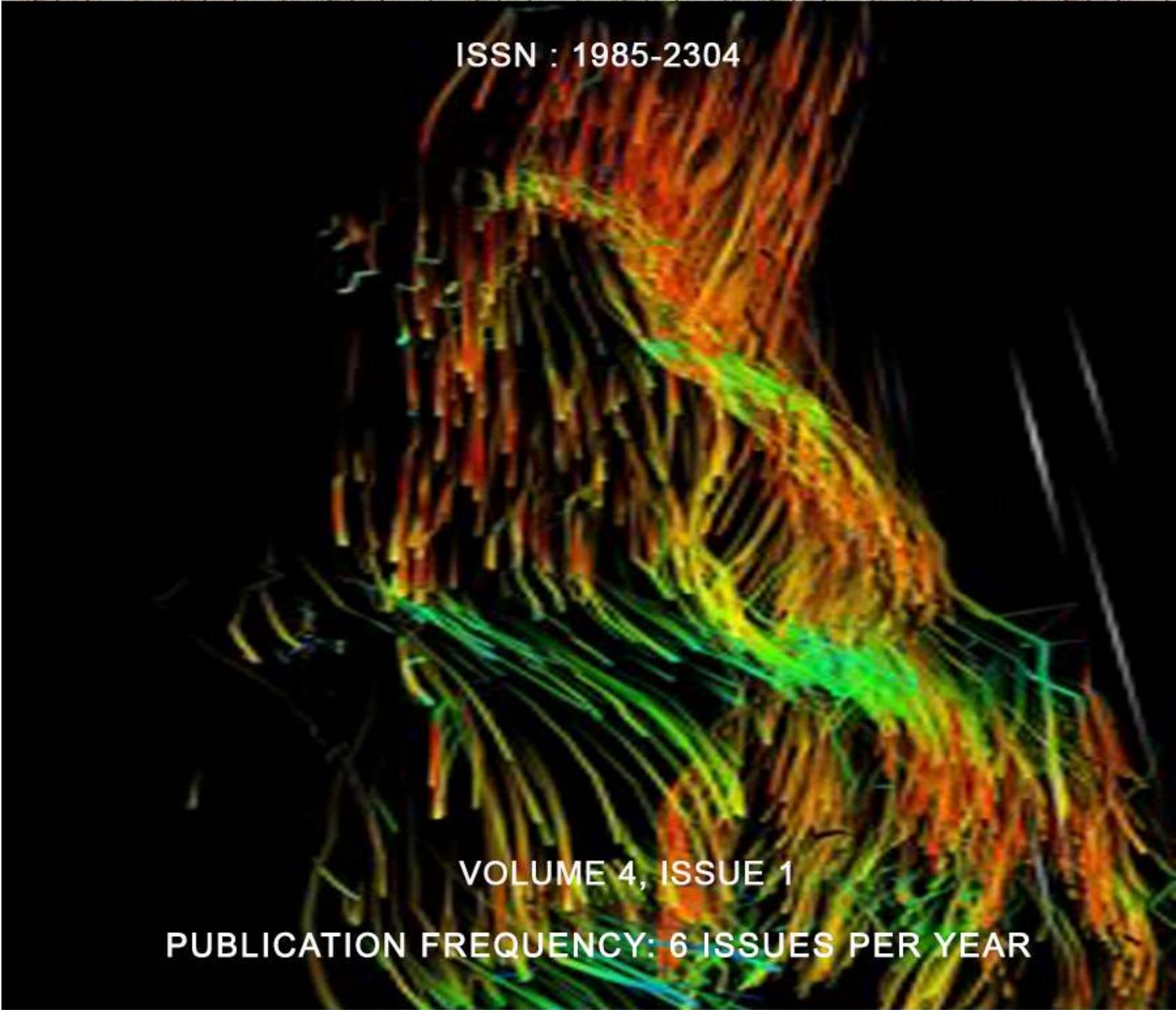


# International Journal of Image Processing (IJIP)



ISSN : 1985-2304



VOLUME 4, ISSUE 1

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

# **International Journal of Image Processing (IJIP)**

**Volume 4, Issue 1, 2010**

**Edited By**  
**Computer Science Journals**  
[www.cscjournals.org](http://www.cscjournals.org)

**Editor in Chief Professor Hu, Yu-Chen**

## **International Journal of Image Processing (IJIP)**

Book: 2010 Volume 4 Issue 1

Publishing Date: 31-03-2010

Proceedings

ISSN (Online): 1985-2304

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

IJIP Journal is a part of CSC Publishers

<http://www.cscjournals.org>

©IJIP Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

**CSC Publishers**

## 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 first 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.

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)

# Editorial Board

## Editor-in-Chief (EIC)

**Professor Hu, Yu-Chen**  
*Providence University (Taiwan)*

## Associate Editors (AEiCs)

**Professor. Khan M. I ftekaruddin**  
*University of Memphis ()*

**Dr. Jane(Jia) You**  
*The Hong Kong Polytechnic University (China)*

**Professor. Davide La Torre**  
*University of Milan (Italy)*

**Professor. Ryszard S. Choras**  
*University of Technology & Life Sciences ()*

**Dr. Huiyu Zhou**  
*Queen's University Belfast (United Kindom)*

## Editorial Board Members (EBMs)

**Professor. Herb Kunze**  
*University of Guelph (Canada)*

**Assistant Professor. Yufang Tracy Bao**  
*Fayetteville State University ()*

**Dr. C. Saravanan**  
*(India)*

**Dr. Ghassan Adnan Hamid Al-Kindi**  
*Sohar University (Oman)*

**Dr. Cho Siu Yeung David**  
*Nanyang Technological University (Singapore)*

**Dr. E. Sreenivasa Reddy**  
*(India)*

**Dr. Khalid Mohamed Hosny**  
*Zagazig University (Egypt)*

**Dr. Gerald Schaefer**  
*(United Kingdom)*

**Dr. Chin-Feng Lee**  
*Chaoyang University of Technology (Taiwan)*

**Associate Professor. Wang, Xiao-Nian**  
*Tong Ji University (China)*

**Professor. Yongping Zhang**  
*Ningbo University of Technology (China )*

# Table of Contents

Volume 4, Issue 1, March 2010.

