwidth has given for applications.

(ii) With a single image in training data set it is possible to define the compression for any other image having any format and of any quality.

(iii) Very easy means available to change the compression ratio. Solution for different compression ratio can be achieve by just changing the number of hidden layer neurons in architecture and apply the learning.

(iv) Very high Security of compressed data automatically appears without involving any further method exclusively. Level of security is so high it is nearly impossible to extract the information from compressed data.

(v) There is possibility to assign a unique compression facility for each member in a small group.

(vi) Very effectively and precisely authentication of image as well as position of alteration can defined, so that receiver could take the decision of acceptance or rejection of received image with quantity and quality of alteration.

(vii) Simple approach and easy implementation makes the solution cost effective.

8. CONCLUSION

In this paper solution of available fundamental challenges in image based applications at present to providing the protection of image data from unauthorized viewer and how to identify any mischievous alteration given in image has presented. Solution has derived by various integral properties like universal approximation, one-way property and one to one mapping associated with neural network. Direct compression method applied to obtain the fixed compression ratio as well as for protection. It has shown that even with a single image, learning of neural network can have generalized compression capability and decompression is nearly impossible without having optimal weights. For authentication purpose a feed forward architecture with compression data taken as input and single output neuron to generate the authentication code has taken. Proposed method is also securing the position of modification in image. Simplicity and efficiency available in proposed solution make it applicable for practical applications not only for internet based application but also for single system/device based applications.

9. REFERENCES

[1] Syed Ali Naqi Gilani , M. Ajmal Bangash ," Enhanced Block Based Color Image Encryption Technique with Confusion". IEEE ,INMIC 2008. pp. 200 – 206.

[2] Srividya.G, Nandakumar.P "A Triple-Key Chaotic Image Encryption Method". International Conference on Communications and Signal Processing (ICCSP), Feb.2011, pp. 266 – 270.

[3] Nidhi S Kulkarni, Indra Gupta and Shailendra N Kulkarni,"A Robust Image Encryption Technique Based on Random Vector " First International Conference on Emerging Trends in Engineering and Technology", ICETET, July 2008, pp. 15 – 19.

[4] Mahmood Al-khassaweneh and Selin Aviyente, "Image Encryption Scheme Based on Using Least square Approximation Techniques ", IEEE International Conference , May 2008 ,pp. 108 – 111.

 [5] Rong-Jian Chen, Wen-Kai Lu and Jui-Lin Lai "Image Encryption Using Progressive Cellular Automata Substitution and SCAN". IEEE International Symposium on Circuits and Systems, ISCAS , Vol. 2, 2005. pp. 1690 - 1693

[6] Shaojiang Deng, Linhua Zhang and Di Xiao,"Image Encryption Scheme Based on Chaotic Neural System "Springer,LNCS, Volume 3497/2005, pp. 810.

[7] Feng Huang and Yong Feng, "Security analysis of image encryption based on two-dimensional chaotic maps and improved algorithm ",Springer, Frontiers of Electrical and Electronic Engineering in China Volume 4, Number 1, 2009, pp.5-9.

[8] Amir Akhavan, Hadi Mahmodi and Afshin Akhshani,"A New Image Encryption Algorithm based on One-Dimensional Polynomial Chaotic Maps",springer,LNCS, 2006, Volume 4263, pp.963-971.

[9] Jun Peng and Du Zhang,"Image Encryption and Chaotic Cellular Neural Network", Springer 2009.pp. 183-213,

[10]. Xingyuan Wang, Xiaojuan Wang, Jianfeng Zhao and Zhenfeng Zhang,"Chaotic encryption algorithm based on alternant of stream cipher and block cipher", Nonlinear Dynamics, Volume 63, Number 4, Springer 2011, pp.587-597.

[11] Eric Kee, Micah K. Johnson, and Hany Farid, "Digital Image Authentication From JPEG Headers", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, vol. 6, n. 3, Septmber 2011, pp. 1066 – 1075.

[12] Li Lizong; Gao Tiegang, Gu Qiaolun. Bi Lei. "An image authentication and verification based on public Key ",International Conference on Artificial Intelligence and Computational Intelligence Nov. 2009 , pp. 389 – 392.

[13] Poonkuntran.S.Rajesh,R.S.,"A messy watermarking for medical image Authentication", International Conference on Communications and Signal Processing (ICCSP), Feb. 2011, pp. 418 - 422.

[14] Mona F.M. Mursi, Ghazy M.R. Assassa, Hatim A. Aboalsamh, Khaled Alghathbar. "A Secure Semi-Fragile JPEG Image Authentication Scheme Based on Discrete Cosine Transform". International Conference on Computing, Engineering and Information 2009, pp. 285 – 291.

[15] Zhenni Peng, Wenbo Liu,"Color image authentication based on spatiotemporal chaos and SVD"Chaos, Solitons & Fractals,Volume 36, Issue 4, May 2008, pp. 946-952.

[16] un-Chou Chuang , Yu-Chen Hu, "An adaptive image authentication scheme for vector quantization compressed image", *Elsevier* Journal of Visual Communication and Image Representation, Volume 22, Issue 5, July 2011, pp. 440-449.

Face Recognition Using Improved FFT Based Radon by PSO and PCA Techniques

Mr. Hamid M. Hasan   Hamid2012net@gmail.com
Electrical Eng. Dept.
Basra University, college of Eng.
Basra , Iraq

Prof. Dr. Waleed A. AL.Jouhar   Profwaleed54@yahoo.com
Electrical Eng. Dept.
Baghdad University, College of Eng.
Baghdad , Iraq

Dr. Majid A. Alwan   Altimimee@yahoo.com
Electrical Eng. Dept
Basra University, college of Eng.
Basra , Iraq

Abstract

Face Recognition is one of the problems which can be handled very well using Hybrid techniques or mixed transform rather than single technique. This paper deals with using of Radon Transform followed by PCA and LDA techniques for Face Recognition. The data used are 2D Face Images from ORL Database. The Radon Transform used is based on FFT slice theorem. The directions along which the Radon transform is performed are selected using PSO in order to achieve a good recognition rate. The best directions selected are less computation expensive as compared to the full set of directions and achieve good recognition rate. The PCA is used to reduce the dimension of the data produced by Radon Transform and the LDA is used to find a set of basis vectors which maximizes the ratio between-class scatter and within–class scatter. In order to verify our method many dataset partitioning scenarios into training set and testing set were conducted. And the maximum recognition rate achieved was 97.5%.

Keywords: Face Recognition (FR), Radon Transform (RT), Fast Fourier Transform (FFT), Principal Component Analysis (PCA), Linear Discrimenant Analysis(LDA) and Particle Swarm Optimization (PSO).

---

1. INTRODUCTION

Face recognition is one of the most important biometrics which seems to be a good compromise between actuality and social reception and balances security and privacy well. It has a variety of potential applications in information security law enforcement and access controls. Face recognition systems fall into two categories: verification and identification. Face verification is 1:1 match that compares a face images against a template face image. On the other hand face identification is 1: N problem that compares a probe face image against all image templates in a face database. Face recognition is a very difficult problem due to a substantial variations in light direction (illumination) , different face poses , diversified facial expressions , Aging (changing the face over time) and Occlusions (like glasses, hair, cosmetics). So the building of an automated system that accomplishes such objectives is very challenging. In last decades many systems with recognition rate greater than 90% has been done however a perfect system with 100% recognition rate remains a challenge. Face recognition algorithms are divided by [1, 2] into three categories as follows:

1. Holistic methods: These methods identify a face using the whole face images as input and extract the overall features.
2. Feature based methods: these methods used the local facial features for recognition (like eyes, mouths, fiducial points. etc.).
3. Hybrid methods: these methods used both feature based and holistic features to recognize a face. These methods have the potential to offer better performance than individuals.

2. A REVIEW OF THE RELATED 2-D FACE RECOGNITION TECHNIQUES

Thomas Heseltine [3] investigated three appearance based approaches for face recognition which are the direct correlation method, the eignface method and fisherface method. the recognition error rate reported is 18% , 20.4% , 17.8% respectively .

M. Chandra Mohan[4] they divide the face into four parts and evaluates the texture features in each part separately the texture features are derived from parameters with different orientations, this makes the face recognition easier and pose, illumination and rotation invariant.

P.Abouzar [5] using WT (Wavelet) and DCT (Discrete Cosine Transform) followed by PCA the proposed algorithm takes advantages of data reduction property of the three transforms. The Support Vector Machine (SVM) was used to classify the images into different classes and the error rate obtained is between 5%-7%.

Zhan Shi[6] they extract a number of features from facial images through taking Trace Transform over different angular directions by using different trace functions then the features are projected into a lower dimensional subspace. The recognition rate achieved is 95% in ORL database.

Laika Karsili [7] used a Radon Transform over the set of angles{0,60,120,180,240,300,360}, then the produced data was reduced using PCA this achieves 70% recognition rate for rank 1 and 95% for rank 4.

Jamal A hmad [8] Investigated the effect of the step size for both the angle and the vector of the radon transform on the performance of a face recognition system based on PCA it is founded that step size of one for both produces recognition rate of 89%.

ZHANG et al, [9] proposed a feature extraction method based on finite Radon transform (FRAT) then used soft threshold (ST) to select main FRAT coefficients. Finally 2DMMC was used to extract features for classification from main FRAT coefficients. They achieved 89.02% recognition rate on ORL database.

Ergun Gumus et al [10] they used Eigenfaces (PCA) and Support Vector Machine (VSM) on ORL database they achieved recognition rate of 91,2% for PCA-AVM (RBF) Radial bases.

Zhang Lin et al [11] used Radon Transform with multiwavelet and PCA on Infrared imaged faces and they achieved 95% classification accuracy with 70 element feature vector.

Yuehui Chen et al [12] proposed using DCT and Hybrid Flexible Neural Tree which was evolved using PSO their experiment on ORL achieved 98.13% recognition rate .

Jian Zhang , Xianyun Fei[13] they used the PSO in order to select the optimum discrimination eigenvectors of PCA and obtain the optimal recognition accuracy simultaneously they validate their method with ORL database with recognition rate of 96%.

Dattatray V. Jadhao and Raghunath S. Holambe[14] they used Radon transform and Fourier transform for face recognition on ORL database and achieved recognition rate of 97.33% the images were classified on the nearest neighbor with 60 features .

Rabab M.Ramadan [15] used PSO to select the efficient features from DCT and DWT and apply their method on ORL with recognition rate 94.7% and 96.8% respectively .

In this paper we used the Radon Transform which is improved by PSO to select the best directions. Then a data reduction is performed using the PCA. A classification basis vectors are derived using LDA which lead to a rank one recognition rate equal to 97.5% and the feature vector size is 35 items per class (person).

3. THE METHOD AND MATERIALS

The method used in this paper for face recognition is depicted in figure (1). It consists the enrollment phase and the testing phase. In the enrollment phase the training set of images are transformed into Radon space using the Radon Transform. The set of directions (angles) along which the transform is performed were calculated using the Particle Swarm Optimization (PSO). Different sets of directions are shown in tables (1, 2, 3). The different sets are for different classifier parameters which yield a good estimated recognition rate. The data generated by Radon Transform are reduced using Principal Components Analysis (PCA). From those reduced data set (i.e. the most effective components) a set of basis vectors which maximizes the ratio between-class scatter and within–class scatter using Linear Discrimenant Analysis (LDA). One basis vector for each class (i.e. Person) is derived and stored in the data base. So for ORL Image database there are 40 basis vectors are stored regardless of the number of images used for each person in the training set. The length of the basis vector depends on the number of components selected by the PCA stage. In the testing phase the input image is transformed into the Radon space using that set of directions which were used in the enrollment phase. The PCA reduction is carried out as same as in the enrollment phase. The resultant vector is projected into the basis vectors stored in the data base that is by inner product method. The highest product value which must be higher than a predetermine threshold measures the similarity between the input image and the specified class. The method which is stated in this paper was evaluated using the ORL data base which contains photographs of faces taken at the Olivetti Research Laboratory in Cambridge between April 1992 and April 1994.There are 10 different images of 40 distinct subjects so there are 400 images in the data base. The images are grayscale with a resolution of 92 x 112. For some of the subjects, the images were taken at different times. There are variations in facial expression and pose variation about 20 degrees and there is some variation in scale of up to about 10%. There are some faces with glasses. Some images are shown in figure (2). In the following sections the discussion of each main part of the method of process is presented.

a:Enrollment Phase   b: Identification Phase
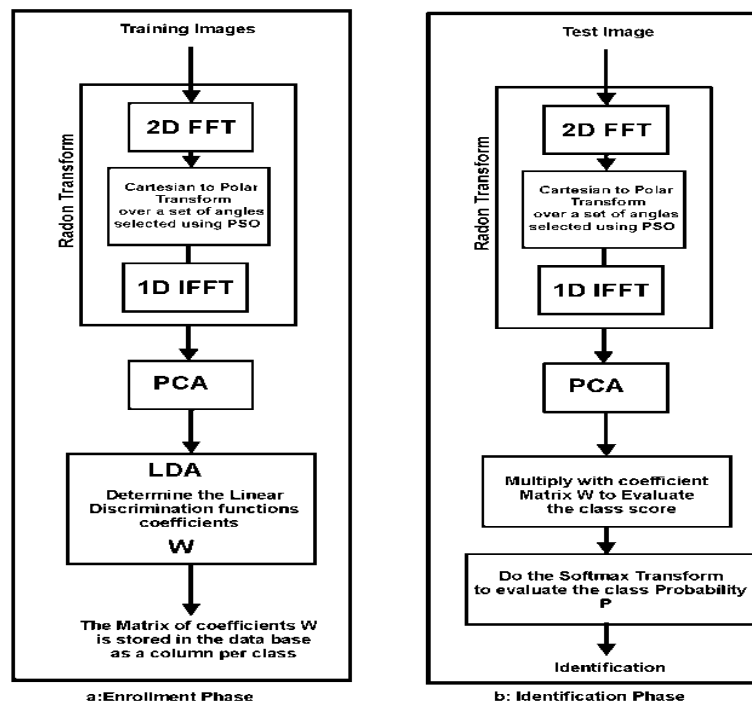
FIGURE 1: Block diagram for the recognition system.

3.1   Radon Transform

There were many applications of Radon transform like Pose Estimation [17],Texture Analysis [18], X-ray crystallography and Nuclear Magnetic Resonance(NMR)[19], Image Matching for translation, rotation and uniform scaling Using Radon Transform[21]. Several definitions of the radon transform existed. A very popular form expresses lines is:-

$$t = x * \cos(\theta) + y * \sin(\theta) \qquad (1)$$

Where θ is the angle and t is the smallest distance to the origin of the coordinate system. The Radon transform for a set of parameters (t, θ) is the line integral through the image f(x, y), where the line is positioned corresponding to the value of (θ and t) in Equation (2).

$$g(t,\theta) \quad = \int\limits_{-\infty}^{+\infty}\!\!\!\int f(x,y)\delta\big(t - x\cos(\theta) - y\sin(\theta)\big)\mathrm{dx\,dy} \qquad (2)$$

In the above equation g(t,θ) is the Radon Transform of the f(x,y) function at a specified value of t and θ . The Radon transform can be calculated using FFT by applying the central-slice theorem [13] which is stated that the 1-D Fourier transform of the integral projection at angle θ is equal to the slice of the 2-D Fourier transform at the same angle. With the central-slice theorem the Radon transform can be computed equivalently with 2-D FFT, a Cartesian-to-polar mapping, and a 1-D FFT. The range of the θ is [0...179] and the range of the variable (t) depends on the dimension of the underling image. In our case the t is in the range [1...112] and θ range is [0...179] that is for full range directions. So each image is represented in Radon space as a vector of dimensions equal to (180 x112) = 20160. In our work a subset of θ is selected from the full range using Particle Swarm Optimization (PSO). The subset of θ values is selected according to the goodness of the recognition rate achieved as the objective function. Using different values of parameters used by the classifier a different subset are selected using PSO. Figure (3) shows a full range of θ Radon transform and figure (4) shows a subset of θ Radon transform.

3.2   Particle Swarm Optimization (PSO)

PSO, first introduced by Kenny and Eberhart in 1995 [22], is one of the evolutionary computation technology based on swarm intelligence. In a PSO system each solution called a "particle", particles fly around in the search space of the problem to look for the optimal solution. Each particle adjusts its position according to the flying experience of its own and the experience of neighboring particles. Each particle updates its velocity and position using the following equation [23]:-

$$V_i(k+1)=V_i(k)+c1 * rand1(\text{ pbest}(k) \text{ - } X_i(k)) + c2* rand2(\text{ gbest}(k) \text{ - } X_i(k)) \qquad (3)$$

$$X_i (k+1)=X_i(k) + V_i(k+1) \qquad (4)$$

Where ,
    V_i is called the velocity for particle i;
    X_i is represent the position of particle i ;
   Pbest   is the best position of ith particle
   gbest   is the global best position ;
    rand   is random variable in [0,1];
 c1 and c2 are the learning factors;
    w   is called the inertia weight

To search for optimal solution, each particle changes its velocity according to equation 3. The value of V_i is clamped to the range [Vmin, Vmax] to avoid excessive roaming of particle outside the search space. Then each particle moves to a new potential solution using equation 4. This process is repeated until a stopping criterion is reached. The above equations (4,5) are the basic

equations and not guarantee the global solution. Using those equations PSO may be trapped in local minima to avoid the local minima there are many variations and improvements to the basic equations have been suggested [24-37]. In this work we used the PSO in order to select the best set of directions (the values of θ) to be used in Radon transform. These set is selected from the full set of θ from 0☐ to 179☐ in step of 1 degree. There are 24 particles in the swarm are used with randomly selected values of θ. These particles moved and evolved towards the best positions that yield a good Recognition rate. The objective function was constructed in order to optimize both the recognition rate and the number of directions (θ) selected as well. Three sets are depicted by tables(1 ,2,3) for three values of N which is the length of the features vector used by the classifier associated with objective function used. In our experiments, the set with N=60 was used. One thing that be mentioned about PSO is its convergence to a steady state of accepted recognition rate very quickly and in stable manner.

3.3 Feature Reduction Using (PCA)

The PCA is a statistical method for reducing data dimensions [37,38]. In PCA the training data is used in obtaining the Eigen basis vectors. Then the training set R and testing set T are projected into those vectors. The PCA can be summarized by the following steps:-

a. Calculate the mean M of the training set and subtract it from the training set:

$$M = \frac{1}{n}\sum_{i=1}^{n} R_i \qquad (5)$$

For all training set perform $R_i = R_i - M$

b. Calculate the Eigenvectors and Eigenvalues of the training set covariance matrix S_t

$$S_t = R \ . R^T \qquad (6)$$

and pick the Eigen vectors corresponding to the N largest Eigen values of S_t. These construct the N principal components matrix (V).

c. The Basis vectors matrix (U) is constructed as:

$$U = V \ . R^T \qquad (7)$$

d. The reduced feature vector is calculated for training and testing data as:

$$W_R = R \ . U \qquad (8)$$

$$W_T = T \ . U \qquad (9)$$

The mean M must be subtracted from the testing data T also. Many experiments were conducted using different numbers of Eigenvectors (N) between (10-60) and the performance are reported in figure(5,7).

3.4 Linear Discrimenant Analysis (LDA)

The LDA finds a set of basis vectors which maximizes the ratio between-class scatter and within –class scatter [37]. Given N samples of C classes, let Ni be the number of samples in the ith class Ci , let Mt be the mean of the whole data set, mi be the mean of the ith class Ci, the between-class scatter matrix is defined by:

$$S_B = \sum_{i=1}^{c} N_i \ (mi - Mt) \ (mi - Mt)^T \qquad (10)$$

And the within-class scatter matrix is defined by

$$S_w = \sum_{i=1}^{c} \sum_{mk \in ci} \quad (mk - mi) \quad (mk - mi)^T \qquad (11)$$

Then the basis vectors is

$$W = \arg\max \left(\frac{|W^T S_B W|}{|W^T S_w W|} \right) \qquad (12)$$

Solving Equation(9) produces a matrix W whose columns are the eigenvectors corresponding to the largest eigenvalues of $S_W^{-1} S_B$. These columns are the linear discriminant functions associated with classes as a one function for each class. These functions are stored in the database in order to perform the classification as following:

Each input vector (T) to be classified is multiplied with W matrix as ($V = T^{-1}W$) the resulting vector is the linear scores of the testing data T. The class probabilities are calculated using the softmax transform as:

$$P \quad = \frac{\exp(V(i))}{\sum_{i \in C} \exp(V(i))} \qquad (13)$$

Where C is the set of classes. Each class represents a person in the data base. The value of P show how much the testing data T is near to a specified class, the higher value is the nearest class so the classification is done. A threshold value P can be designated with in order to reject the unknown or to do misclassification.

4. THE EXPERIMENTS

The developed face recognition method was applied to the ORL database. Five experiments were conducted. Each with different partitioning scenario to the data set and the performance was evaluated against the number of eigenvectors used see figure (5). The set of eigenvectors is {10,20,25,30,35,40,45,50,55,60}. The set of directions (angles) used for radon transform is shown in table (2). The five scenarios are:

a. Scenario #1.
In this scenario the ten images for each person are divided as 5 images for training and 5 images for testing. The images were randomly selected. The maximum recognition rate was 93% that is when 25 eigenvectors are selected. See figure (5). The Boxplot is shown in figure (6). It is appear that the median is 90.5%, and the 75th percentile is around 92.5%.

b. Scenario #2.
In this scenario the ten images for each person are divided as 9 images for training and one image for testing. The images were randomly selected. The maximum recognition rate was 97.5% that is when 35 eigenvectors are selected. See figure (5). See Boxplot at figure (6). It appears that the median is 95%, and 75th percentile around 97%.

c. Scenario #3.
In this scenario the ten images for each person are divided as 8 images for training and 2 images for testing. The images were randomly selected. The maximum recognition rate was 97.5% that is when 35 eigenvectors are selected. See figure (5). See Boxplot at figure (6). It appears that the median is 95%, and 75th percentile around 97.2%.

d. Scenario #4.
In this scenario the ten images for each person are divided as 7 images for training and 3 images for testing. The images were randomly selected. The maximum recognition rate was 97.5% that is when 50 eigenvectors are selected. See figure (5). See Boxplot at figure (6). It appears that the median is 96.7%, and 75th percentile around 96.7%.

e. Scenario #5.
In this scenario the ten images for each person are divided as 6 images for training and 4 images for testing. The images were randomly selected. The maximum recognition rate was 95.63% that is when 60 eigenvectors are selected. See figure (5). See Boxplot at figure (5). It appears that the median is 94.4%, and 75th percentile around 94.9%.

5. DISCUSSION

In this work the using of Radon transform improves the performance of PCA+LDA techniques in face recognition. That is if compared with mentioned literature [3, 5, 7, 8, 10, and 13]. The using of PSO for selecting the best directions (angles) used by Radon transform give better performance as compared to [7,9]. This means that the subset selected using PSO is better than the one selected in [7]. From the different scenarios conducted it appears that the performance is improved with the increased number of images per person in the training set this is clear by scenario #2 ,#3,#4 . From Boxplot in figure (7), which shows the performance against the number of eigenvectors used over all scenarios in our experiment, it is clear that the number of eigenvectors between (30-40) gain the good performance. It is also clear that the (35) eigenvectors is the best.

6. CONCLUSIONS

A Face Recognition method has been described in this work. The core of this work is to apply the PCA+LDA in Radon space rather than directly to the images. The images are transformed using Radon Transform with a specified angles (directions) set determined using PSO in order to achieve good recognition rate with less computation expensive. The Radon transform used in this method was FFT based. The full range Radon transform is computational expensive if it is performed for angles from 0 degree to 180 degree and for a large number of offsets. To reduce the computations required a subset of angles and offset must be selected. So the PSO was used to select that subset and maintaining a good recognition rate. This method was verified on ROL data base using five different scenarios for training set selection. The best recognition rate was 97.5% when only 35 eigenvectors are used. The number of eigenvectors determines the length of the signature vector that be used for each person in the data base. The recognition rate and the size of signature that represent each person as well as the computation of Radon transform that a achieved in this method is better than the related works stated in the literature review in section 2.

2 : Samples from ORL Database

77 angles calculated using PSO N=60							
2	4	11	19	21	23	26	28
29	35	36	39	40	41	42	43
44	46	47	51	52	57	58	59
61	62	65	68	71	73	75	79
80	81	83	84	85	89	90	92
93	95	96	98	99	100	106	108
111	112	119	120	122	126	131	134
124	125	126	130	132	134	135	137
135	140	141	144	145	146	149	155
156	161	162	174	178			

TABLE 1 : angles calculated using PSO N=60

84 angles calculated using PSO N=10							
1	6	8	10	11	12	14	15
17	20	25	28	30	32	33	35
47	49	52	53	54	56	58	59
60	61	62	63	64	68	69	70
71	72	73	75	79	80	84	85
92	95	96	98	103	104	105	108
109	110	114	115	116	119	121	122
124	125	126	130	132	134	135	137
138	142	143	146	151	152	153	154
156	159	161	166	168	169	170	171
172	173	175	179				

TABLE 2: angles calculated using PSO N=10

88 angles calculated using PSO N=32							
1	3	4	6	10	11	16	23
24	25	26	27	28	31	32	33
34	35	36	38	42	43	46	47
48	51	52	53	55	57	58	60
61	62	64	73	79	82	86	87
89	91	92	97	98	99	100	101
103	106	107	108	109	112	118	119
120	121	122	123	124	125	126	128
129	134	136	140	141	142	145	147
149	150	151	152	153	156	158	162
163	164	166	168	171	175	176	177

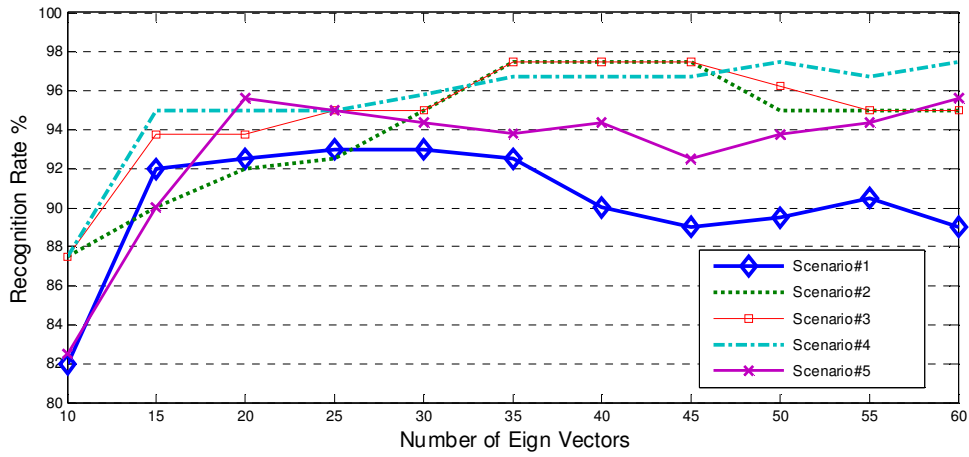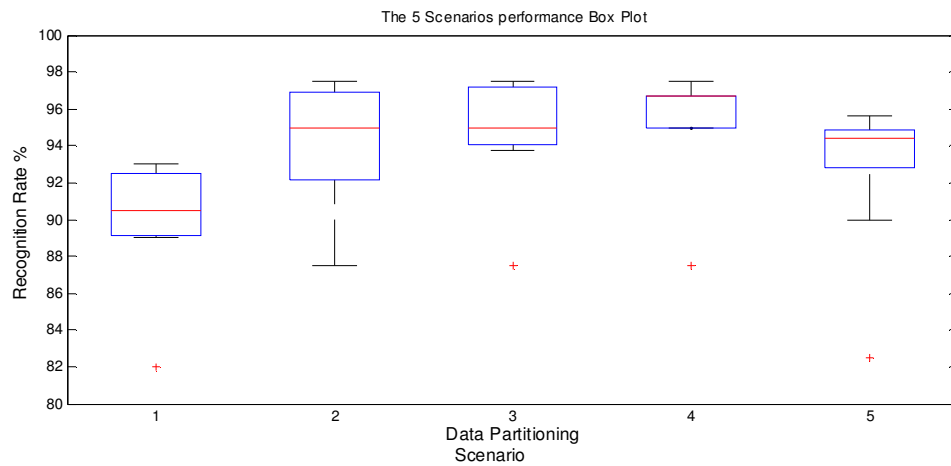**TABLE 3**: angles calculated using PSO N=32
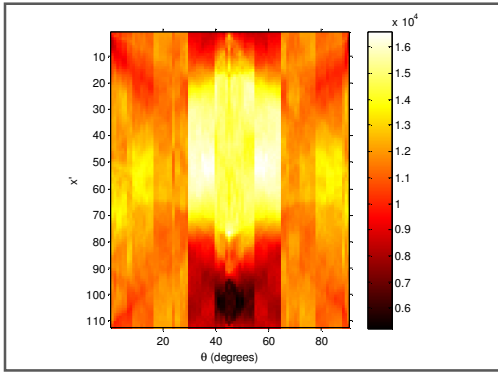
FIGURE 4 :FFT Based Radon Transform using 90 angles transform for an image calculated using PSO
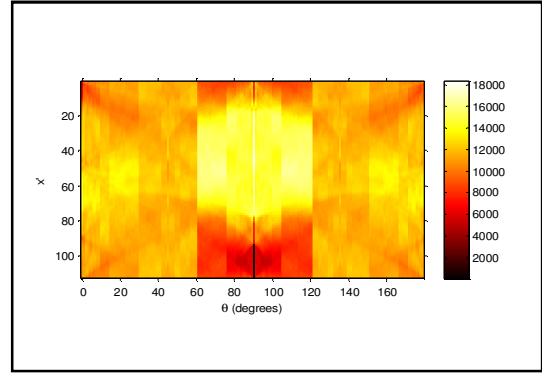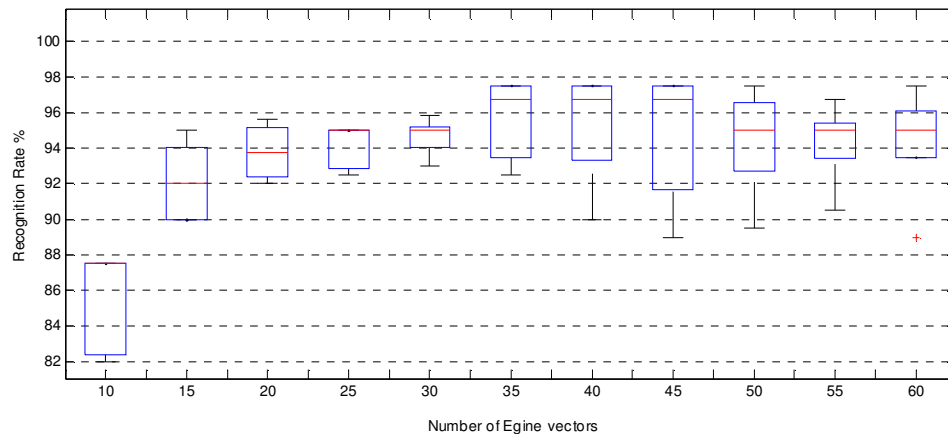
FIGURE 3: FFT Based Radon for an image

FIGURE 5: Performance against #of eigenvectors for different scenarios.

FIGURE 6: Box Plot show the performance for each scenario.

FIGURE 7: Box plot shows the performance against #of eigenvector cross the 5 scenarios.