## Pages

- |         |  |
|---------|--|
| 1– 10   | Two-level Vector Quantization Method for Codebook Generation using Kekere's Proportionate Error Algorithm<br><b>Tanuja, H. B. Kekre</b>  |
| 12 - 23 | Fusion of Wavelet Coefficients from Visual and Thermal Face Images for Human Face Recognition – A Comparative Study<br><b>Mrinal Kanti Bhowmik, Debotosh Bhattacharjee, Mita Nasipuri, Dipak Kumar Basu , Mahantapas Kundu</b> |
| 24 - 34 | Smart Video Surveillance System<br><b>Md.Hazrat Ali, Fadhlan Hafiz, A.A Shafie</b>   |
| 35 - 52 | Script Identification of Text Words from a Tri-Lingual Document Using Voting Technique<br><b>M C Padma, P. A. Vijaya</b>   |
| 53 - 65 | Wavelet Packet Based Features for Automatic Script Identification  |
| 66 - 76 | Image Denoising Using Earth Mover's Distance and Local Histograms<br><b>Nezamoddin N. Kachoui</b>  |

77 – 88

Digital Image-Adaptive Watermarking Based on BSS and Wavelet  
Domain Embedding

**Sangeeta d. Jadhav, anjali s. bhalchandra**

## Two-level Vector Quantization Method for Codebook Generation using Kekre's Proportionate Error Algorithm

**Dr. H. B. Kekre**

Senior Professor,  
MPSTME, SVKM's NMIMS University  
Mumbai, 400-056, India.

hbkekcre@yahoo.com

**Tanuja K. Sarode**

Ph.D. Scholar, MPSTME, SVKM's NMIMS University  
Assistant Professor, TSEC, Bandra (W),  
Mumbai, 400-050, India.

tanuja\_0123@yahoo.com

---

### Abstract

Vector Quantization is lossy data compression technique and has various applications. Key to Vector Quantization is good codebook. Once the codebook size is fixed then for any codebook generation algorithm the MSE reaches a value beyond which it cannot be reduced unless the codebook size is increased. In this paper we are proposing two-level codebook generation algorithm which reduces mean squared error (MSE) for the same codebook size. For demonstration we have used codebooks obtained from Kekre's Proportionate error (KPE) algorithm. The proposed method is can be applied to any codebook generation algorithm.

**Keywords:** Vector Quantization, Codebook, Data Compression, Encoding.

---

## 1. INTRODUCTION

The increasing demand in various fields has made the digital image compression very vigorous in the area of research. Compression techniques reduce the amount of data needed to represent an image so that images can be economically transmitted and stored. Vector quantization (VQ) is one of the non lossless data compression techniques. VQ has been used in number of applications, like speech recognition and face detection [3], [5], pattern recognition [4], speech data compression [30], image segmentation [31-34], Content Based Image Retrieval (CBIR) [35], [36], Face recognition[40].

VQ is a mapping function which maps k-dimensional vector space to a finite set  $CB = \{C_1, C_2, C_3, \dots, C_N\}$ . The set CB is called as codebook consisting of N number of codevectors and each codevector  $C_i = \{c_{i1}, c_{i2}, c_{i3}, \dots, c_{ik}\}$  is of dimension k. Good codebook design leads to less distortion in reconstructed image. Codebook can be designed in spatial domain by clustering algorithms [1], [2], [6], [7], [27-29], [37-39].

For encoding, image is fragmented into non overlapping blocks and each block is then converted to the training vector  $X_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ . The codebook is searched for the closest codevector  $C_{min}$  by computing squared Euclidean distance as presented in equation (1) between vector  $X_i$  and all the codevectors of the codebook CB. This method is called as exhaustive search (ES).

$$d(X_i, C_{min}) = \min_{1 \leq j \leq N} \{d(X_i, C_j)\} \text{ Where } d(X_i, C_j) = \sum_{p=1}^k (x_{ip} - c_{jp})^2 \quad (1)$$

Exhaustive Search (ES) method gives the optimal result at the end, but it intensely involves computational complexity. Observing equation (1) to obtain one nearest codevector for a training vector computations required are  $N$  Euclidean distance where  $N$  is the size of the codebook. So for  $M$  image training vectors, will require  $M*N$  number of Euclidean distances computations. It is obvious that if the codebook size is increased the distortion will decrease with increase in searching time.

Various encoding methods are given in literature: Partial Distortion search (PDS)[7], nearest neighbor search algorithm based on orthonormal transform (OTNNS) [8]. Partial Distortion Elimination (PDE) [9], triangular inequality elimination (TIE) [10], mean distance ordered partial codebook search (MPS) algorithm [11], fast codebook search algorithm based on the Cauchy-Schwarz inequality (CSI) [12], fast codebook search based on subvector technique (SVT) [13], the image encoding based on L2-norm pyramid of codewords [14] and the fast algorithms using the modified L2-norm pyramid (MLP) [15], fast codeword search algorithm based on MPS+TIE+PDE proposed by Yu-Chen, Bing-Hwang and Chih-Chiang (YBC) in 2008 [16], Kekre's fast search algorithms [17], [18], [19] and others [20], [21], [22], are classified as partial search methods. Some of the partial techniques use data structure to organize the codebook for example tree-based [23], [24] and projection based structure [25], [26]. All these algorithms lessen the computational cost needed for VQ encoding keeping the image quality close to Exhaustive search algorithm.

To generate codebook there are various algorithms. It is observed that for the same codebook size the distortion obtained from codebook generation algorithms varies. However the minimum error is not achieved. Once the codebook size is fixed then for all these algorithms the MSE reaches a value beyond which it cannot be reduced because the codevectors in the codebook have not reached their optimal position. Hence Two-level codebook generation algorithm which minimizes the distortion further is proposed. For demonstration codebooks obtained from Kekre's Proportionate Error (KPE) [29] algorithm is used and the results are compared with well known LBG codebook.. The method proposed is quite general and is applicable to any codebook generation algorithm.

## 2. KEKER'S PROPORTIONATE ERROR ALGORITHM (KPE)

Let  $T = \{X_1, X_2, \dots, X_M\}$  be the training sequence consisting of  $M$  source vector. Assume that source vector is of length  $K$ ,  $X_m = \{x_{m,1}, x_{m,2}, \dots, x_{m,K}\}$  for  $m=1, 2, \dots, M$ . In this algorithm initial codevector is computed by taking the mean of all the training vectors  $X_i$  for  $i=1, 2, \dots, M$ . Thus initially the codebook contains only one codevector. Then two vectors from the codevector are computed by adding proportionate error instead of adding constant. From the codevector proportions between the members of vector is calculated.

Let  $k$  be the length of codevector,

$C = \{c_1, c_2, \dots, c_k\}$  be the codevector, and

$E = \{e_1, e_2, \dots, e_k\}$  be the error vector

$c_j = \min\{c_i / i = 1, 2, \dots, k\}$  where  $j$  is the index of the member of vector whose value is minimum among the vector members.

Then assign  $e_j = 1$  and if  $c_i / c_j \leq 10$  then assign  $e_i = c_i / c_j$

else assign  $e_i = 10$  for  $i \neq j$  and  $i=1, 2, \dots, k$ .

Two vectors  $v_1$  and  $v_2$  are formed by adding the error vector  $E$  to codevector  $C$  and by subtracting the error vector  $E$  from codevector  $C$  respectively. Euclidean distance between the all the training vectors  $X_i$  with  $v_1$  and with  $v_2$  are computed

i.e.  $d_1 = \|v_1 - X_i\|_2$  and  $d_2 = \|v_2 - X_i\|_2$  for  $i=1, 2, \dots, M$

if  $d_1 < d_2$  then  $X_i$  is put in cluster1 else  $X_i$  is put in cluster2 and two clusters are created.

From each cluster codevector is computed by taking the mean of all the vectors in the cluster.

Thus the codebook size is increased to two.

The above procedure is repeated for each of the codevector and that codebook size is increased to four. This procedure is repeated till the codebook size is increased to the size specified by the user or MSE is reduced to minimum permissible value.