7. REFERENCES

[1] Nilima B. Kachare, Vandana S. Inamdar, 2010, " Survey of Face Recognition Tecchniques", International Journal of Computer Applications(0975-8887), Volum 1-No.19,2010.

[2] Patil A.M., Kolhe S.R. and Patil P.M., 2010,"2D Face Recognition Techniques:A Survay", International Journal of Machine Intelligence.

[3] Thomas Heseltine, Nick Pears,Jim Austin,Zezhi Chen, 2003, "Face Recognition: A Comparison of appearance-Based approaches", Proc.VIIth Digital Image Computing :Techniques and Applications.,Sun C., Talbot H.,Ourselin S. and Adriaanen T. (10-12 Dec,2003,Sydny).

[4] M. Chandra Mohan, V. Vijaya Kumar,K.V.Subbaiah,(2010)," A New Method of Face Recognition Based on Texture Feature Extraction on Individual Components of Face",International Journal of signal and Image Processing (Vol 1-2020/ISS.2)pp.69-74.

[5] P.Abouzar, Yousefi,S.K.Setarehdan,(2007), "Hybrid WT Based-DCT Based Face Recogntion" , 2007 IEEE International Conference on Signal Processing and communications(ICSPC 2007). 24-27 November 2007, Dubai,United arab Emirates.

[6] Zhan Shi, Minghui Du, Rongbing Huang,(2010),"A Trace Transform based on subspace method for Face Recognition", 2010 International Conference on Computer Application and System Modeling (ICCASM 2010).

[7] Laika Karsili and Adnan Acan,2007,"A Radon Transform and PCA Hybrid for High Performance Face Recognition", IEEE International Symposium on Signal Processing and Information Technology.

[8] Jamal A hmad Dargham et al. (2010) "Radon transform for face recognition", Artif Life Robotics(2010) 15:359-362,ISAROB 2010.

[9] ZHANG Yuhua,WANG Xin,(2010),"Study of Finite Radon Transform in Face Recognition",2010 Second International Conference on Computer Modeling and Simulation.

[10] Ergun Gumus, et al., "Eigenfaces and Support Vector Machine Approaches for Hybrid Face Recognition", The Online Journal on Electronics and Electrical Engineering (OJEEE) Vol(2)-No.4.

[11] Zhang Lin,et al. "Infrared Face Recognition Based On Radon and Multiwavelet Transform" ,Proceedings of ICCTA 2009.

[12] Yuehui Chen, Shuyan Jiang, Ajith Abraham, " Face Recognition Using DCT and Hybrid Flexible Neural Tree",2005IEEE,Development Program of Shandong under contract number SDSP2004-0720-03.

[13] Jian Zhang,Xiany un Fei, " A New Method for Face Recognition Based on PCA Optimize Strategy"; 2010 International Conference on Computer Application and System Modeling (ICCASM)2010.

[14] Dattatr V.Jadhao, Raghunath S.Holambe; "Feature Extraction and Dimensionality Reduction Using Radon and Fourier Transform with Application to Face Recognition", International Conference on Computational Intelligence and Multimedia Application 2007.

[15] Rabab M. Ramadan and Rehab F. Abdel Kader; " Face Recognition Using Particle Swarm Optimization-Based Selected Features", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No. 2, June 2009.

[16] Daming Shi. , Liying Zheng, and Jigang Liu," Advanced Hough Transform Using A Multilayer Fractional Fourier Method", IEEE Transactions on Image Processing, VOL, 19. NO, 6, JUNE 2010.

[17] Patrick Etyngier et al., "Radon Space and Adaboost for Pose Estimation", Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06) 2006 .

[18] Mahmoud R.HEJAZI and YO Sung HO," Texture Analysis Using Modified Discrete Radon Transform", IEICE TRANS.INF. & SYST., VOL.E90-D, NO.2 FEBRUARY 2007.

[19] S. Venturras, I. Flaounas ," Study of Radon Transformation and Application of its Inverse to NMR", Paper for " Algorithms in Molecular Biology" Course Assoc Prof. I. Emiris, 4 July, 2005.

[20] Ming Jiang,Chih ting Wu," Wavelet Based Local Tomography" Mathematical Methods in Medical Imaging , Final Project, Math-6792 Spring 2003.

[21] Jiangsheng You, Weiguo Lu, et al; " Image Maching for translation, rotation and uniform scaling by the Radon Transform",1998 IEEE.

[22] Alec Banks, et al. " A review of particle swarm optimization. Part II: hybridization, combinatorial, multicriteria and constrained optimization, and indicative applications", Nat Comput (2008) 7:109-124, DOI 10.1007/s11047-007-9050-z.

[23] Shih wei Lin and shih chieh Chen, " PSOLDA: A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis", Applied Soft Computing 9 (2009) 1008-1015.

[24] Millie Pant et al, "A New Quantum Behaved Particle Swarm Optimization",GECCO'08,July 12-16, 2008 Atlanta, Georgia. USA.

[25] Leandro dos Santos Coelho," A quantum particle swarm optimizer with chaotic mutation operator", Chaos, Solutions and Fractals 37 (2008) 1409-1418.

[26] O. Togla altinoz, et al. " Chaos Particle Swarm Optimization PID Controller for the Inverted Pendulum System", 2nd International Conference on Engineering Optimization, September 6-9, 2010, Lsbon, Portugal.

[27] Leandro dos Santos Coelho and Viviana Cocco Mariani, " A novel chaotic particle swarm ooptimization approach using Henon map and implicit filitering local search for load dispatch", Chaos, Solutions and Fractals 39(2009) 510-518.

[28] Qing Zhang, et al. " Fast Multi swarm Optimization with Cauchy Mutation and Crossover operation", Puplications of China University of Geosciences, School of Computer, Wuhan, P.R.China, 430074.

[29] Yanjun Yan and Lisa ann Osadciw, "Varying Dimensional Particle Swarm Optimization", 2008 IEEE swarm Intelligence symposim , St. Louis Mo USA, September 21-23,2008.

[30] R. V. Kulkarni and G.K. Venayagamoorthy, " An Estimation of Distribution Improved Particle Swarm Optimization Algorithm", ISSNIP 2007.

[31] Yanj Yan, Ganapathi Kamath and Lisa ann Osadciw, " Feature Selection Optimization by Discrete Particle swarm Optimization for Face recognition", Syracuse University , Syracuse , NY, USA 13244.

[32] Hong Pan, LiangZhengXia, and Truong Q.Nguyen," Robust Object detection Scheme using feature selection", Proceeding of 2010 IEEE 17th International Conference on Image Processing , September 26-29, 2010, Hong Kong.

[33] Osslan Osiris Aergara Villegas and Viancy Guadalupe," a Novel Evolutionary Face algorithm Using Particle Swarm Optimization ", 2009 Fith International Conference on Signal Image Technology and Internet Based Syetems.

[34] Lanzarini Laura , et al. " Face Recognition Using SIFT and Binary PSO Descriptors", 2010 Proceedings of the ITI 2010 32th Int. Conf. on Information technology Interfaces, June 21-24,2010, Cavtat, Croatia.

[35] Rajinda Senaratne, et al. " Face Recognition by Extending Elastic Bunch Graph Matching with Particle Swarm Optimization", Journal of Multimedia , VOL. 4, No. 4, August 2009.

[36] Ming Li, et al. " Application of Improved CPSO-SVM Approach in Face Recognition", 2009 Internaltional Conference on Artificial Intelligence and Com[utational Intelligence.

[37] Xiaorong Pu, Zhang YI, Zhongjie Fang," Holistic and partial facial features fution by binary particle swarm optimization", Neural Comput & Applic (2008) 17:481-488.

[38] Belhumeur PN, Hespanala JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711-720.

Rafia Mumtaz, Mustafa Mumtaz, Atif Bin Mansoor & Hassan Masood

Computer Aided Visual Inspection of Aircraft Surfaces

Rafia Mumtaz   *rafia.mumtaz@seecs.edu.pk*
School of Electrical Engineering and Computer Science
National University of Sciences and Technology
Sector H-12, Islamabad, Pakistan

Mustafa Mumtaz   *mustafa672@ieee.org*
College of Aeronautical Engineering
National University of Sciences and Technology
Sector H-12, Islamabad, Pakistan

Atif Bin Mansoor   *atif-cae@nust.edu.pk*
College of Aeronautical Engineering
National University of Sciences and Technology
Sector H-12, Islamabad, Pakistan

Hassan Masood   *hassan13204@yahoo.com*
College of Aeronautical Engineering
National University of Sciences and Technology
Sector H-12, Islamabad, Pakistan

Abstract

Non Destructive Inspections (NDI) plays a vital role in aircraft industry as it determines the structural integrity of aircraft surface and material characterization. The existing NDI methods are time consuming, we propose a new NDI approach using Digital Image Processing that has the potential to substantially decrease the inspection time. Automatic Marking of cracks have been achieved through application of Thresholding, Gabor Filter and Non Subsampled Contourlet transform. For a novel method of NDI, the aircraft imagery is analyzed by three methods i.e Neural Networks, Contourlet Transform (CT) and Discrete Cosine Transform (DCT). With the help of Contourlet Transform the two dimensional (2-D) spectrum is divided into fine slices, using iterated directional filterbanks. Next, directional energy components for each block of the decomposed subband outputs are computed. These energy values are used to distinguish between the crack and scratch images using the Dot Product classifier. In next approach, the air-craft imagery is decomposed into high and low frequency components using DCT and the first order moment is determined to form feature vectors.A correlation based approach is then used for distinction between crack and scratch surfaces. A comparative examination between the two techniques on a database of crack and scratch images revealed that texture analysis using the combined transform based approach gave the best results by giving an accuracy of 96.6% for the identification of crack surfaces and 98.3% for scratch surfaces.

Keywords: Computer Vision, Gabor Filter, Contourlet Transform, Non Subsampled Contourlet Transform, Discrete Cosine Transform, Neural Networks.

1. INTRODUCTION
Vision is the most advanced of our senses, so the concept of keeping or storing one of these senses i.e. the visual sense, seems interesting to all the human beings. Today images are being used in almost all applications of daily day life and research. Satellite and space imagery, industrial radiographs, radars and photoreconnaissance, infrared studies, mapping, pollution analysis etc use images as a basic tool for their study. With the advancement in other fields of study and research, image processing developed itself from optical (analog) processing to DIGITAL IMAGE PROCESSING (DIP).Visual inspection of aircraft is widely used for ensuring

structural integrity of surface and its substructures. Visual inspector examines an aircraft for defects such as cracks, corrosion, damaged rivets, bird hit, lightning strike etc [1]. The inspection may suffer due to many reasons like competence level of inspector, lack of interest boredom or delicate nature of defect etc. Enhanced visual inspection could allow the inspector to safely, quickly and accurately perform the necessary visual inspection.

Non Destructive Inspection (NDI) techniques are used in aerospace industry for analyzing the aircraft surface and sub-surface defects. These techniques along with the visible cracks detect the microscopic cracks too. As stated in [2], the importance of aircraft surface inspections lies from the fact that a typical heavy inspection of a commercial aircraft is 90% visual and 10% NDI. Visual inspection helps in isolating the surface of the aircraft that may suffer from any failure. Commonly employed NDI techniques are Dye Penetrant Method, Fluorescent Penetrant Inspection, Magnetic Particle Inspection (MPI), Eddy Current Losses, Radiography, and Ultrasonic Inspection. These NDI techniques suffer from more down time and require large and costly setups. Robotics Institute of Carnegie Mellon University has carried out a research in development of a robot known as Crown Inspection Mobile Platform (CIMP) [2], [3]. This robot is capable of moving over the aircraft body and transmitting live stereoscopic imagery of the aircraft to the control center. The control center applies image enhancement and understanding algorithm to highlight areas of cracks and scratches. [4] made a similar research by acquiring aircraft images by a robot and subsequently processing them by Neural Networks. Training was performed by giving different images of multiple fasteners used in a modern aircraft. The technique resulted in differentiating healthy and crack areas of the aircraft surface.

This research is aimed to investigate various image processing algorithms to develop a cost effective and computationally efficient computer aided visual inspection of aircraft surface. The project involves real data collection by identifying and photographing crack susceptible areas of aircraft. In Section 2 Suggestions are formalized for inclusion of surface imaging in existing periodic inspections. Automatic marking of cracks is investigated through different methods like thresholding techniques of Otsu and Entropy, Gabor Filter for texture analysis, Nonsubsampled Contourlet transform and Neural Network as described in Section 3. Adaptive models have been optimized to address random illumination variations. A new visual method is devised to differentiate between crack and scratch other than the existing NDI techniques. Methods like Neural Network classifier, Discrete Cosine transform in collaboration with Dot Product classifier and energy calculations via Contourlet transform are applied in order to minimize False Alarm Rate. This method in elaborated in Section 4. Chapter is concluded in Section 5.

2. SUGGESTED SURFACE IMAGING
In today's aviation industry weekly and periodic inspection is carried out in accordance with predefined manuals and work cards. These work cards provide the requirements of the applicable aircraft scheduled inspection and maintenance requirements. The manual provides a checklist form and is use as a guide in performing the inspection to ensure that no item is overlooked. The inspection requirements are tested in such a manner as to establish which equipment is to inspected, when it is to be inspected, and what conditions are to be sought. In scope, the requirements are designed to direct the attention of maintenance personnel to components and areas where defects are suspected to exist as a result of usage under normal operating conditions. They are not intended to provide coverage for normal routine cleaning, washing, etc nor are they designed to lead to the detection of isolated discrepancies that are the result of carelessness or poor maintenance practices.

A new criteria is defined to carry out periodic inspection that is performed through imaging which will have the benefit of better record keeping and trend analysis of aircrafts. Various areas of aircraft that are prone to cracks were identified and were photographed weekly. The Identified surfaces prone to cracks were photographed weekly. Suggested scale and angle were defined and maintained while photographing.