### 3. TWO-LEVEL CODEBOOK GENERATION ALGORITHM

First the image is spitted in non-overlapping blocks of 2x2 pixels (each pixel consisting of R, G, and B color component). Hence we get vector of dimension 12. Codebooks of sizes 256, 512 and 1024 are obtained using following Two-level codebook generation algorithm.

Let  $F$  be the input image and  $N$  be the codebook size.

1. The  $x\%$  (where  $x = 100, 75, 70, 69, 68, 67, 50$ ) of size of codebook is generated using KPE algorithm.
2. Image  $\hat{F}$  is reconstructed using codebook obtained in step 1.
3. Generate error image  $E = F - \hat{F}$ .
4. Construct codebook of size  $N-(N*x)/100$  for an error image  $E$  using KPE.
5. Reconstruct error image  $\hat{E}$  using codebook obtained from step 4.
6. Regenerate final image by adding  $\hat{F}$  and  $\hat{E}$ .

The method is general and can be applied to any codebook generation algorithm. For illustration of this method KPE codebook generation algorithm is used and results are compared with well known LBG algorithm[2], [39].

### 4. RESULTS

The Two-level VQ algorithms are implemented on Intel processor 1.66 GHz, 1GB RAM machine to obtain results. We have tested these algorithms on seven images namely Flower, Ganesh, Scenary, Strawberry, Tajmahal and Tiger each of size 256X256 color images. The images selected correspond to different classes as shown in Figure 1.

Table I, II, III shows the performance comparison of MSE for different sharing between level 1 and level 2 using KPE codebooks of varying sizes 256, 512 and 1024 respectively.

Figure 1. shows the six color training images namely Ganesh, Strawberry, Scenary, Tajmahal and Tiger of size 256X256 on which this algorithm is tested.

Figure 2, 3 and 4. shows the plot of Average MSE for different sharing between level 1 and level 2 using LBG and KPE codebooks of sizes 256, 512 and 1024 respectively.

Figure 5. shows the results of Tajmahal image using Two-level technique.



Figure 1. Six color Training images of size 256x256

Images		CB Size 256							
		% of level 1	100	75	70	69	68	67	50
		CB Size for level 1	256	192	179	176	174	171	128
Flower	MSE after level 1	314.99	217.90	218.89	219.05	219.06	219.20	219.39	
	MSE after level 2	314.99	112.42	106.65	106.80	103.95	103.89	<b>89.93</b>	
	PSNR	23.15	27.62	27.85	27.84	27.96	27.97	28.59	
Ganesh	MSE after level 1	767.77	500.84	514.81	515.17	518.11	519.31	539.25	
	MSE after level 2	767.77	392.40	388.61	379.74	373.24	360.28	<b>307.40</b>	
	PSNR	19.28	22.19	22.24	22.34	22.41	22.56	23.25	
Scenery	MSE after level 1	406.39	204.48	207.30	208.02	208.26	209.05	211.99	
	MSE after level 2	406.39	132.90	135.62	134.33	134.57	135.10	<b>116.36</b>	
	PSNR	22.04	26.90	26.81	26.85	26.84	26.82	27.47	
Strawberry	MSE after level 1	338.06	239.81	242.08	242.40	242.61	243.42	247.22	
	MSE after level 2	338.06	192.31	190.14	190.81	187.86	188.14	<b>168.14</b>	
	PSNR	22.84	25.29	25.34	25.32	25.39	25.39	25.87	
Tajmahal	MSE after level 1	364.59	266.81	270.00	270.59	270.60	271.15	280.39	
	MSE after level 2	364.59	191.79	195.44	188.20	190.22	192.86	<b>158.67</b>	
	PSNR	22.51	25.30	25.22	25.38	25.34	25.28	26.13	
Tiger	MSE after level 1	463.80	374.79	387.73	388.62	389.62	390.93	401.15	
	MSE after level 2	463.80	293.76	276.89	267.07	270.88	269.95	<b>245.79</b>	
	PSNR	21.47	23.45	23.71	23.86	23.80	23.82	24.23	
Average	MSE after level 1	442.60	300.77	306.80	307.31	308.04	308.84	316.57	
	MSE after level 2	442.60	219.26	215.56	211.16	210.12	208.37	<b>181.05</b>	

	<b>PSNR</b>	21.88	25.13	25.20	25.27	25.29	25.31	25.92
<b>Percentage MSE Reduction</b>		50.46	51.30	52.29	52.53	52.92	59.09	

**TABLE 1.** Comparison of MSE for different sharing between level 1 and level 2 using KPE codebook of size 256

Images		CB Size 512						
	% of level 1	100	75	70	69	68	67	50
	CB Size for level 1	512	384	358	353	348	343	256
	CB size for level 2	0	128	154	159	164	169	256
Flower	MSE after level 1	249.18	149.72	151.50	152.16	152.31	152.44	152.81
	MSE after level 2	249.18	85.23	84.94	79.83	78.83	77.78	<b>56.32</b>
	PSNR	24.17	28.82	28.84	29.11	29.16	29.22	30.62
Ganesh	MSE after level 1	654.09	402.08	418.00	420.02	421.03	423.28	440.35
	MSE after level 2	654.09	272.96	256.55	260.02	256.35	254.12	<b>204.44</b>
	PSNR	19.97	23.77	24.04	23.98	24.04	24.08	25.03
Scenary	MSE after level 1	296.83	156.28	160.23	160.49	161.32	161.99	165.37
	MSE after level 2	296.83	102.56	100.29	99.55	101.65	96.54	<b>73.12</b>
	PSNR	23.41	28.02	28.12	28.15	28.06	28.28	29.49
Strawberry	MSE after level 1	277.28	197.55	199.55	200.01	200.53	200.76	203.34
	MSE after level 2	277.28	125.42	123.32	121.46	120.53	118.86	<b>87.12</b>
	PSNR	23.70	27.15	27.22	27.29	27.32	27.38	28.73
Tajmahal	MSE after level 1	279.01	222.44	226.39	227.13	227.74	229.23	233.75
	MSE after level 2	279.01	116.71	110.26	113.15	110.72	118.67	<b>77.71</b>
	PSNR	23.67	27.46	27.71	27.59	27.69	27.39	29.23
Tiger	MSE after level 1	432.18	285.36	295.17	297.56	298.57	300.36	313.92
	MSE after level 2	432.18	168.46	175.61	173.58	170.13	174.37	<b>131.46</b>
	PSNR	21.77	25.87	25.69	25.74	25.82	25.72	26.94
Average	MSE after level 1	364.76	235.57	241.81	242.90	243.58	244.68	251.59
	MSE after level 2	364.76	145.22	141.83	141.27	139.70	140.06	<b>105.03</b>
	PSNR	22.78	26.85	26.94	26.98	27.02	27.01	28.34
<b>Percentage MSE Reduction</b>		60.19	61.12	61.27	61.70	61.60	71.21	