Before explaining the criteria it is necessary to mention here that illumination cannot be controlled while taking the photographs outside. Illumination variance is not only due to climatic change but also with time due to changing position of sun. Therefore it is highly recommended that photography should be carried out in an indoor environment like a hanger where the image quality is less prone to be degraded, moreover the illumination variance though still there; can be neglected safely.
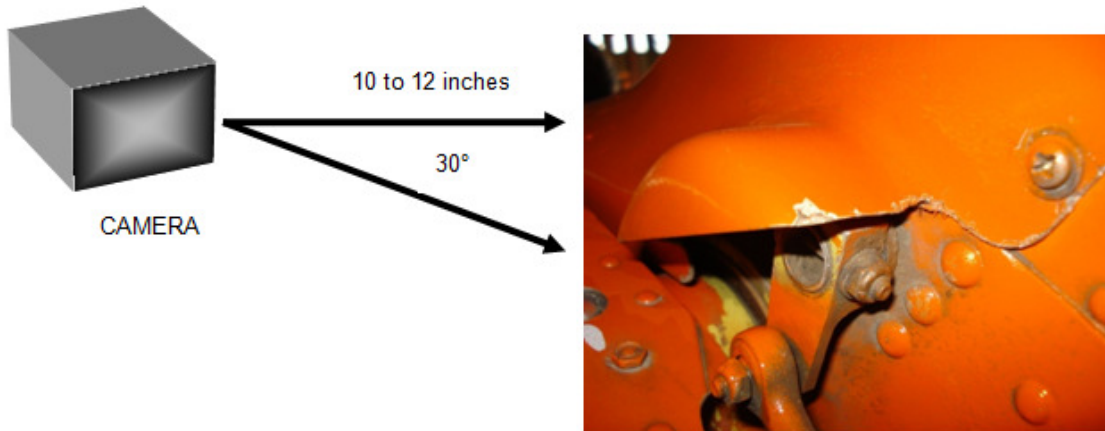
FIGURE 1: Elevator Surface
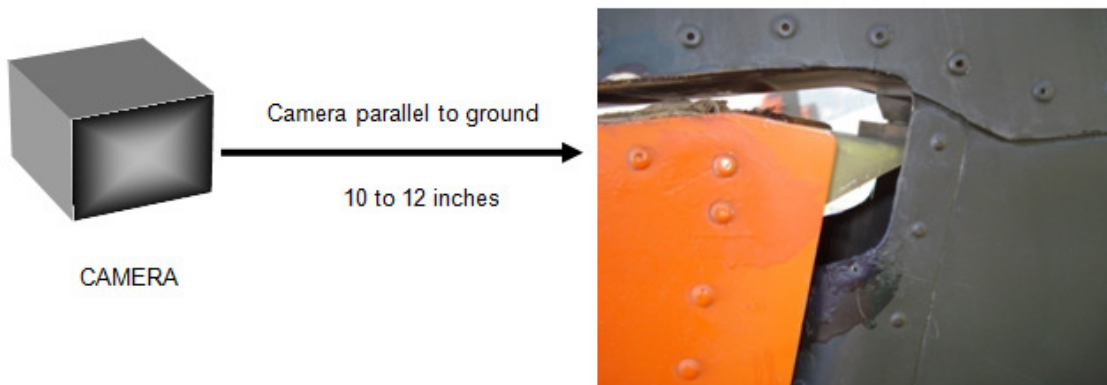
FIGURE 2: Rudder Surface
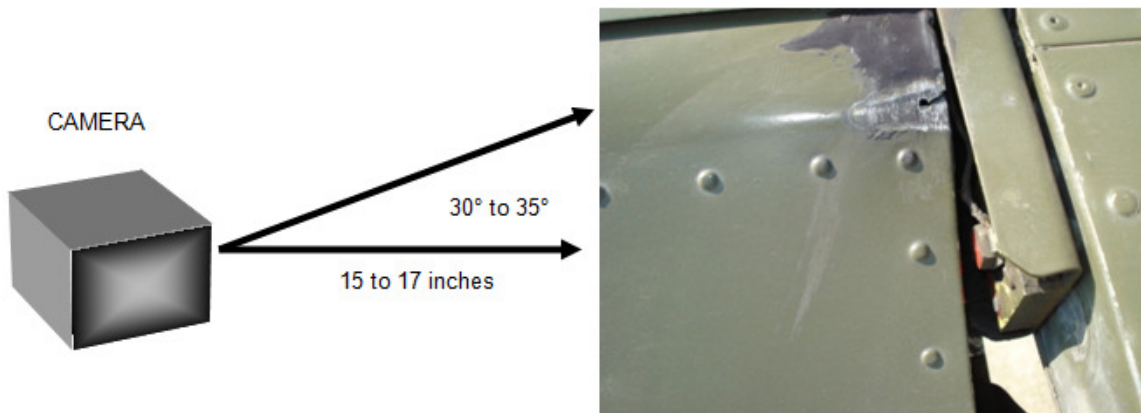
FIGURE 3: Flap Surface

These criteria can be demonstrated by showing different images of aircraft surfaces prone to stress during flight and subsequently generates surface cracks. Figure 1 shows how an aircraft elevator is photographed with suggested angle and distance. Figure 2 shows rudder surface while Figure 3 shows the flap surface. On similar basis photography methods are devised and formulated for wing root, tail cone, engine hot and cold flaps, engine jet pipe liners and nose cone surfaces.

3. AUTOMATIC MARKING OF CRACKS

One of the main task in aiding visual inspection of aircraft is to automatically mark the cracked area of aircraft surface. The automatic marking of marks helps the user to instantaneous get the information of cracked surface from the surface imaging data set area but also to find out which orientation it is following. It helps the inspector to carry out inspection which raises safety issues for the inspector, is time consuming, and suffers at times from being ineffective due to inspector fatigue or boredom [1]. Automatic marking could allow the inspector to safely, quickly and accurately perform the necessary visual inspection.

Automatic marking of cracks have been performed through various techniques. These methods are based on thresholding based, texture based, transform based and classifier based models. The applications of various techniques were due to nature and quality of images which vary basically because of illumination.
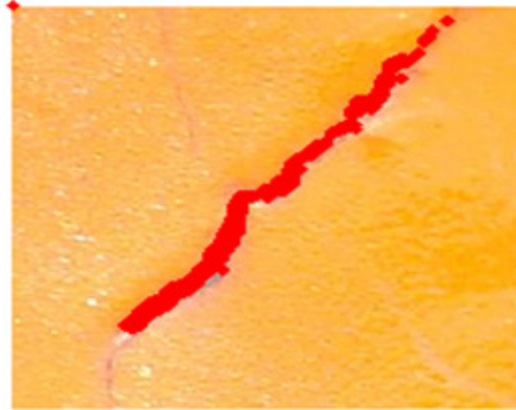
3.1 Thresholding Techniques

Thresholding is an important technique for image segmentation that tries to identify and extract a target from its background on the basis of the distribution of gray levels or texture in image objects. Most Thresholding techniques are based on the statistics of the one-dimensional (1D) histogram of gray levels and on the two dimensional (2D) co-occurrence matrix of an image [5], [6], [7]. Automatic marking of cracks is dramatically affected by changes in the illumination conditions of image captured. In general the variation between images of different cracks captured in the same conditions is smaller than that of the same crack taken in a variety of environments. Two methods are used for thresholding images. These are Otsu Method of Thresholding and Entropy Method of Thresholding.

3.1.1   Otsu Method of Thresholding

Nobuyuki Otsu in 1979 proposed an algorithm for automatic threshold selection from a histogram of image. The algorithm is based on discriminant analysis and uses the zeroth- and the first order cumulative moments of the histogram for calculating the value of the thresholding level. This is an unsupervised method of thresholding. An optimal threshold is selected by the discriminant criterion, namely, so as to maximize the separability of the resultant classes in gray levels [8]. After thresholding the image, the binary image obtained has black background with crack information in white pixels. The coordinates of white pixels were then extracted and were plotted on the original image with a pseudo color as shown in Figure 4.
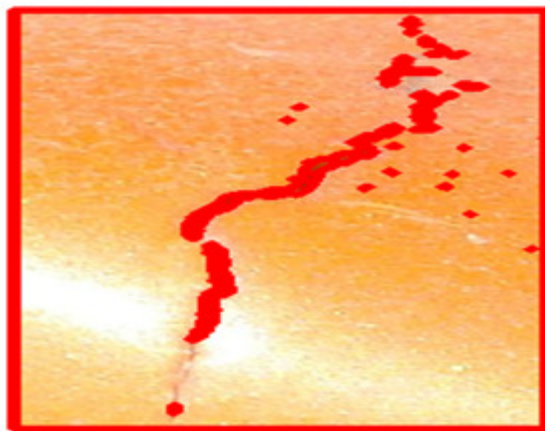
FIGURE 4: Marking of Crack by Otsu Method of Thresholding

3.1.2   Entropy Method of Thresholding

The entropy-based techniques have proven to be successful and reasonably robust [9], however they suffer various limitations and are sensitive to noise. These techniques rely on ontourlet the total entropy of both the object and background regions to find the appropriate threshold. Some of these methods also make use of the pixels' spatial information. The method is very similar to Otsu's method. Rather than maximizing the inter-class variance, it maximizes the inter-class entropy. Entropy is a measure of the uncertainty of an event taking place. It can be calculated it as:

$$S = -\text{sum } (p) \times \log (p) \tag{1}$$

so it is very straightforward to do using the histogram data. P is the probability of a pixel grayscale value in the image, and sum is the Greek capital sigma. It is customary to use log in base 2. Figure 5 is an example of entropy principle of thresholding.

FIGURE 4: Marking of Crack by Entropy Method of Thresholding

3.2 Texture Analysis

The human's capability to distinguish perceptually different textures is difficult to reproduce using machine vision due the variety of textural patterns and illumination conditions. Textures are modeled as a pattern dominated by a narrow band of spatial frequencies and orientations. Textures are used extensively by the human visual system to perform tasks such as the segmentation of scenes into distinct objects and the analysis of surface geometries. Each texture can thus be thought of as containing a narrow range of frequency and orientation components. By filtering the image with band-pass filter tuned to the dominant frequency and orientation component of the textures, it is possible to locate each texture. Therefore Gabor filter was the

best option in this case because Gabor filters are band-pass filters with tune able center frequency, orientation and bandwidth, properly tuned Gabor filters react strongly to specific textures and weakly to all others [10].
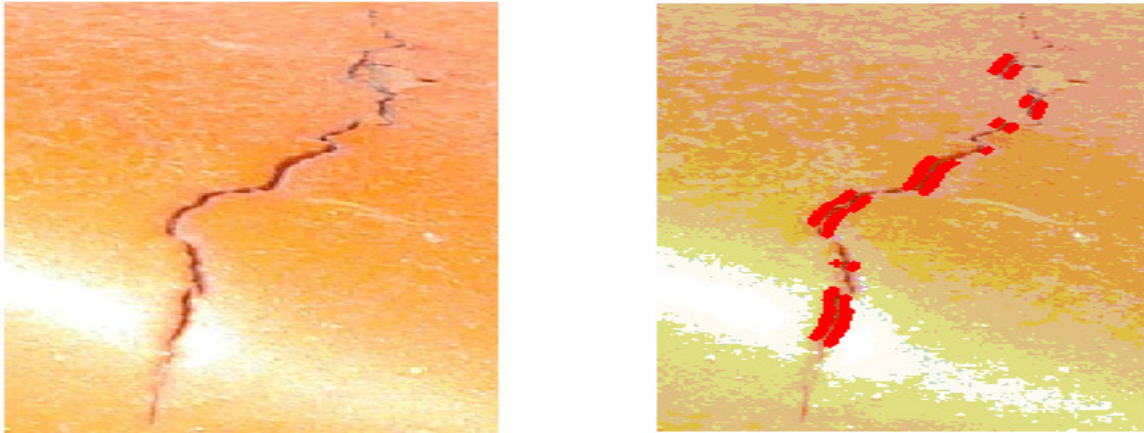
3.2.1   Gabor Filter

Gabor filters are a traditional choice for obtaining ontourle frequency information. They offer the best simultaneous localization of spatial and frequency information. [11] suggests that natural images are better coded by filters that have Gaussian transfer functions when viewed on the logarithmic frequency scale. (Gabor functions have Gaussian transfer functions when viewed on the linear frequency scale). On the linear frequency scale the log-Gabor function has a transfer function as shown in equation 2.

$$G(w) = e^{(-\log(w/w_o)^2) / (2(\log(k/w_o)^2))} \qquad (2)$$

where w_o is the filter's centre frequency. To obtain constant shape ratio filters the term k/w_o must also be held constant for varying w_o. In the frequency domain the even symmetric filter is represented by two real-valued log-Gaussian bumps symmetrically placed on each side of the origin. The odd-symmetric filter is represented by two imaginary valued log-Gaussian bumps anti-symmetrically placed on each side of the origin. The final transfer function is obtained by adding both even and odd transfer functions. Cracks have different orientations and have a set range of frequency. Therefore Gabor filter can give good results after finding orientation through extensive testing. But for every image we have to set the frequency and orientation information separately and that cannot be used by other images. Let us have a look at this image in Figure 4 where Gabor filter gave a good result in marking the crack when set to specific frequency and orientation In Figure 5, when same frequency and orientation was applied, the frequency of Gabor filter does not created any problem but only a mismatch in orientation created abnormal behavior. The crack which has the orientation similar to the crack in Figure 4 was detected rest was washed out.
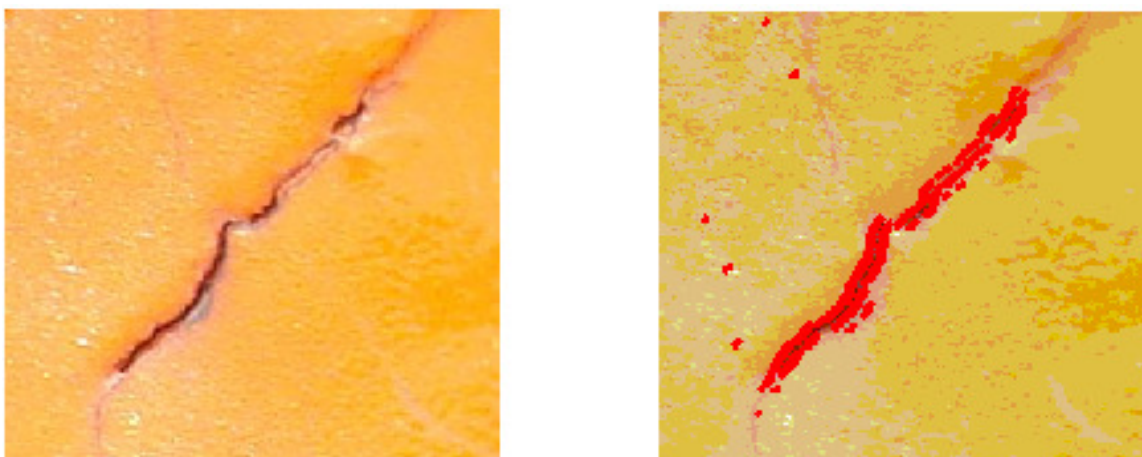
FIGURE 4: Marking of Crack by Gabor Filter

**FIGURE 5:** Marking of Crack by Gabor Filter

3.3 Transform Domain

The frequency spectrum of a signal is basically the frequency components (spectral components) of that signal. The frequency spectrum of a signal shows what frequencies exist in the signal [12]. Intuitively, it is known that the frequency is something to do with the change in rate of something. If something changes rapidly, it is of high frequency, where as if this variable does not change rapidly, i.e., it changes smoothly, it is of low frequency. If this variable does not change at all, it has zero frequency, or no frequency.