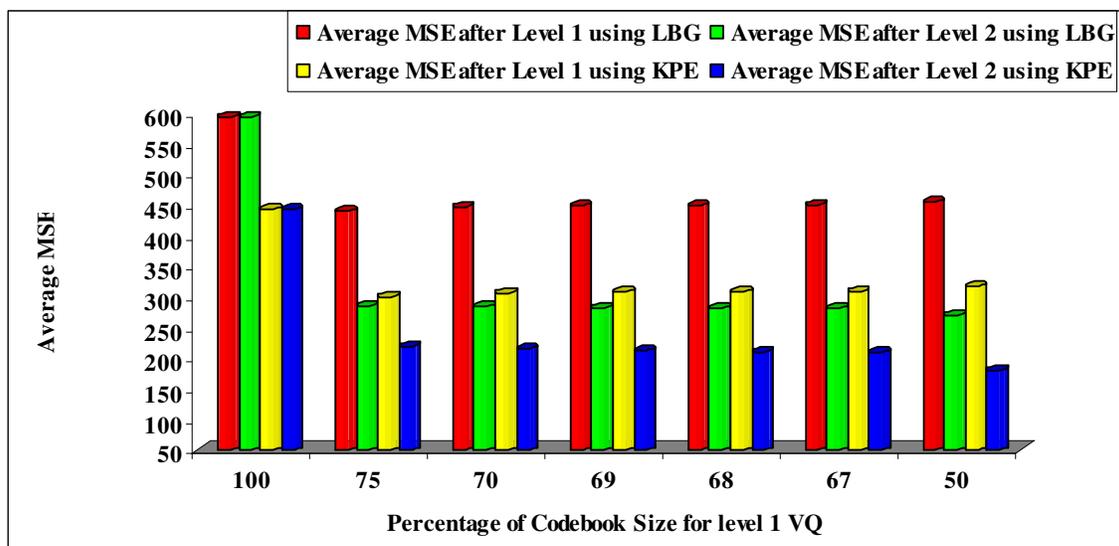
**TABLE 2.** Comparison of MSE for different sharing between level 1 and level 2 using KPE codebook of size 256

Images		CB Size 1024						
	% of level 1	100	75	70	69	68	67	50
	CB Size for level 1	1024	768	716	706	696	686	512
	CB size for level 2	0	256	308	318	328	338	512
Flower	MSE after level 1	167.20	98.92	101.35	101.63	101.74	102.01	102.80
	MSE after level 2	167.20	47.03	47.94	47.95	47.29	46.30	<b>33.35</b>
	PSNR	25.90	31.41	31.32	31.32	31.38	31.47	32.90
Ganesh	MSE after level 1	544.79	296.69	307.67	309.55	310.71	312.00	322.13
	MSE after level 2	544.79	160.06	164.04	163.50	161.24	160.06	<b>125.42</b>
	PSNR	20.77	26.09	25.98	26.00	26.06	26.09	27.15
Scenary	MSE after level 1	189.24	106.26	107.78	108.30	108.52	108.84	111.21
	MSE after level 2	189.24	45.91	48.60	49.58	49.28	48.48	<b>34.36</b>
	PSNR	25.36	31.51	31.26	31.18	31.20	31.27	32.77
Strawberry	MSE after level 1	233.57	157.25	158.90	159.09	159.56	160.10	162.81
	MSE after level 2	233.57	67.74	67.87	67.62	68.29	67.42	<b>49.90</b>
	PSNR	24.45	29.82	29.81	29.83	29.79	29.84	31.15
Tajmahal	MSE after level 1	230.30	169.30	173.25	173.64	174.41	174.02	176.35

	<b>MSE after level 2</b>	230.30	65.48	64.15	64.96	66.43	65.17	<b>46.34</b>
	<b>PSNR</b>	24.51	29.97	30.06	30.00	29.91	29.99	31.47
<b>Tiger</b>	<b>MSE after level 1</b>	345.72	193.81	198.90	199.54	201.05	201.86	212.35
	<b>MSE after level 2</b>	345.72	92.48	91.33	91.04	90.95	90.06	<b>68.50</b>
	<b>PSNR</b>	22.74	28.47	28.52	28.54	28.54	28.59	29.77
<b>Average</b>	<b>MSE after level 1</b>	285.14	170.37	174.64	175.29	176.00	176.47	181.28
	<b>MSE after level 2</b>	285.14	79.78	80.66	80.78	80.58	79.58	<b>59.65</b>
	<b>PSNR</b>	23.96	29.55	29.49	29.48	29.48	29.54	30.87
<b>Percentage MSE Reduction</b>			72.02	71.71	71.67	71.74	72.09	79.08

**TABLE 3.** Comparison of MSE for different sharing between level 1 and level 2 using KPE codebook of size 256

**Note:** The highlighted figures in all these tables are for minimum value of MSE. It is observed that most of them are close to 50% allocation of codebook to level 1. This indicates that codebook be divided in the ratio 1:1 between level 1 and level 2. Further it is observed that MSE reduces by 59%, 71% and 79% for the codebook size 256, 512 and 1024 respectively.



**Figure 2.** Plot of Level 1 and level 2 average MSE Vs percentage of codebook size for level 1 VQ using LBG and KPE for codebook size 256.

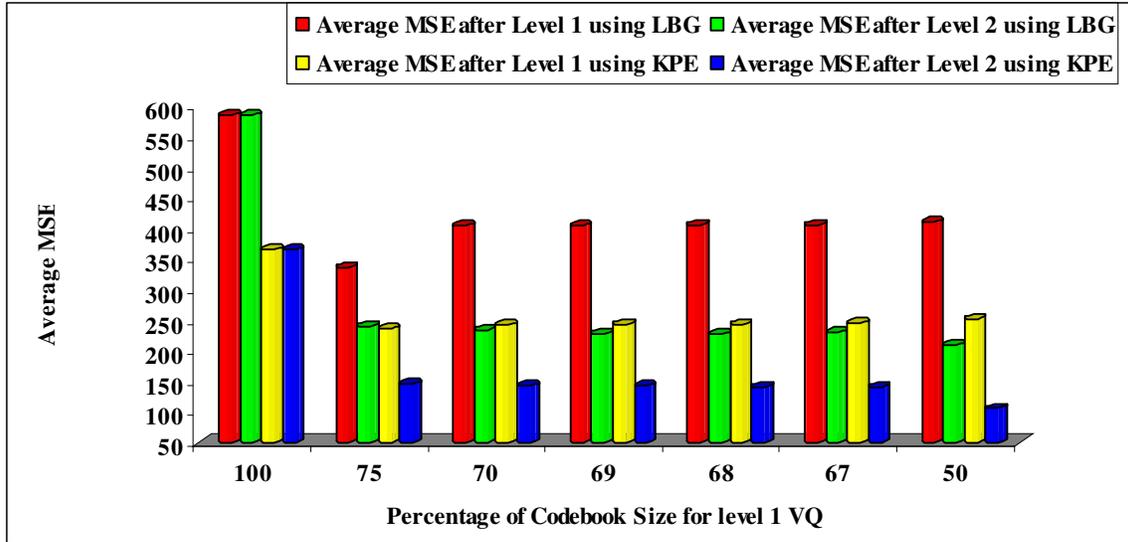


Figure 3. Plot of Level 1 and level 2 average MSE Vs percentage of codebook size for level 1 VQ using LBG and KPE for codebook size 512.