Often times, the information that cannot be readily seen in the time-domain can be seen in the frequency domain.By looking at the advantage of frequency information several transforms have been proposed for image signals that have incorporated directionality and multiresolution and hence, could more efficiently capture edges in natural images. Many transform like Wavelet, Contourlet and Nonsubsampled Contourlet are there to provide multiresolution and multidirectional information and each of them of them have their own limitations. In case of Wavelet which is a more general transform have three major flaws. It is shift sensitive because it implies that DWT coefficients fail to distinguish between input- signal shifts [13]. It lacks directional information as natural images contain number of smooth regions and edges with random orientations which affects the optimal representation of natural images. Finally it has absence of phase information which is sensitive to our visual system. Owing to the geometric information, the Contourlet transform achieves better results than discrete wavelet transform in image analysis applications such as denoising and texture retrieval. It is proposed by Minh Do and Martin Vetterli [14] and provides sparse representation at both spatial and directional resolutions. Contourlet transform uses a structure similar to that of curvelets [15]. The pyramidal filter bank structure of the Contourlet transform has very little redundancy, which is important for compression applications. However, Shift sensitivity is an undesirable property because it implies that the transform coefficients fail to distinguish between input- signal shifts. Shift-invariance is desirable in image analysis applications such as edge detection, contour characterization, and image enhancement.
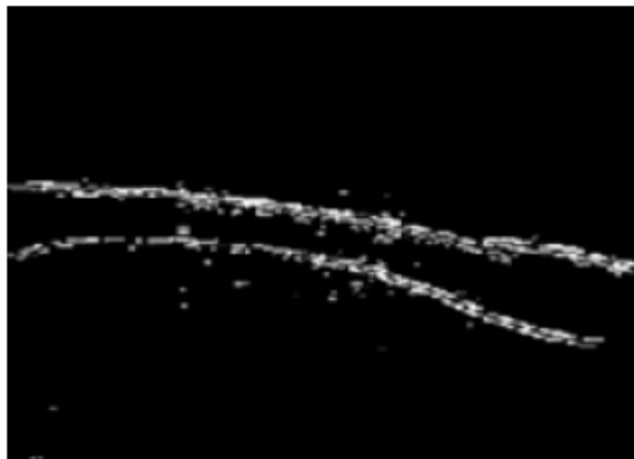
3.3.1 Non Subsampled Contourlet Transform

A shift-invariant version of the Contourlet transform is proposed which is built upon iterated non sub sampled filter banks to obtain a shift-invariant directional multiresolution image representation [16]. Shift invariance is a desirable property in many image processing applications [17]. Due to no down and up sampling NSCT has more redundancy but by allowing redundancy, it is possible to enrich the set of basic functions so that the representation is more efficient in capturing some signal behavior. Figure 6 depicts two channel non subsampled pyramid and directional filter banks used in non sub-sampled contourlet transform. Figure 7 shows how an input image is split into high pass subbands and low pass subbands by a non subsampled pyramid and then a directional filterbank decomposes a high subband into several directional subbands.
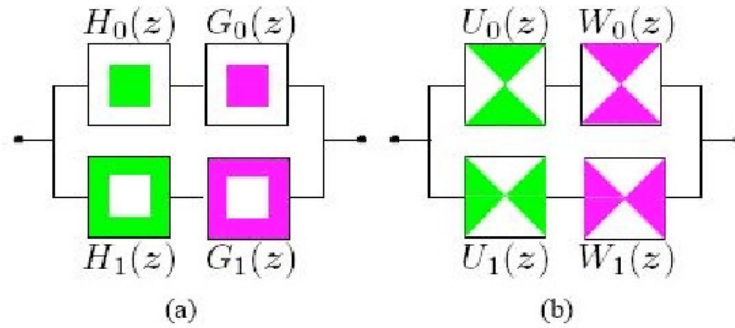
FIGURE 6: (a) Non Subsampled Pyramid (b) Non Subsampled Directional Filter Bank [16]
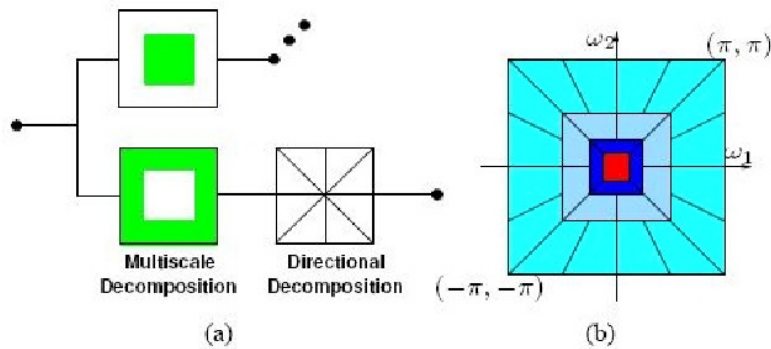
FIGURE 7: The nonsubsampled contourlet transform: (a) Block diagram. First, a nonsubsampled pyramid split the input into a lowpass subband and a highpass subband. Then a nonsubsampled DFB decomposes the highpass subband into several directional subbands. The scheme is iterated repeatedly on the lowpass subband. (b) Resulting frequency division, where the number of directions is increased with frequency [16].

The crack image was decomposed to 3 levels with 8 different directional sub-band. The 3rd level decomposition with 8 different directional sub-band was extracted which contains crack information in different direction. The selected sub-band were added which contains maximum crack information through an automatic thresholding technique devised separately. Automatic thresholding was devised in such a manner that it counts average amount of white pixels in the sub-band which are representing crack and if the white pixels are less than a specified value, which was calculated through extensive testing, they are discarded because they corresponds to noise. Then a union operation is applied between added coefficients of all the sub-bands and selected added coefficients.

FIGURE 8: Marking of Crack by NSCT
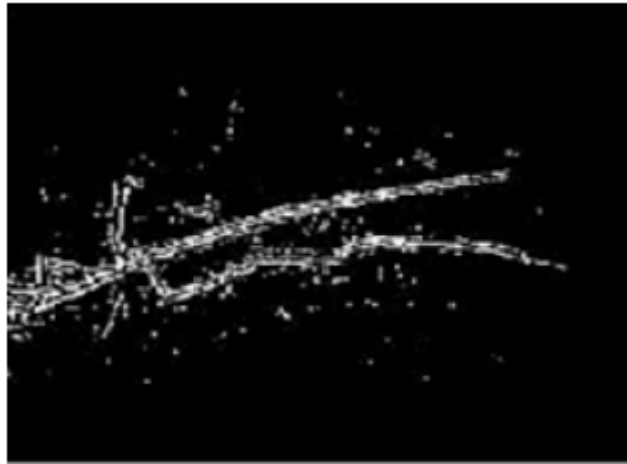
FIGURE 9: Marking of Crack by NSCT

From Figure 8 and Figure 9 it shows that Nonsubsampled Contourlet transform identified the crack in a right fashion and crack was not washed out at any location. NSCT gave finer results for automatically mark the cracked area of aircraft surface.
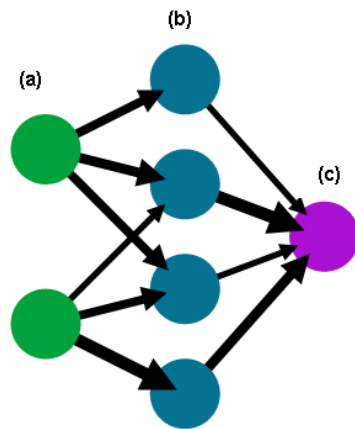
4. DIFFERENTIATING CRACK AND SCRATCH

To address so as how to differentiate between crack and scratch is quite challenging and obviously a difficult task. One must be careful not to evaluate on the basis of how this surface looks like because when the matter comes for aircraft there is no way to take any chance otherwise results could be hazardous and may lead the aircraft to crash. Sometime a surface seems to be scratch, after testing verifies to be crack and vice versa.

In this section a new method of differentiating between crack and scratch is proposed which is totally different from the existing NDI (Non-Destructive Inspection) techniques used in Aviation Industry. The desired surface to be judged as crack or scratch is performed through image processing techniques. Machine vision is used to make a decision whether the input surface is a crack or scratch [18]. The existing methods use in Aviation industry are time consuming and requires a lot of time for maintenance which makes the aircraft to remain ground for a longer period but if the required image of the defected place is available then it will definitely reduce the maintenance time causing the aircraft utilization to its maximum limits.

Three methods have been used to differentiate between crack and scratch. They are (a) Application of Neural Network (b) Energy Calculation (c) Discrete Cosine Transform with Dot Product Classifier
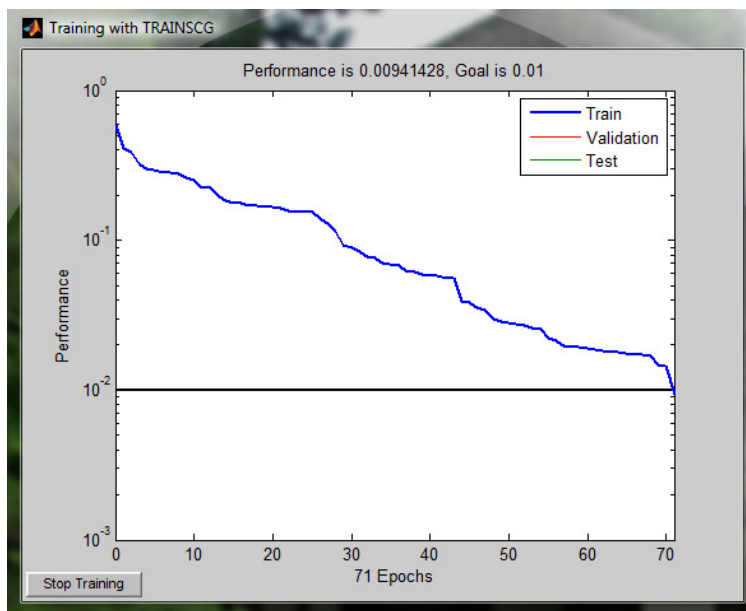
4.1 Application of Neural Network

Artificial neural networks are made up of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons) [19]. Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. A single neuron may be connected to many other neurons and the total number of neurons and connections in a network may be extensive. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. A simple Neural Network is shown in Figure 10.

FIGURE 10: A simple Neural Network (a) Input Layer (b) Hidden Layer (c) Output Layer

4.1.1   Proposed Algorithm

To train the network so as to differentiate whether the incoming image is a scratch or crack, 300 samples of crack and 300 samples of scratch were taken to train the network. The orientation and width of crack and scratch were random. The network was then trained accurately without any saturation. Training was improved by increasing the number of neurons which causes more distance between values so as to reduce overlapping regions, decreasing the Goal so as to minimize false alarm rate (FAR) and increasing number of Epochs which corresponds to number of iterations. Performance curve achieved for this network is shown in Figure 11. Cracks were given a weight of +1 while scratches were given a weight of -1. Training is said to be successful if performance curve touches the goal without any saturation.

FIGURE 11: Performance Curve of Neural Network

4.1.2 Testing and Evaluation

For detailed testing and evaluation of the algorithm another set of 300 cracks and 300 scratch images were taken to validate the method. The algorithm gave 19% FAR for crack and 11% for scratch. Although the FAR was quite high but the process of differentiating between crack and scratch via machine vision was a success.

4.2 Energy Calculation

Another approach used to differentiate between crack and scratch was performed with the help of energy calculation. It was assumed that energy of crack should be higher than that of scratch therefore search to find an optimum threshold was necessary to distinguish both features. Energy method was performed through Contourlet Transform [20].

4.2.1 Proposed Algorithm

The images are decomposed into sub-bands at four different resolution levels. At each resolution level `k' the images are decomposed in 2^n sub-bands where 'n' is the order of the directional filter. The highest resolution level (level 1) corresponds to the actual size of image i.e. 128 x 128. The next resolution level is determined by the expression 2^{N-1} where N in this case is 7. This gives us an image of size 64 x 64 at level 2. Similarly the image is further reduced by subsampling at levels 3 and 4 and generating images of sizes 32 x 32 and 16 x 16 respectively. We have empirically chosen to apply a 5th order filter at resolution level 1 thus giving a total of 32 subbands. By applying a 4th order filter at resolution level 2, 16 subband outputs are obtained. Similarly resolution levels 3 and 4 gave 8 and 4 subbands respectively. Resultantly, 60 valued feature vector is calculated by finding the directional energies in respective sub-bands. Decomposition at three resolution levels is shown in Figure 12.

$E_{k\theta}$, defined as the Energy value in directional sub-band $S_{k\theta}$ at S^{th} resolution level is given by:

$$E_{k\theta} = \Sigma_{S_{k,\theta}} |F_{k,\theta}(x,y) - \overline{F_{k,\theta}}| \qquad (3)$$

where $\overline{F_{k,\theta}}$ is the mean of pixel values of $F_{k,\theta}(x,y)$ in the sub-band $S_{k,\theta}$. $F_{k,\theta}(x,y)$ is the contourlet coefficient value at position (x,y). Additionally, the directional sub-bands vary from 0 to $2^n - 1$. The normalized energy value $\overline{E_{k\theta}}$ of the subband θ at k^{th} resolution level is defined as:

$$\overline{E_{k\theta}} = \frac{E_{k\theta}}{\Sigma_{\theta=0}^{2^n-1} E_{k\theta}} \qquad (4)$$

Taking the constant F_{max} value equal to maximum intensity level of 255, the feature value $F_{k\theta}$ is calculated as:

$$F_{k\theta} = F_{max} \times \overline{E_{k\theta}} \qquad (5)$$

We have evaluated the performance of our proposed algorithm by using the Contourlet Toolbox available at [21]. We used PKVA (Ladder Filter) as the selected filter for our algorithm. Feature vectors are calculated for the images and are stored in a gallery.
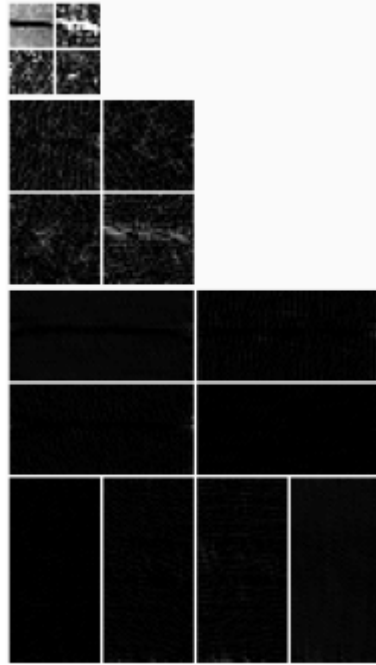
FIGURE 12: Sub band decomposition at three resolution levels

4.2.2 Experimental Results

Proposed algorithm has been implemented in MATLAB on a 1.5 GB RAM, 1.67 GHz Intel Core Duo processor PC. The data set of 600 images of crack and 600 images of scratch were divided into two parts. Therefore, out of 600 images of crack and scratch each, 300 images were used for the purpose of training respectively. These training images were subjected to the Contourlet transform as described in the previous section and their feature vectors were stored separately. The rest 300 images of crack and 300 images of scratch were used for the purpose of validation. These test images were subjected to Contourlet Transform and their feature vectors were passed through the Dot Product Classifier with the feature vectors of training images. The dot product giving the highest result with the training images was finalized to give decision of the crack or scratch. It was observed that the directional energy components of the image of crack and scratch are highly overlapping resulting in classification errors as depicted in Figure 13. This method was able to identify 200 images of scratch and 225 images of crack out of 300 images of crack and scratch respectively.

4.3 Discrete Cosine Transform

Discrete Cosine Transform packs image into its low frequency components [22]. DCT has many applications in the field of Image processing. It bears the property of de-correlation, energy compaction, seperability which means that 1-D DCT can be applied to rows and then columns of an image. DCT has vast applications in the field of feature extraction and pattern recognition. Two dimensional DCT is defined as:

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)v}{2N} \qquad (6)$$

For u,v = 0,1,2,3.........N-1 and $\alpha(u)$ and $\alpha(v)$ are defined as

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} \ for\ u = 0 \\ \sqrt{\frac{2}{N}} \ for\ u \neq 0 \end{cases} \qquad (7)$$
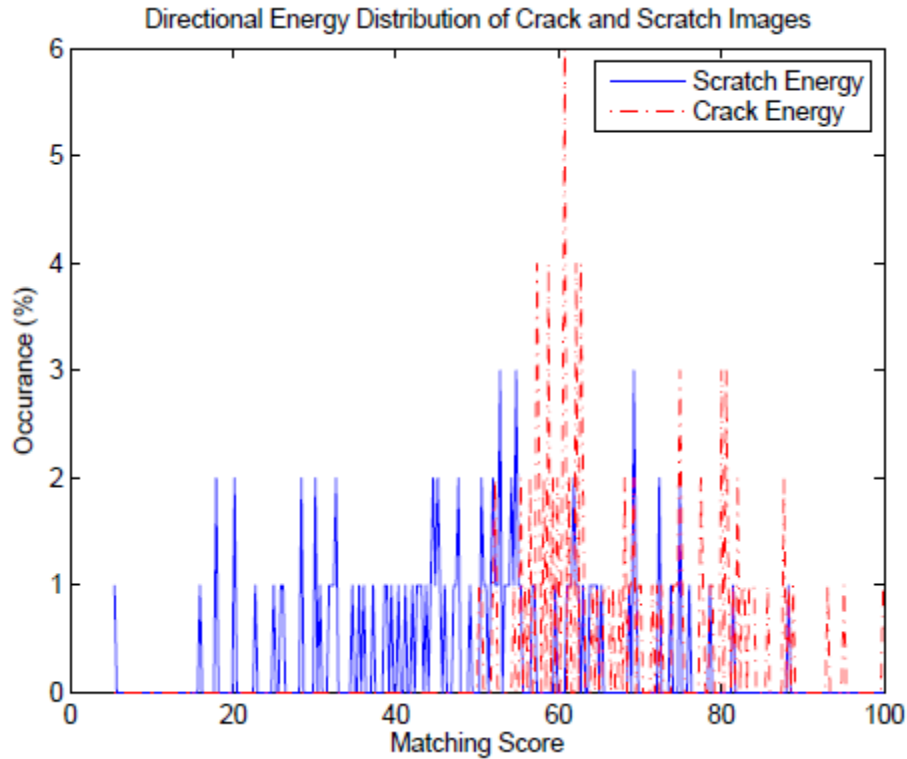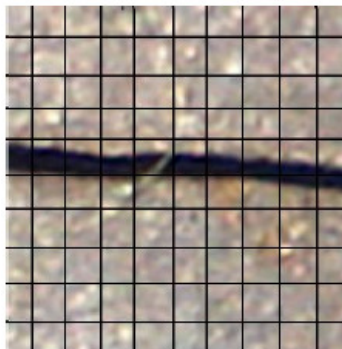
FIGURE 13: Directional Energy distribution of Sub bands

4.3.1 Application of DCT

The images that were previously used for training the CT were utilized for feature extraction by the DCT. Each image of crack and scratch was decomposed into 100 segments each of size 10 X 10 pixels as shown in Figure 14.