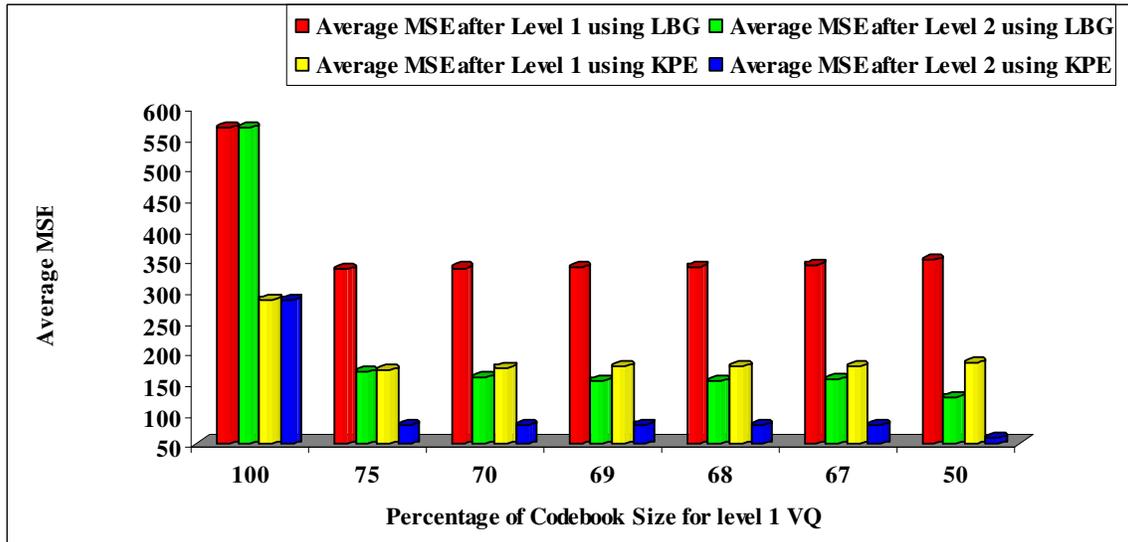
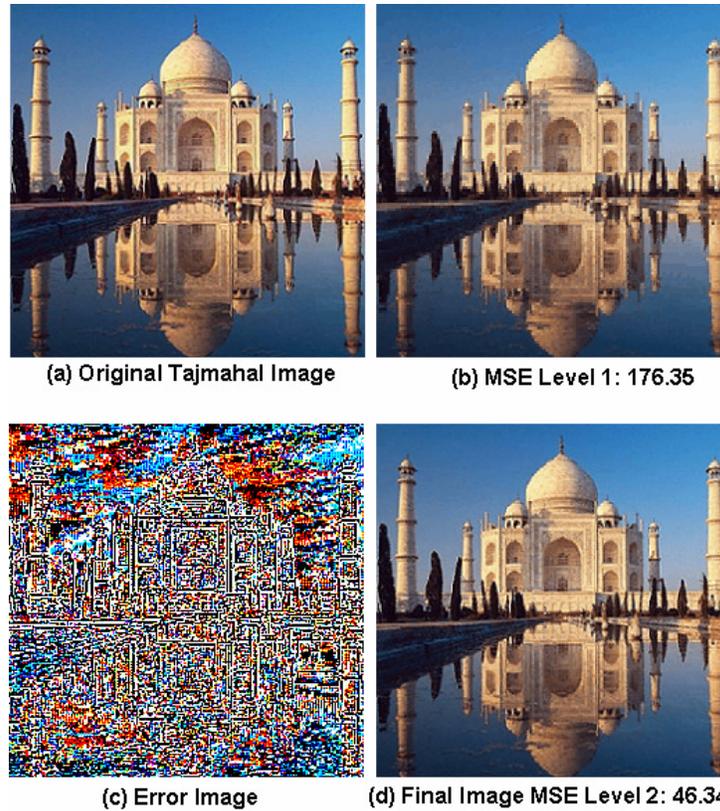


Figure 4. Plot of Level 1 and level 2 average MSE Vs percentage of codebook size for level 1 VQ using LBG and KPE for codebook size 1024.



**Figure 5.** Tajmahal image reconstructed after two-level VQ and error images after level 1 and level 2, image size 256x256 codebook size 512 for level 1 and 512 for level 2.

## 5. CONCLUSIONS

Paper presents a novel error minimization technique using two-level VQ. It is observed that once we fix the codebook size and then we use any VQ algorithm the MSE obtained for that size cannot be reduced further unless the codebook size is increased. Here in this paper we present a novel idea of splitting the codebook in two levels in the ratio L1:L2. Any codebook algorithm can be used for level 1 and the error image is obtained. On this error image same VQ algorithm is used in level 2. It is observed that this method drastically reduces MSE as compared to allocation of entire codebook to level 1. Minimum MSE obtained by this algorithm is based on ratio L1:L2 and is image dependent. However size allocation in the ratio of 1:1 for level 1 and level 2 on the average gives the best results reducing MSE by 79 percent for the codebook size 1024. Further it is observed that MSE reduction obtained using KPE codebook with respect to LBG is 33% to 53% for the codebook size 256 to 1024 respectively.

## 6. REFERENCES

1. R. M. Gray, "Vector quantization", IEEE ASSP Mag., pp. 4-29, Apr. 1984.
2. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans.Commun., vol. COM-28, no. 1, pp. 84-95, 1980.
3. Chin-Chen Chang, Wen-Chuan Wu, "Fast Planar-Oriented Ripple Search Algorithm for Hyperspace VQ Codebook", IEEE Transaction on image processing, vol 16, no. 6, June 2007.