FIGURE 14: Division of Image into 100 pieces

The DCT of individual segment is calculated and values near or equal to zero are discarded. Standard deviations for the rest of the coefficient values were calculated. This procedure is applied to 100 segments of an image to generate a feature vector of 100 length.

4.3.2 Experimental Results

The size of the database of the crack and scratch images for the purpose of training and validation remains same as that of Contourlet Transform. Each of the image of crack and scratch used for the purpose of training was passed through DCT as described in the previous section

and its feature vector was stored separately. Similarly the images used for evaluation were subjected to DCT and their feature vectors were passed through the Dot Product Classifier with the feature vectors of training images. The dot product with the highest result was finalized to give decision of the crack or scratch. The DCT was able to identify 285 images of crack and 292 images of scratch out of 300 images of crack and scratch respectively leading to an accuracy of 95% and 97.3% respectively.

4.4 Combination of Transforms

To further enhance our approach, the feature vectors obtained from both the transforms of the training images were concatenated and stored in a separate gallery. The images used for validation were first passed through the Contourlet Transform and Discrete Cosine Transform as described in previous sections and their concatenated feature vectors were subjected to the Dot product classifier with rest of the feature vectors of the training images. The dot product with the highest result was observed with the training images to finalize the decision of crack or scratch. This technique resulted in better identification result, giving a higher recognition rate i.e. 96.6% for the crack surfaces and 98.3% for the scratch surfaces. This method identifies 290 images of crack and 295 images of scratch out of 300 images of crack and scratch surfaces respectively. Table 1 summarizes the results for the three approaches.

Test	CT	DCT	Combination of Transforms
Crack Images Identified	225	285	290
Scratch Images Identified	200	292	295
Recognition Rate of Crack Images (%)	75	95	96.6
Recognition Rate of Scratch Images (%)	66.66	97.3	98.3

TABLE 1: Comparison of Three approaches on a database of 300 crack and 300 scratch aircraft surfaces

5. CONCLUSION

The methods described depict intelligent utilization of Digital Image Processing for specified application. The proposed computer aided techniques are not only power efficient but they also consume less amount of time for maintenance. The inspection through image processing does not have the element of fatigue or boredom. Enhancement of aircraft imagery helps the user to have a broader view of image that have been either degraded due to poor lighting conditions, blurness or addition of noise. A new visual method is devised to differentiate between crack and scratch to supplement existing NDI techniques. The differentiation between crack and scratch can be improved by varying the parameters of Neural Network or by increasing the data set of images for training the Neural Network or DCT with Dot product classifier. The proposed technique was capable of differentiating crack and scratch with 96.6% accuracy for crack and 98.3% accuracy for scratch. Suggestion are included to have surface imaging in existing periodic inspection for better record keeping and trend analysis of aircrafts. Various critical areas of aircraft that are susceptible to cracks were identified and were photographed with suggested scale and angle. Surface imaging along with automatic marking of cracks helps in analyzing the stress and strain analysis of aircrafts thus identifying aircrafts which requires more maintenance time. These methods in general reduce maintenance time thus ensuring maximum utilization of aircraft for flying.

6. REFERENCES

[1] Gunatilake P., Siegel M., Jordan A., and Podnar G. Image Understanding Algorithms for Remote Visual Inspection of Aircraft Surfaces, In: Machine Vision Applications in Industrial inspection V, page numbers (2-13), 1997

[2] Siegel M. and Gunatilake P. Enhanced Remote Visual Inspection of Aircraft Skin, In: Proc. Intelligent NDE Sciences for Aging and Futuristic Aircraft Workshop, page numbers (101-112), 1997

[3]   Siegel M., Gunatilake P. and Podnar. Robotic assistants for Aircraft Inspectors, In: Proc. IEEE Instrumentation and Measurement Magazine, Vol 1, page numbers (16-30), 1998

[4]   Alberts C J, Carroll C W ,Kaufman W M , Perlee C J , and Siegel M W . Automated Inspection of Aircraft, In: Technical report, no. DOT/FAA/AR-97/69, Carnegie Mellon Research Institute, Pittsburgh, PA 15230-2950, USA, 1998

[5]   Liao P. S., T. S. Chen and P. C. Chung. A Fast Algoritm for Multi Level Thresholding, In: Journal of Information Science and Engineering 17, page numbers (713-727), 2001

[6]   Lee S. U. and S. U. Chung. A Comparitive Performance Study of Several Global Thresholding Techniques for Segmantation, In: Computer Vision Graphics Image Processing, Volume 52, page numbers (171-190), 1990

[7]   Tsai D. M. and Chen Y. H. A Fast Histogram Clustering Approach for Multi Level Thresholding, In: Pattern Recognition Letters, Volume 13, Number 4, page numbers (245-252), 1992

[8]   Otsu Nobuyuki. A Threshold Selection Method for Gray Level Histogram, In: IEEE Transaction on System, Man and Cybernetics, Volume SMC-9, Number 1, 1979

[9]   Kapur J. N., P. K. Sahoo and A. K. C. Wong. A New Method for Gray-level Picture Thresholding using Histogram, In: Computer Vision, Graphics and Image Processing, Volume 29, Issue 3, page numbers (273-285), 1985

[10]   Hammouda Khaled. Texture Segmentation using Gabor Filters, In: IEEE Journal "Transform", Volume 26, Issue 6, page numbers (1-8), 2000

[11]   Field D. J. Relation between the Statics of Natural Images and Response Properties of Cortical Cells, In: Journal Optical Society of America, page numbers (2379-2394), 1987

[12]   Cherng Shen. The Analysis of Osteoblast Cellular Response to the reaction of Electromagnetic Field at 2.4 GHz, In: Jounal of American Science, page numbers (48-50), 2005

[13]   Ceylan M., Ceylan R., Ozbay Y. and Kara S. Application of Complex Discrete Wavelet Transform in Classification of Doppler Signals using Complex-Valued Artificial Neural network, In: Artifical Intelligence in Machines, Volume 44, Issue 1, page numbers (65-76), 2008

[14]   Do M. N. Multi Resolution Image Representation, PhD Thesis EPFL, Lausanne, Switzerland, 2001

[15]   Emmanuel J. Candes and David L. Donoho. Curvelets-A surprisingly effective non-adaptive representation for objects with Edges, In: Curve and Surface Fitting, publisher: Vanderbilt Univ. Press, Nashville, TN, 1997

[16]   Cunha A. L., Zhou J.,Do M. N. The Nonsubsampled Contourlet Transform: Theory, Design and Applications, In: IEEE Transaction on Image Processing, 2005

[17]   Simoncelli E. P., Freeman W. T., Adelson E. H., Heeger D. J. Shiftable Multiscale Transforms, In: IEEE Transaction on Information Theory, Volume 38, number. 2, page numbers (587–607), 1992

[18]   Mumtaz M., Mansoor Atif B. and Masood H. A New Approach to Aircraft Surface Inspection Based on Directional Energies of Texture, In: International Conference on Pattern Recognition, ISSN: 1051-4651, page numbers (4404-4407), 2010

[19]   Artificial   Neural   Network   Available   at   weblink: http://www.usegnu.net/projects/files/ANN_Project.pdf

[20]   Do M. N., Vetterli M. Contourlets In: Beyond wavelet, J. Stoeckler and G.V. Welland, Eds. Academic Press, New, 2003

[21]   Do M. N. (2003): Available at weblink: http://www.ifp.uiuc.edu/~minhdo/software/

[22]   Rao, Kamisetty Ramamohan and Yip, P. Discrete Cosine Transform: Algorithms, Advantages, Applications, In: Academic Press, ISBN-13: 9780125802031 ISBN: 012580203X, NV, USA, 1990

Data-Driven Motion Estimation with Spatial Adaptation

Alessandra Martins Coelho   *alessandra.coelho@ifsudestemg.edu.br*
Instituto Federal de Educacao, Ciencia e
Tecnologia do Sudeste de Minas Gerais
(IF SEMG),Rio Pomba, MG, Brazil

Vania Vieira Estrela   *vestrela@id.uff.br*
Departamento de Telecomunicacoes,
Universidade Federal Fluminense (UFF),
Niteroi, RJ, Brazil

Abstract

Besides being an ill-posed problem, the pel-recursive computation of 2-D optical flow raises a wealth of issues, such as the treatment of outliers, motion discontinuities and occlusion. Our proposed approach deals with these issues within a common framework. It relies on the use of a data-driven technique called Generalized Cross Validation (GCV) to estimate the best regularization scheme for a given moving pixel. In our model, a regularization matrix carries information about different sources of error in its entries and motion vector estimation takes into consideration local image properties following a spatially adaptive. Preliminary experiments indicate that this approach provides robust estimates of the optical flow.

Keywords: Motion Estimation, Generalized Cross Validation, Video Processing, Computer Vision, Regularization.

1. INTRODUCTION

Motion estimation is very important in multimedia video processing applications. For example, in video coding, the estimated motion is used to reduce the transmission bandwidth. The evolution of an image sequence motion field can also help other image processing tasks in multimedia applications such as analysis, recognition, tracking, restoration, collision avoidance and segmentation of objects [6, 7, 10].

In coding applications, a block-based approach [7] is often used for interpolation of lost information between key frames. The fixed rectangular partitioning of the image used by some block-based approaches often separates visually meaningful image features. If the components of an important feature are assigned different motion vectors, then the interpolated image will suffer from annoying artifacts. Pel-recursive schemes [2,3,6] can theoretically overcome some of the limitations associated with blocks by assigning a unique motion vector to each pixel. Intermediate frames are then constructed by resampling the image at locations determined by linear interpolation of the motion vectors. The pel-recursive approach can also manage motion with subpixel accuracy. The update of a motion estimate is based on the minimization of the displaced frame difference (DFD) at a pixel. In the absence of additional assumptions about the pixel motion, this estimation problem becomes "ill-posed" because of the following problems: a) occlusion; b) the solution to the 2-D motion estimation problem is not unique (aperture problem); and c) the solution does not continuously depend on the data due to the fact that motion estimation is highly sensitive to the presence of observation noise in video images.

We propose to solve optical flow (OF) problems by means of a framework that combines the Generalized Cross Validation (GCV) and a regularization matrix Λ. Such approach accounts better for the statistical properties of the errors present in the scenes than the solution proposed by Biemond [1] where a scalar regularization parameter was used.

Alessandra Martins Coelho & Vania Vieira Estrela

We organized this work as follows. Section 2 provides some necessary background on the pel-recursive motion estimation problem. Section 3 introduces our spatially adaptive approach. Section 4 describes the Ordinary Cross Validation. Section 5 deals with the GCV technique. Section 6 defines the metrics used to evaluate our results. Section 7 describes the experiments used to access the performance of our proposed algorithm. Finally, Section 8 presents some conclusions.

2. PEL-RECURSIVE DISPLACEMENT ESTIMATION

2.1. Problem Characterization

The displacement of a picture element (pel) between adjacent frames forms the displacement vector field (DVF) and its estimation can be done using at least two successive frames. The DVF is the 2-D motion resulting from the apparent motion of the image brightness (OF) where a displacement vector (DV) is assigned to each image pixel.

A pixel belongs to a moving area if its intensity has changed between consecutive frames. Hence, our goal is to find the corresponding intensity value $I_k(r)$ of the k-th frame at location $r = [x, y]^T$, and $d(r) = [d_x, d_y]^T$ the corresponding (true) DV at the working point r in the current frame. Pel-recursive algorithms minimize the DFD function in a small area containing the working point assuming constant image intensity along the motion trajectory. The DFD is defined by

$$\Delta(r; d(r)) = I_k(r) - I_{k-1}(r - d(r)) \tag{1}$$

and the perfect registration of frames will result in $I_k(r) = I_{k-1}(r - d(r))$. The DFD represents the error due to the nonlinear temporal prediction of the intensity field through the DV. The relationship between the DVF and the intensity field is nonlinear. An estimate of $d(r)$, is obtained by directly minimizing $\Delta(r, d(r))$ or by determining a linear relationship between these two variables through some model. This is accomplished by using the Taylor series expansion of $I_{k-1}(r - d(r))$ about location $(r - d^i(r))$, where $d^i(r)$ represents a prediction of $d(r)$ in i-th step. This results in

$$\Delta(r, r - d^i(r)) = -u^T \nabla I_{k-1}(r - d^i(r)) + e(r, d(r)), \tag{2}$$

where the displacement update vector $u = [u_x, u_y]^T = d(r) - d^i(r)$, $e(r, d(r))$ represents the error resulting from the truncation of the higher order terms (linearization error) and $\nabla = [\partial/\partial_x, \partial/\partial_y]^T$ represents the spatial gradient operator. Applying Eq. (2) to all points in a neighborhood R containing N pixels gives

$$z = Gu + n, \tag{3}$$

where the temporal gradients $\Delta(r, r - d^i(r))$ have been stacked to form the $N \times 1$ observation vector z containing DFD information on all the pixels in R, the $N \times 2$ matrix G is obtained by stacking the spatial gradient operators at each observation, and the error terms have formed the $N \times 1$ noise vector n which is assumed Gaussian with $n \sim N(0, \sigma_n^2 I)$. Each row of G has entries

$\left[g_{xi}, g_{yi} \right]^T$, with $i = 1, ..., N$. The spatial gradients of I_{k-1} are calculated through a bilinear interpolation scheme [2].

2.2. Regularized Least-Squares Estimation

The pel-recursive estimator for each pixel located at position r of a frame can be written as

$$d^{i+1}(r) = d^i(r) + u^i(r) \tag{4}$$

where $u^i(r)$ is the current motion update vector obtained through a motion estimation procedure that attempts to solve Eq. (3), $d^i(r)$ is the DV at iteration i and $d^{i+1}(r)$ is the corrected DV. The regularized minimum norm solution to the previous expression, that is

$$\hat{u}(\Lambda) = \hat{u}_{RLS}(\Lambda) = (G^T G + \Lambda)^{-1} G^T z \tag{5}$$

is also known as Regularized Least-Squares (RLS) solution. In order to improve the RLS estimate of the motion update vector, we propose a strategy which takes into consideration the local properties of the image. It is described in the next section.
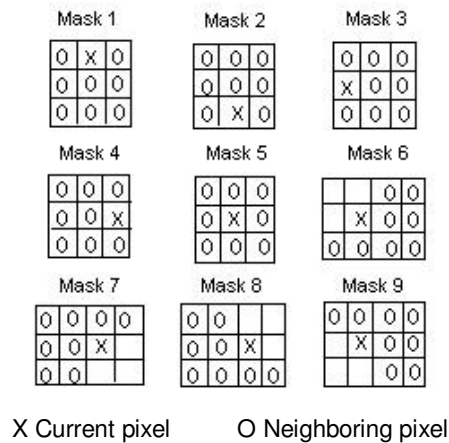

X Current pixel   O Neighboring pixel

FIGURE 1: Neighborhood geometries.