4. Ahmed A. Abdelwahab, Nora S. Muharram, "A Fast Codebook Design Algorithm Based on a Fuzzy Clustering Methodology", *International Journal of Image and Graphics*, vol. 7, no. 2 pp. 291-302, 2007.
5. C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264–277, Sep. 1999.
6. A.Gersho, R.M. Gray.: 'Vector Quantization and Signal Compression', Kluwer Academic Publishers, Boston, MA, 1991.
7. C. D. Bei and R. M. Gray.: 'An improvement of the minimum distortion encoding algorithm for vector quantization', *IEEE Trans. Commun.*,vol. 33, No. 10, pp. 1132–1133, Oct. 1985.
8. Z. Li, and Z.- M. Lu. : 'Fast codevector search scheme for 3D mesh model vector quantization', *Electron. Lett.*, vol. 44, No. 2, pp. 104-105, Jan 2008.
9. C. Bei, R. M. Gray, "An improvement of the minimum distortion encoding algorithm for vector quantization", *IEEE Trans. Commun.*, Vol. 33, no. 10, pp. 1132–1133, Oct 1985.
10. S. H. Huang, S. H. Chen, "Fast encoding algorithm for VQ-based image coding", *Electron. Lett.* Vol. 26, No. 19, pp. 1618–1619, 1990.
11. S. W. Ra, J. K. Kim, "A fast mean-distance-ordered partial codebook search algorithm for image vector quantization", *IEEE Trans. Circuits-II*, vol. 40, No. 9, pp. 576–579, 1993.
12. K .S. Wu, J.C. Lin, "Fast VQ encoding by an efficient kick-out condition", *IEEE Trans. Circ. Syst. Vid.*, vol.10, No. 1, pp. 59–62, 2000.
13. J .S. Pan, Z.M. Lu, S.H. Sun, "An efficient encoding algorithm for vector quantization based on subvector technique", *IEEE Trans. Image Process.* Vol 12, No.3, pp. 265–270, 2003.
14. B. C. Song, J.B. Ra, "A fast algorithm for vector quantization using L2-norm pyramid of codewords", *IEEE Trans. Image Process.* Vol. 4, No.12, pp. 325–327, 2002.
15. Z. Pan, K. Kotani, T. Ohmi, "Fast encoding method for vector quantization using modified L2-norm pyramid", *IEEE Signal Process. Lett.* Vol. 12, issue 9, pp. 609–612, 2005.
16. Y. Chen, B. Hwang, C. Chiang, "Fast VQ codebook search algorithm for grayscale image coding", *Image and Vision Compu.*, vol. 26, No. 5, pp. 657-666, May 2008.
17. H. B. Kekre, Tanuja K. Sarode, "Centroid Based Fast Search Algorithm for Vector Quantization", *International Journal of Imaging (IJI)*, Vol 1, No. A08, pp. 73-83, Autumn 2008, available: <http://www.ceser.res.in/iji.html>
18. H. B. Kekre, Tanuja K. Sarode, "Fast Codevector Search Algorithm for 3-D Vector Quantized Codebook", *WASET International Journal of cal Computer Information Science and Engineering (IJCISE)*, Volume 2, No. 4, pp. 235-239, Fall 2008. available: <http://www.waset.org/ijcise>.
19. H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Search Algorithm for Vector Quantization using Sorting Technique", *ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009)*, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available on online ACM portal.
20. C. C. Chang and I. C. Lin, "Fast search algorithm for vector quantization without extra look-up table using declustered subcodebooks," *IEE Proc. Vis., Image, Signal Process.*, vol. 152, No. 5, pp. 513–519, Oct.2005.
21. C. M. Huang, Q. Bi, G. S. Stiles, and R. W. Harris, "Fast full-search equivalent encoding algorithms for image compression using vector quantization," *IEEE Trans. Image Process.*, vol. 1, No. 3, pp. 413–416, Jul. 1992.
22. Y. C. Hu and C. C. Chang, "An effective codebook search algorithm for vector quantization", *Imag. Sci. J.*, vol. 51, No. 4, pp. 221–234, Dec. 2003.
23. R. M. Gray and Y. Linde, "Vector quantization and predictive quantizers for gauss-markov sources," *IEEE Trans. Commun.*, vol. 30, No. 2, pp. 381–389, Feb. 1982.
24. S. J. Wang and C. H. Yang, "Hierarchy-oriented searching algorithms using alternative duplicate codewords for vector quantization mechanism," *Appl. Math. Comput.*, vol. 162, No. 234, pp. 559–576, Mar. 2005.
25. S. C. Tai, C. C. Lai, and Y. C. Lin, "Two fast nearest neighbor searching algorithms for image vector quantization," *IEEE Trans. Commun.*, vol. 44, No. 12, pp. 1623–1628, Dec. 1996.
26. C. C. Chang, D. C. Lin, and T. S. Chen, "An improved VQ codebook search algorithm using principal component analysis," *J. Vis. Commun. Image Represent.*, vol. 8, No. 1, pp. 27–37, Mar. 1997.

27. H. B. Kekre, Tanuja K. Sarode, "New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Engineering and Technology, vol.1, No.1, pp. 67-77, September 2008.
28. H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Computer Science and Information Technology, Vol. 1, No. 1, pp: 7-12, Jan 2009.
29. H. B. Kekre, Tanuja K. Sarode, "An Efficient Fast Algorithm to Generate Codebook for Vector Quantization," First International Conference on Emerging Trends in Engineering and Technology, ICETET-2008, held at Rasoni College of Engineering, Nagpur, India, 16-18 July 2008, Available at online IEEE Xplore.
30. H. B. Kekre, Tanuja K. Sarode, "Speech Data Compression using Vector Quantization", WASET International Journal of Computer and Information Science and Engineering (IJCISE), vol. 2, No. 4, 251-254, Fall 2008. available: <http://www.waset.org/ijcise>
31. H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Algorithm for Vector Quantization", International Journal of Computer Science (IJCS), Vol. 3, No. 4, pp. 287-292, Fall 2008. available: <http://www.waset.org/ijcs>.
32. H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Vector Quantization", Journal of Advances in Engineering Science, Section C, Number 3, 2008.
33. H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Vector Quantization Techniques Based on Energy Ordering Concept" International Journal of Computing Science and Communication Technologies (IJCST) Volume 1, Issue 2, January 2009.
34. H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Fast Codebook Generation Algorithm Based on Energy Ordering Concept", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), 23-24 Jan 2009, Fr. Conceicao Rodrigues College of Engg., Mumbai. Available on online ACM portal.
35. H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), Volume 9, Issue 5, pp.: 1-8, September 2009. Available online at <http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html>.
36. H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html).
37. H. B. Kekre, Tanuja K. Sarode, "Vector Quantized Codebook Optimization using K-Means", International Journal on Computer Science and Engineering (IJCSE) Vol.1, No. 3, 2009, pp.: 283-290, Available online at: [http://journals.indexcopernicus.com/abstracted.php?level=4&id\\_issue=839392](http://journals.indexcopernicus.com/abstracted.php?level=4&id_issue=839392).
38. H. B. Kekre, Tanuja K. Sarode, "Multilevel Vector Quantization Method for Codebook Generation", International Journal of Engineering Research and Industrial Applications (IJERIA), vol. 2, no. V, 2009, ISSN 0974-1518, pp.: 217-231. Available online at: [http://www.ascent-journals.com/ijeria\\_contents\\_Vol2No5.htm](http://www.ascent-journals.com/ijeria_contents_Vol2No5.htm).
39. H. B. Kekre, Tanuja K. Sarode, "Bi-level Vector Quantization Method for Codebook Generation", Second International Conference on Emerging Trends in Engineering and Technology, at G. H. Rasoni College of Engineering, Nagpur on 16-18 December 2009 pp.:866-872, this paper is available on IEEE Xplore.
40. H. B. Kekre, Kamal Shah, Tanuja K. Sarode, Sudeep D. Thepade, "Performance Comparison of Vector Quantization Technique – KFCG with LBG, Existing Transforms and PCA for Face Recognition", International Journal of Information Retrieval (IJIR), Vol. 02, Issue 1, pp.: 64-71, 2009.

## Fusion of Wavelet Coefficients from Visual and Thermal Face Images for Human Face Recognition – A Comparative Study

### Mrinal Kanti Bhowmik

Lecturer, Department of Computer Science and Engineering  
Tripura University (A Central University)  
Suryamaninagar – 799130, Tripura, India

mkb\_cse@yahoo.co.in

### Debotosh Bhattacharjee

Reader, Department of Computer Science and Engineering  
Jadavpur University  
Kolkata – 700032, India

debotosh@indiatimes.com

### Mita Nasipuri

Professor, Department of Computer Science and Engineering  
Jadavpur University  
Kolkata – 700032, India

mitanasipuri@gmail.com

### Dipak Kumar Basu

Professor, AICTE Emeritus Fellow, Department of Computer Science and Engineering  
Jadavpur University  
Kolkata – 700032, India

dipakkbasu@gmail.com

### Mahantapas Kundu

Professor, Department of Computer Science and Engineering  
Jadavpur University  
Kolkata – 700032, India

mkundu@cse.jdvvu.ac.in

---

### Abstract

In this paper we present a comparative study on fusion of visual and thermal images using different wavelet transformations. Here, coefficients of discrete wavelet transforms from both visual and thermal images are computed separately and combined. Next, inverse discrete wavelet transformation is taken in order to obtain fused face image. Both Haar and Daubechies (db2) wavelet transforms have been used to compare recognition results. For experiments IRIS Thermal/Visual Face Database was used. Experimental results using Haar and Daubechies wavelets show that the performance of the approach presented here achieves maximum success rate of 100% in many cases.

**Keywords:** Thermal image, Daubechies Wavelet Transform, Haar Wavelet Transform, Fusion, Principal Component Analysis (PCA), Multi-layer Perceptron, Face Recognition.

---

## 1. INTRODUCTION

Research activities on human face recognition are growing rapidly and researchers have expanded the application potential of the technology in various domains [1, 2, 3, 4, 5]. It is still a very challenging task to perform efficient face recognition in the outside environments with no

control on illumination [6]. Accuracy of the face recognition system degrades quickly when the lighting is dim or when it does not uniformly illuminate the face [7]. Light reflected from human faces also varies depending on the skin color of people from different ethnic groups. Fusion techniques have many advantages over the thermal and visual images to improve the overall recognition accuracy. As a matter of fact fusion of images has already established its importance in case of image analysis, recognition, and classification. Different fusion levels are being used, for example, low-level fusion, high-level fusion etc. Low-level data fusion combines the data from two different images of same pose and produces a new data that contains more details. High-level decision fusion combines the decisions from multiple classification modules [8], [9]. Decision fusion can be accomplished with majority voting, ranked-list combination [10], and the use of Dempster-Shafer theory [11]. Several fusion methods have been attempted in face recognition. Few of them are shown at Table I.

**TABLE 1:** Different Fusion Methods.

Author	Technique	References
I. Pavlidis	Dual band Fusion System	[6]
I. Pavlidis	Near Infrared Fusion Scheme	[12]
J. Heo	Data and Decision Fusion	[11]
A. Gyaourova	Pixel based fusion using Wavelet	[13]
A. Selinger	Appearance based Face Recognition	[14]
A. Selinger	Fusion of co-registered visible and LWIR	[14]
M. Hanif	Optimized Visual and Thermal Image Fusion	[15]

One of the fusion technique mentioned above, A. Gyaourova et al [13] tried to implement pixel-based fusion at multiple resolution levels. It allows features with different spatial extend to be fused at the resolution at which they are most salient. In this way, important features appearing at lower resolutions can be preserved in the fusion process. This method contains two major steps: one is fusion of IR and visible images and other one is recognition based on the fused images. Fusion is performed by combining the coefficient of Haar wavelet by decomposition of a pair of IR and visible images having equal size. They used Genetic Algorithms (GAs) to fuse the wavelet coefficients from the two spectra. They use GAs for fusion was based on several factors. First, the search space for the image fusion. Second, the problem at hand appears to have many suboptimal solutions. They shown the effectiveness of the fusion solution by GAs and they perform the recognition using eigenfaces. They used Euclidian distance to recognize the PCA applied by projected data. A. Selinger et al [14] they have compared different algorithms like eigenfaces (PCA), Linear Discriminate Analysis (LDA) and many more for multiple appearance-based face recognition methodology. All the experimental data were captured with a visible CCD array and LWIR microbolometer and used n-fold cross validation technique for the performance. I. Pavlidis et al [6], they propose a novel imaging solution which address the problems like lightning and disguise effect in facial image. They demonstrate a dual-band fusion system in the near infrared which segment human faces much more accurately than traditional visible band face detection systems and they have also represent with theoretical and experimental arguments that the upper band of the near infrared is particularly advantageous for disguise detection purpose. In their face detection system is the only dual band system in the near infrared range. For this method they calls two cameras one for upper-band (within a range of 0.8 - 1.4  $\mu\text{m}$ ) and lower-band (within a range of 1.4 - 2.2 $\mu\text{m}$ ). They generated fused image by using weighted difference method of the co-registered imaging signals from the lower and upper-band cameras because of the abrupt change in the reflectance for the human skin around 1.4  $\mu\text{m}$  the fusion has as a result the intensification of the humans face silhouettes and other exposed skin parts and the diminution of the background. They used Otsu thresholding algorithm for perfect image segmentation. The dual-band near-infrared face detection method could be used in combination with a traditional visible spectrum face recognition method. This approach will provide a much more accurate face segmentation result than a traditional visible band method. G. Bebis et al [16] in this paper they have extended their previous technique which is mentioned at [13]. In this paper, they have introduced another fusion scheme as feature-based fusion which is operates in

the eigenspace domain. In this case also they employ a simple and general framework based on Genetic Algorithms (GAs) to find an optimum fusion strategy and for this feature level fusion they use eigenspace domain, which involves combining the eigenfeatures from the visible and IR images. First they compute two eigenspaces, of visible IR face images. They generate the fused image by applying the GAs to the eigenfeatures from the IR and visible eigenspace. In another scheme, M. Hanif et al [15] proposed data fusion of visual and thermal images using Gabor filtering technique. It has been found that by using the proposed fusion technique Gabor filter can recognize face even with variable expressions and light intensities, but not in extreme condition. They have designed the data fusion technique in the spatial and DWT domain. For the spatial domain they have used a weight to generate the fusion image. But in case of DWT image fusion, they first decompose both thermal and visual image up to level-n and then choose the absolute maximum at every pixel for both the thermal and visual image to generate the fusion image and this work is implemented using equinox face database. Jingu Heo et al [11] describe two types of fusion method as Data fusion and Decision Fusion. In data fusion, it produces an illumination-invariant face images by integrating the visual and thermal images. A weighted average technique is used for data fusion. In the data fusion they first detect the eyeglasses and remove the eyeglasses by a fitting an eclipse and then merge it with visual image. They used n non-iterative eclipse fitting algorithm for it. They have implemented Decision fusion with highest matching score (Fh) and Decision fusion with average matching score (Fa). Comparison results on three fusion-based face recognition techniques showed that fusion-based face recognition techniques outperformed individual visual and thermal face recognizers under illumination variations and facial expressions. From them Decision fusion with average matching score consistently demonstrated superior recognition accuracies as per their results. The organization of the rest of this paper is as follows: in section 2, the overview of the system is discussed, in section 3 experimental results and discussions are given. Finally, section 4 concludes this work.

## 2. SYSTEM OVERVIEW

The block diagram of the system is given in Figure 1. All the processing steps used in this work are shown in the block diagram. In the first step, decomposition of both the thermal and visual images up to level five has been done using Haar and Daubechies wavelet transformations. Then fused images are generated using respective decomposed visual and thermal images of two different wavelets. These transformed images separated into two groups namely, training set and testing set are fed into a multilayer perceptron for classification. Before, feeding them into the network, those are passed through Principal Component Analysis (PCA) for dimensionality reduction.