3. SPATIALLY ADAPTIVE NEIGHBORHOODS

Aiming to improve the estimates given by the pel-recursive algorithm, we introduced an adaptive scheme for determining the optimal shape of the neighborhood of pixels with the same DV used to generate the overdetermined system of equations given by Eq. (3). More specifically, the masks in Fig. 1 show the geometries of the neighborhoods used.

Errors can be caused by the basic underlying assumption of uniform motion inside R (the smoothness constraint), by not grouping pixels adequately, and by the way gradient vectors are estimated, among other things. It is known that in a noiseless image containing pixels in textured areas most errors, when estimating motion occur close to motion boundaries. This information leads to a hypothesis testing (HT) approach to determine the most appropriate neighborhood shape for a given pixel. The best neighborhood from the finite set of templates shown in Fig. 1, according to the smallest $|DFD|$ criterion, in an attempt to adapt the model to local features associated to motion boundaries.

4. ORDINARY CROSS VALIDATION (OCV)

Cross validation has been proven to be a very effective method of estimating the regularization parameters [4,5,8], which in our work are the entries of Λ, without any prior knowledge on the noise statistics. The degree of smoothing of the solution $\hat{u}(\Lambda)$, in Eq. (5), is dictated by the regularization matrix Λ.

OCV divides the data into two disjoint subsets obtained from the original observations: an estimation/prediction set and a validation set. In the context of neural networks, the former is called "training set". Let M be the number of observations used for the validation set, where $M \geq 1$, and N be the total number of observations. In our particular problem, N is the mask size and z is the entire observation set, that is an N-dimensional measurement vector. For each set $j = 1, ..., M$, where j is the size of the validation set, the minimum mean-square error (MSE) is calculated using the left out data set, that is, the remaining $(N - j)$ observations, and varying the regularization matrix Λ. In other words, for each value of j, a corresponding set with $(N - j)$ elements is used to predict the j data points left out.

OCV is defined as the average of all the MSE's evaluated over all possible $\binom{N}{j}$ combinations of validation sets. For the case $j = 1$, that is the validation set has only one element, the OCV or prediction MSE is given by

$$OCV(\Lambda) = \frac{1}{N} \sum_{i=0}^{N-1} \left[z_i - \hat{u}_i \right]^2 ,$$

where z_i is the i-th entry of the observation vector z, $i = 0, ..., (N-1)$ as follows: $z = \begin{bmatrix} z_0 \\ M \\ z_i \\ M \\ z_{N-1} \end{bmatrix}$,

z_{-i} the vector obtained after making the i-th entry of z equal to zero, that is $z_{-i} = \begin{bmatrix} z_0 \\ M \\ z_{i-1} \\ 0 \\ z_{i+1} \\ M \\ z_{N-1} \end{bmatrix}$, and

$\hat{u}_i = [G(G^T G + \Lambda)^{-1} G^T z_{-i}]_i$ is the estimate of point z_i using vector z_{-i}.

The previous OCV equation averages all the MSE's obtained by leaving each of the entries of z out. Therefore, the data division into validation and estimation/prediction sets is done in an alternate fashion. All the data is used for both purposes. This technique is also called predictive sample reuse or leave-one-out principle. The idea behind OCV is to perform a data-driven consistency check that, essentially, measures the adequacy of a parameter set via the model ability to predict some of the observations based on the other ones. Hence, the optimum value of the regularization parameters for N samples is the one that minimizes the mean-square error $OCV(\Lambda)$. The previous expression has to be further manipulated in order to express the OCV in terms of Λ. The optimum Λ, that is $\hat{\Lambda}$, is the one that minimizes the following function:

$$OCV(\Lambda)=\frac{1}{N}\|H(\Lambda)\{I-A(\Lambda)\}z\|^2 ,$$

where $\quad H(\Lambda)=diag\left\{\dfrac{1}{1-\{g_0^T B^{-1} g_0\}},L,\dfrac{1}{1-\{g_{N-1}^T B^{-1} g_{N-1}\}}\right\} ,$

$A(\Lambda)=G[G^T G+\Lambda]^{-1}G^T$, and $\quad B=B(\Lambda)=(G^T G+\Lambda)$.

The main advantage of the OCV is its systematic way of determining the regularization parameter directly from the observed data. However, it presents the following drawbacks:

(i) It uses a noisy performance measure, Mean Squared Error (MSE). This means that since we are looking at the average value of the MSE over several observation sets re-sampled from the original z , we can only guarantee the OCV estimator of Λ is going to be a good predictor when $N\gg 1$.

(ii) It treats all data sets equally. In terms of image processing, we expect close neighbors of the current pixel to behave more similarly to it (in most of the cases) than pixels that are more distant from it. Of course, this is not the case with motion boundaries, occlusion and transparency.

5. THE GENERALIZED CROSS VALIDATION (GCV)

The OCV does not provide good estimates of Λ [5, 8]. A modified method called GCV function gives more satisfactory results. GCV is a weighted version of the OCV, and it is given by

$$GCV(\Lambda)=\frac{1}{N}\sum_{i=1}^{N}[z_i-\hat{z}_i]^2 w_i(\Lambda) , \tag{6}$$

where the weights w_i are defined as follows:

$$w_i(\Lambda)=\left\{\frac{[1-a_{ii}(\Lambda)]}{\left(1-\frac{1}{N}Tr[A(\Lambda)]\right)}\right\}^2 , \text{ and} \tag{7a}$$

$$A(\Lambda)=G\left(G^T G+\Lambda\right)^1 G^T \tag{7b}$$

with $a_{ii}(\Lambda)$ being the diagonal entries of matrix $A(\Lambda)$ as defined in Eq. (7b). The main shortcoming of OCV is the fact that OCV is not invariant to orthonormal transformations. In other words, if data $z'=\Gamma z$ is available, where Γ is an $N\times N$ orthonormal matrix, and z' is the observation vector corresponding to the linear model given by

$$z'=G'u'+n'=\Gamma\{Gu+n\} . \tag{8}$$

Therefore, the *OCV*, and, consequently the regularization matrix Λ, depends on Γ. *GCV* on the other hand is independent of Γ. Thus, the $GCV(\Lambda)$ is a better criterion for estimating the regularization parameters [5, 8]. So, Eq. (6) can also take the form

$$GCV(\Lambda) = \frac{1}{N} \frac{\left\| [\mathbf{I} - \mathbf{A}(\Lambda)]\mathbf{z} \right\|^2}{\left[\dfrac{1}{N} Tr\{\mathbf{I} - \mathbf{A}(\Lambda)\} \right]^2} \ . \tag{9}$$

5.1 Regularization Matrix Determination

The *GCV* function for the observation model in Eq. (3) is given in closed form by Eq. (9). Let us call $\hat{u}_{gcv}$ the solution for Eq. (3) when an optimum parameter set (the entries of the regularization matrix) Λ_{GCV} is found by means of the GCV. Then, Eq. (5) becomes

$$\hat{u}_{\mathrm{GCV}} = \left(G^T G + \Lambda_{GCV} \right)^{-1} G^T z \tag{10}$$

5.2 The GCV-based Estimation Algorithm

For each pixel located at $r = (x, y)$ the GCV-based algorithm is given by the following steps:

1) Initialize the system: $d°(r)$, $m \leftarrow 0$ (m = mask counter), and $i \leftarrow 0$ (i = iteration counter).
2) If $|DFD| < T$, then stop. T is a threshold for $|DFD|$.
3) Calculate G^i and z^i for the current mask and current initial estimate.
4) Calculate Λ^i by minimizing the expression

$$GCV(\Lambda^i) = \frac{1}{N} \frac{\left\| \left[I - A(\Lambda^i) \right] z^i \right\|^2}{\left[\dfrac{1}{N} Tr\{ I - A(\Lambda^i) \} \right]^2} \ , \tag{11}$$

where

$$A(\Lambda^i) = G^i \left[\left(G^i \right)^T G^i + \Lambda^i \right]^{-1} \left(G^i \right)^T \ . \tag{12}$$

5) Calculate the current update vector:

$$u^i = \left[\left(G^i \right)^T G^i + \Lambda^i \right]^{-1} \left(G^i \right)^T z^i \ . \tag{13}$$

6) Calculate the new DV:

$$d^{i+1}(r) = d^i(r) + u^i(r) \tag{14}$$

7) For the current mask m :

If $\left\| d^{i+1}(r) - d^i(r) \right\| \le \varepsilon$ and $|DFD| < T$, then stop.

If $i < (I-1)$ where I is the maximum number of iterations allowed, then go to step 3 with $i \leftarrow i+1$ and use $d^i(r) \leftarrow d^{i+1}(r)$ as the new initial estimate.

Otherwise, try another mask: $m \leftarrow m+1$. If all masks where used and no DV was found, then set $d^{i+1}(r) = 0$.

(a)   (b)   (c)

FIGURE 2: Frames from the video sequences used for tests: (a) synthetic frame, (b) Mother and Daughter and (c) Foreman

6. METRICS TO EVALUATE THE EXPERIMENTS
The motion field quality is accessed using the four metrics [2, 3] described below applied to the video sequences shown in Fig. 2.

6.1 Mean Squared Error (MSE)
Since the MSE provides an indication of the degree of correspondence between the estimates and the true value of the motion vectors, we can apply this measure to two consecutive frames of a sequence with known motion. We can evaluate the MSE in the horizontal (MSE_x) and in the vertical (MSE_y) directions as follows

$$MSE_x = \frac{1}{RC}\sum_{r \in S}[d_x(r) - \hat{d}_x(r)]^2 \text{, and} \tag{15}$$

$$MSE_y = \frac{1}{RC}\sum_{r \in S}[d_y(r) - \hat{d}_y(r)]^2 \text{,} \tag{16}$$

where S is the entire frame, r represents the pixel coordinates, R and C are, respectively, the number of rows and columns in a frame, $d(r) = (d_x(r), d_y(r))$ is the true DV at r, and $\hat{d}(r) = (\hat{d}_x(r), \hat{d}_y(r))$ its estimation.

6.2 Bias
The bias gives an idea of the degree of correspondence between the estimated motion field and the original optical flow. It is defined as the average of the difference between the true DV's and their predictions, for all pixels inside a frame S, and it is defined along the x and y directions as

$$bias_x = \frac{1}{RC}\sum_{r \in S}[d_x(r) - \hat{d}_x(r)] \tag{17}$$

and

$$bias_y = \frac{1}{RC}\sum_{r \in S}[d_y(r) - \hat{d}_y(r)] \text{.} \tag{18}$$

Ean-squared Displaced Frame Difference

This metric evaluates the behavior of the average of the squared displaced frame difference ($\overline{DFD}^2$). It represents an assessment of the evolution of the temporal gradient as the scene evolves by looking at the squared difference between the current intensity $I_k(r)$ and its predicted value $I_{k-1}(r-d(r))$. Ideally, the $\overline{DFD}^2$ should be zero, which means that all motion was identified correctly ($I_k(r) = I_{k-1}(r-d(r))$ for all r's). In practice, we want the $\overline{DFD}^2$ to be as low as possible. Its is defined as

$$\overline{DFD}^2 = \frac{\sum_{k=2}^{K}\sum_{r\in S}[I_k(r)-I_{k-1}(r-d(r))]^2}{RC(K-1)} , \tag{19}$$

where K is the length of the image sequence.

6.3 Improvement in Motion Compensation

The average improvement in motion compensation $\overline{IMC}(dB)$ between two consecutive frames is given by

$$\overline{IMC}_k(dB) = 10\log_{10}\left\{\frac{\sum_{r\in S}[I_k(r)-I_{k-1}(r)]^2}{\sum_{r\in S}[I_k(r)-I_{k-1}(r-d(r))]^2}\right\} , \tag{20}$$

where S is the frame being currently analyzed. It shows the ratio in decibel (dB) between the mean-squared frame difference ($\overline{FD}^2$) defined by

$$\overline{FD}^2 = \frac{\sum_{r\in S}[I_k(r)-I_{k-1}(r)]^2}{RC} , \tag{21}$$

and the $\overline{DFD}^2$ between frames k and $(k-1)$.

As far as the use of the this metric goes, we chose to apply it to a sequence of K frames, resulting in the following equation for the average improvement in motion compensation:

$$\overline{IMC}(dB) = 10\log_{10}\left\{\frac{\sum_{k=2}^{K}\sum_{r\in S}[I_k(r)-I_{k-1}(r)]^2}{\sum_{k=2}^{K}\sum_{r\in S}\left[I_k(r)-I_{k-1}(r-d(r))\right]^2}\right\} . \tag{22}$$

When it comes to motion estimation, we seek algorithms that have high values of $\overline{IMC}(dB)$. If we could detect motion without any error, then the denominator of the previous expression would be zero (perfect registration of motion) and we would have $\overline{IMC}(dB) = \infty$.
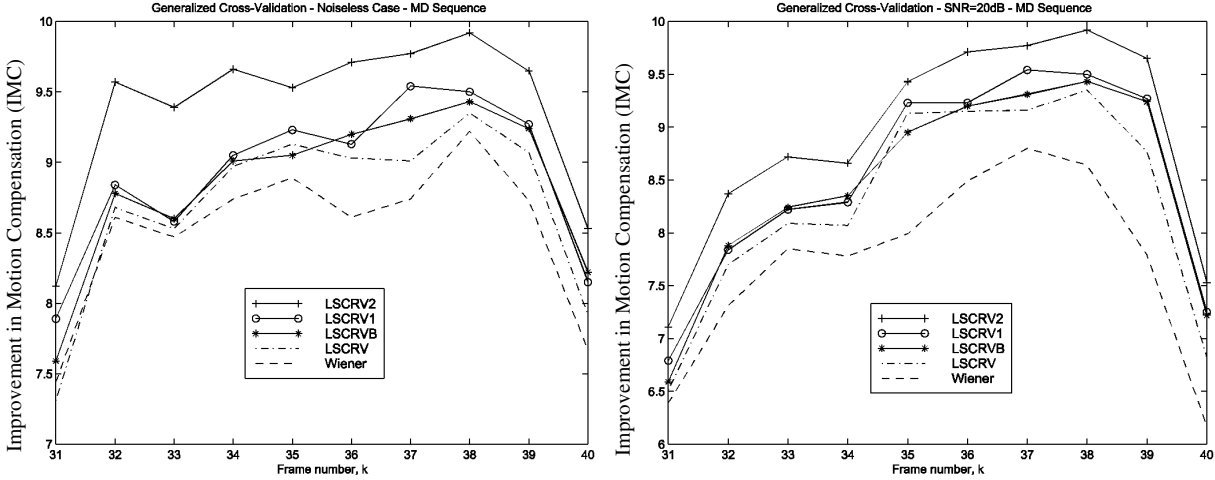
FIGURE 3: $\overline{IMC}(dB)$ for the noiseless (left) and noisy (right) cases for the MD sequence.
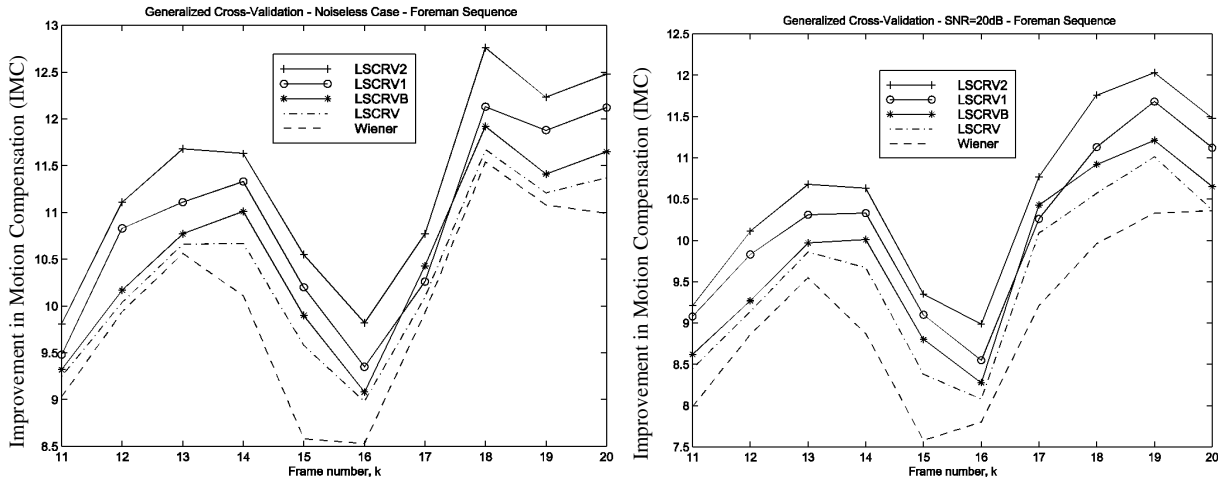

FIGURE 4: $\overline{IMC}(dB)$ for frames 11-20 of the noiseless (left) and noisy with $SNR = 20dB$ (right) for the Foreman sequence.

7. IMPLEMENTATION

In this section, we present several experimental results that illustrate the effectiveness of the GCV approach and compare it with the Wiener filter [2, 3 ,7] similar to the one in [1] given by

$$\hat{u}_{Wiener} = \hat{u}_{LMMSE} = \left(G^T G + \mu I\right)^{-1} G^T z \ , \tag{23}$$

where $\mu=50$ was chosen for all pixels in an entire frame. All sequences are *144×176*, *8*-bit (QCIF format). The algorithms were applied to three image sequences: one synthetically generated, with known motion; the "Mother and Daughter" (MD) and the "Foreman". For each sequence, two sets of experiments are analyzed: one for the noiseless case and the other for a sequence whose frames are corrupted by a signal-to-noise-ratio (SNR) equal to 20 dB. The SNR is defined as

$$SNR = 10\log_{10} \frac{\sigma^2}{\sigma_c^2} .$$ (24)

where σ^2 is the variance of the original image and σ_c^2 is the variance of the noise corrupted image [8].

	Wiener	LSCRV	LSCRVB	LSCRV1	LSCRV2
MSE$_x$	0.1548	0.1534	0.1511	0.1493	0.1440
MSE$_y$	0.0740	0.0751	0.0753	0.0754	0.0754
bias$_x$	0.0610	0.0619	0.0599	0.0581	0.0574
bias$_y$	-0.0294	-0.0291	-0.0294	-0.0294	-0.0293
$\overline{IMC}(dB)$	19.46	19.62	19.74	19.89	20.38
$\overline{DFD}^2$	4.16	4.05	3.921	3.76	3.35

TABLE 1: Comparison between GCV implementations and the Wiener filter. $SNR = \infty$.

	Wiener	LSCRV	LSCRVB	LSCRV1	LSCRV2
MSE$_x$	0.2563	0.2544	0.2446	0.2437	0.2373
MSE$_y$	0.1273	0.1270	0.1268	0.1257	0.1254
bias$_x$	0.0908	0.0889	0.0883	0.0881	0.0852
bias$_y$	-0.0560	-0.0565	-0.0564	-0.0561	-0.0553
$\overline{IMC}(dB)$	14.74	14.83	14.98	15.15	15.32
$\overline{DFD}^2$	12.24	12.02	11.60	11.16	10.78

TABLE 2: Comparison between GCV implementations and the Wiener filter. $SNR = 20dB$.

7.1 Programs and Experiments

The following GCV-based programs were developed:

a) **LSCRV:** λ is a scalar; a non-causal *3×3* mask, centered at the pixel being analyzed.

b) **LSCRVB:** λ is a scalar; we tried all nine masks.

c) **LSCRV1:** $\Lambda = \Lambda_{GCV} = diag(\lambda_1, \lambda_2)$, where λ_1 and λ_2 are scalars, is a matrix; a non-causal *3×3* mask centered at the pixel being analyzed.

d) **LSCRV2:** Λ is a matrix; we tried all nine masks.

Results from the proposed algorithms are compared to the ones obtained with the Wiener (LMMSE) filter from Eq. (23) in the subsequent experiments.

Experiment 1. In this sequence, there is a moving rectangle immersed in a moving background. In order to create textures for the rectangle and its background (otherwise motion detection would not be possible), the following auto-regressive model was used:

$$I(m,n) = \tfrac{1}{3}\big[I(m,n-1) + I(m-1,n) + I(m-1,n-1)\big] + n_i(m,n) ,$$ (25)

where $i = 1,2$. For the background $(i = 1)$, n_1 is a Gaussian random variable with mean $\mu_1 = 50$ and variance $\sigma_1^2 = 49$. The rectangle $(i = 2)$ was generated with $\mu_2 = 100$ and variance $\sigma_2^2 = 25$. All pixels from the background move to the right, and the displacement from frame *1* to frame *2* is

$d_b(r) = (d_{bx}(r), d_{by}(r)) = (2, 0)$. The rectangle moves in a diagonal fashion from frame *1* to *2* with $d_r(r) = (d_{rx}(r), d_{ry}(r)) = (1, 2)$.

Table 1 shows the values for the *MSE*, bias, $\overline{IMC}$ (dB) and $\overline{DFD}^2$ for the estimated optical flow using the Wiener filter and the four programs mentioned previously when no noise is present.
All the algorithms employing the *GCV* show improvement in terms of the metrics used. When we compare LSCRV with the Wiener filter, we see that with a regularization matrix of the form $\boldsymbol{\Lambda} = \lambda\boldsymbol{I}$, whose regularization parameter λ is determined by means of the minimization of the GCV function and using the same *3×3* mask as the Wiener, the improvements are small (we discuss some of our findings about the drawbacks of the GCV at the end of this article). The performance of the GCV implementation increases with the spatially adaptive approach a scalar regularization parameter λ (algorithm LSCRVB) when compared to the OLS. Now, when we compare the performance of the previous algorithms with the case where we have a more complex regularization matrix $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2\}$, that is, the implementation LSCRV1, then get even more improvements, although we have a single mask. Finally, using both the spatially adaptive approach and $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2\}$ we get the best results (the $\overline{IMC}(dB)$ goes up almost *1 dB* on the average).

Table 2 shows the values for the MSE, bias, $\overline{IMC}(dB)$ and $\overline{DFD}^2$ for the estimated optical flow using the Wiener filter and the four programs mentioned previously with two noisy frames ($SNR = 20dB$).

The results for both the noiseless and noisy cases present better values of $\overline{IMC}(dB)$ and $\overline{DFD}^2$ as well as MSE's and biases for all algorithms using the GCV. The best results in terms of metrics and visually speaking are obtained with the LSCRV2 algorithm ($\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2\}$ and multi-mask strategy). For the noisy case, it should be pointed out the considerable reduction of the interference of noise when it comes to the motion in the background and inside the object. For this algorithm, even the motion around the borders of the rectangle is clearer than when the LMMSE estimator is used.

Experiment 2. Fig. 4 presents the values of the improvement in motion compensation for frames *31* to *40* of the MD sequence for the noiseless and noisy ($SNR = 20dB$) cases, respectively, for all algorithms investigated. Here we concentrate our analysis on the performance of LSCRV2, which is the algorithm that gave us the best results. The LSCRV2 algorithm provides, on the average, $1.5dB$ higher $\overline{IMC}(dB)$ than the LMMSE algorithm for the noiseless case. The $\overline{IMC}(dB)$ for the noisy case is not as high as in the previous situation. Their qualitative performance can be observed in Fig. 3. By visual inspection, the noiseless case does present dramatic differences between both motion fields. For the noisy case, we were able of capturing the motion relative to the rotation of the mother's head, although incorrect displacement vectors were found in regions were there is no texture at all such as the background, for instance, but there is less noise than when we use the Wiener filter.

Experiment 3. Fig. 4 demonstrates results obtained for frames *11-20* of the "Foreman" sequence. Some frames of this sequence show abrupt motion changes. One can see that all the algorithms based on GCV outperform the LMMSE. This sequence shown very good values for the $\overline{IMC}(dB)$ for both the noiseless and the noisy cases. As one can see by looking at the plots for the errors in the motion compensated frames, the algorithm LSCRV2 performs better than the Wiener filter visually speaking.

8. CONCLUSION AND DISCUSSIONS

This work addresses some issues related to the application of two adaptive pel-recursive techniques to solve the problem of estimating the DVF. We analyzed the issue of robust estimation of the DVF between two consecutive frames, concentrating our attention on the effect of noise on the estimates. The observation z is subjected to independent identically distributed (i.i.d.) zero-mean additive Gaussian noise n. This entire work considered n and the update vector u as the only random signals present, as well as z, since it is obtained from a linear combination of u and n. Robustness to noise was achieved by means of regularization and by making the regularization parameters dependent on data.

In our case, the regularization matrix is no longer of the form λI, where λ is a scalar, but has a more general form Λ. The entries of Λ form a set of regularization parameters and such a formulation allows us more possibilities when it comes to find the best smoothed estimate.

A spatially adaptive approach was introduced and it consists of using a set of masks, each one representing a different neighborhood and yielding a distinct estimate. The final estimate is the one that provided the smallest $|DFD|$. The results from some experiments demonstrated the advantages of employing multiple masks.

A strategy for choosing the regularization parameter without knowledge of the noise statistics was introduced: the GCV. It depends solely on the observations. Two cases were explored: scalar λ and $\Lambda = diag\{\lambda_1, \lambda_2\}$. The best results were obtained using the GCV with $\Lambda = diag\{\lambda_1, \lambda_2\}$. All implementations of the GCV presented in this article performed better than the LMMSE technique. The drawbacks exhibited by the GCV in the context of motion estimation/detection were also analyzed. GCV evaluates the variability of the regression results when some subset of observations is omitted from the original set z. The criterion to select the best Λ is the minimum prediction MSE. Since the regularization matrix Λ is related to the autocorrelation of the data, there is some implicit stochastic knowledge in our model. The main advantage of our approach is the fact we use the GCV to choose the best set of regularization parameters and, then, we use these values to calculate the update vector estimate $\hat{u}_{GCV}$. This take on GCV brings in an implicitly Bayesian touch because the entries of Λ are actually variance ratios. We improved the GCV performance via the introduction of local adaptability (through the use of multiple neighborhoods).

It should be pointed out that our GCV model can handle the motion estimation/detection problem well and, as expected, a more complex Λ gives better result than a scalar regularization parameter.

The proposed method provides an automatic, data-based selection of the regularization parameter by means of the minimization of Eq. (14). The GCV is a non-parametric estimation method that does not require any knowledge about the probability density functions of the model variables, although the regularization parameters sought are related to the covariance matrices of u and n. It relies solely on the minimization of a function obtained from the weighted sum of squared prediction errors.

However, the technique presents some drawbacks [8]. This technique works very well in most of the cases (approximately 95% of the time), but due to the volume of minimizations done, the GCV failed to produce good estimates at all points because of one of the following situations:

a) $GVC(\Lambda)$ has multiple minima;
b) There is no minimum such that all entries of Λ are positive;
c) The minimum is hard to be found (no convergence);
d) The global minimum of the *GCV* results in a undersmoothed solution; a local minimum can be better; or
e) We may have found a saddle point.

Our spatially adaptive scheme indeed improves the behavior of the routines based on GCV around motion borders due to the fact that it seeks the neighborhood which provides the best system of equations according to the smoothness constraint assumption.

An interesting problem we are currently investigating, is a more intelligent way of choosing a neighborhood upon which to build our system of equations. We are also looking at more complex regularization matrices.

REFERENCES

[1]   J. Biemond, L. Looijenga, D. E. Boekee, R. H. J. M. Plompen, "A pel-recursive Wiener-based displacement estimation algorithm," Signal Proc., 13, 1987, pp. 399-412.

[2]   J. C. Brailean, A. K. Katsaggelos, "Simultaneous   recursive displacement estimation and restoration of noisy-blurred image sequences," IEEE Trans. Image Proc., Vol. 4, No. 9, 1995, pp. 1236-1268.

[3]   V. V. Estrela, N. P. Galatsanos, "Spatially-adaptive regularized pel-recursive motion estimation based on cross-validation," ICIP 98 Proceedings (2), 1998, pp. 200-203.

[4]   N. P. Galatsanos, A. K. Katsaggelos,   "Methods   for   choosing   the regularization parameter and estimating the noise variance in image restoration and their relation," IEEE Trans. Image Proc., Vol. 1, No. 3, 1992, pp. 322-336.

[5]   G. H. Golub, M. Heath, G.Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics, Vol. 21, No. 2, 1979, pp. 215-223.

[6]   A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, New Jersey, 1989.

[7]   A. M. Tekalp, Digital Video Processing, Prentice-Hall, New Jersey, 1995.

[8]   A. M. Thompson, J. C. Brown, J. W. Kay, D. M. Titterington, "A study of methods for choosing the smoothing parameter in image restoration by regularization," IEEE Trans. P.A.M.I., Vol. 13, No. 4, 1991, pp. 326-339.

[9]   G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia, 1990.

[10]   J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Performance of optical flow techniques," International Journal of Computer Vision, 12, 1994, pp. 43–77.

[11]   MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11 WG11N6828, 2004.

[12]   T. Ebrahimi, Y. Abdeljaoued, R. Figueras i Ventura, and O. Divorra Escoda, "MPEG-7 Camera," IEEE Proc. Int. Conference on Image Processing (ICIP), Thessaloniki, Greece, Oct. 2001.

[13]   R. Kapela and A. Rybarczyk, "Real-time shape description system based on MPEG-7 descriptors," Journal of Systems Architecture, Volume 53, Issue 9, 2007, pp. 602-618.

[14]   R. Kapela, and A. Rybarczyk, "The neighboring pixel representation for efficient binary image processing operations," International Symposium on Parallel Computing in Electrical Engineering (PARELEC'06), 2006, pp. 396-404.

[15]   A. Petrovic, O. Divorra Escoda and P. Vandergheynst, "Multiresolution segmentation of natural images: from linear to non-linear scale-space representations," IEEE Transactions on Image Processing, Vol. 13, No 8, 2004, pp. 1104-1114.

Alessandra Martins Coelho & Vania Vieira Estrela

[16]   A. Bab-Hadiashar, N. Gheissari, and D. Suter, "Robust Model Based Motion Segmentation," ICPR 2002, Quebec, Canada, 2002, pp. 753-757.

[17]   N. Gheissari, and A. Bab-Hadiashar, "Motion analysis: model selection and motion segmentation," Proceedings of 12th International Conference on Image Analysis and Processing, 2003, pp. 442-448.

INSTRUCTIONS TO CONTRIBUTORS

The *International Journal of Image Processing (IJIP)* aims to be an effective forum for interchange of high quality theoretical and applied research in the Image Processing domain from basic research to application development. It emphasizes on efficient and effective image technologies, and provides a central forum for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing.

We welcome scientists, researchers, engineers and vendors from different disciplines to exchange ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 6, 2012, IJIP appears in more focused issues. Besides normal publications, IJIP intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

LIST OF TOPICS
The realm of International Journal of Image Processing (IJIP) extends, but not limited, to the following:

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission

- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography
- Image capturing, databases

- Image processing applications

- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography

- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing

- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors
- Photopolymers
- Protocols for packet video
- Retrieval and multimedia

- Storage and transmission
- Video coding algorithms and technologies for ATM/p

CALL FOR PAPERS

Volume: 6 - **Issue:** 3 - June 2012

i. Paper Submission: March 31, 2012   **ii. Author Notification:** May 15, 2012

iii. Issue Publication: June 2012

CONTACT INFORMATION