### 2A. Wavelet Transforms and Fusion

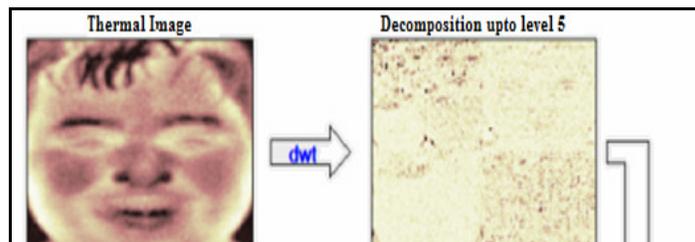
Haar wavelet is recognized as the first known wavelet. It is same as Daubechies wavelet (db1). The Haar wavelet is proposed by Alfred Haar [17]. Haar wavelet is a certain sequence of function.

The Haar wavelet's mother wavelet function  $\psi(t)$  can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2. \\ -1 & 1/2 \leq t < 1. \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

and its scaling function  $\phi(t)$  can be described as

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$



**FIGURE 1:** System Overview.

It is an orthogonal and bi-orthogonal wavelet. The technical disadvantage of Haar wavelet is that it is not continuous, and therefore not differentiable. In functional analysis, the Haar systems denotes the set of Haar wavelets

$$\{t \mapsto \psi_{n,k}(t) = \psi(2^n t - k); n \in \mathbb{N}, 0 \leq k < 2^n\} \quad (3)$$

The Haar system (with the natural ordering) is further a Schauder basis for the space  $L_p[0, 1]$  for  $1 \leq p < +\infty$ . This basis is unconditional for  $p > 1$  [18], [19].

The Haar transform is the simplest of the wavelet transforms. This transform cross-multiplies a function against the Haar wavelet with various shifts and stretches. The Haar transform is derived from the Haar matrix. An example of 3x3 Haar matrixes is shown below:

$$H_1 = \frac{1}{\sqrt{4}} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & -\mathbf{1} & -\mathbf{1} \\ \sqrt{2} & -\sqrt{2} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (4)$$

The Haar transform can be thought of as a sampling process in which rows of the transform matrix act as samples of finer and finer resolution.

Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. In two

dimensions, the difference of information between the approximations  $A_{2^{j+1}}f$  and  $A_{2^j}f$  is therefore characterized by the sequences of inner products

$$D_{2^j}^1 f = (\langle f(x, y), \Psi_{2^j}^1(x - 2^{-j}n, y - 2^{-j}m) \rangle)_{(n,m) \in Z^2} \quad (5)$$

$$D_{2^j}^2 f = (\langle f(x, y), \psi_{2^j}^2(x - 2^{-j}n, y - 2^{-j}m) \rangle)_{(n,m) \in Z^2} \quad (6)$$

$$D_{2^j}^3 f = (\langle f(x, y), \psi_{2^j}^3(x - 2^{-j}n, y - 2^{-j}m) \rangle)_{(n,m) \in Z^2} \quad (7)$$

Each of these sequences of inner products can be considered as an image. In Figure 2(a) showing level 2 decomposition,  $H_2$  gives the vertical higher frequencies (horizontal edges),  $V_2$  gives the horizontal higher frequencies (vertical edges), and  $D_2$  gives higher frequencies in both directions (corners). Let us suppose that initially we have an image  $A_1 f$  measured at the resolution 1. For any  $J > 0$ , this discrete image can be decomposed between the resolution 1 and  $2^J$ , and completely represented by the  $3J + 1$  discrete images

$$(A_{2^{-j}}f, (D_{2^j}^1 f)_{-j \leq j \leq -1}, (D_{2^j}^2 f)_{-j \leq j \leq -1}, (D_{2^j}^3 f)_{-j \leq j \leq -1}) \quad (8)$$

This set of images is called an orthogonal wavelet representation in two dimensions [20]. The image  $A_{2^{-j}}f$  is a coarse approximation, and the images  $H_2$  give the image details for different orientations and resolutions. If the original image has  $N^2$  pixels, each image  $A_{2^j}f$ ,  $H_2$ ,  $V_2$ ,  $D_2$  has  $2^j * N^2$  pixels ( $j < 0$ ). The total number of pixels of an orthogonal wavelet represents is therefore equal to  $N^2$ . It does not increase the volume of data. This is due to the orthogonality of the representation. Such decompositions were done through Matlab and shown in equations 9 and 10.

$$[cA, cH, cV, cD] = \text{dwt2}(X, \text{'wname'}) \quad (9)$$

$$[cA, cH, cV, cD] = \text{dwt2}(X, \text{Lo\_D}, \text{Hi\_D}) \quad (10)$$

Equation (9), 'wname' is the name of the wavelet used for decomposition. We used 'haar' and 'db2' as wavelet. Equation (10) Lo\_D (decomposition low-pass filter) and Hi\_D (decomposition high-pass filter) wavelet decomposition filters [21], [22], [23], [35], [39].

The more the decomposition scheme is being repeated, the more the approximation of images concentrates in the low frequency energy. During the decomposition process it actually down-sampling the rows and columns of an image. Firstly it down-samples the rows (keep one out of two) and then down-samples the columns (keeps one out of two). The downsampling process for both the images is same. After decomposing the thermal and visual image we have to generate the decomposed fused image by providing fusion methods. Then, at each level select the absolute 'maximum' at every pixel from both the images for approximate co-efficient (cA) and absolute 'minimum' at every pixel from both the images for three details co-efficient (cH, cV, cD). Let us take an example, if T is the thermal and V is the visual image with same pose and illumination than

$$\text{'max': } D = \text{abs}(T) \geq \text{abs}(V); C = T(D) + V(\sim D) \quad (11)$$

$$\text{'min': } D = \text{abs}(T) \leq \text{abs}(V); C = T(D) + V(\sim D) \quad (12)$$

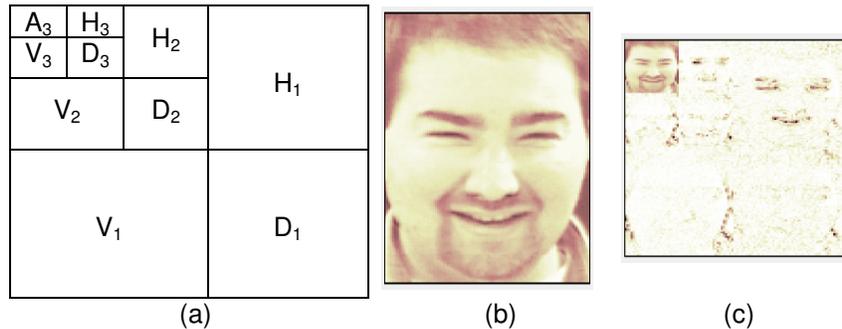
To generate the fused image (C) of co-efficient it will add that value which is deducted from the visual image during the calculation of absolute value (D) of thermal (T) and visual (V) image for all the coefficients using fusion method shown at equation (11) & (12).

Consequently the reconstruction process is performed using inverse of wavelet transformations to generate synthesized fused image.

$$X = \text{idwt2}(cA, cH, cV, cD, 'wname') \tag{13}$$

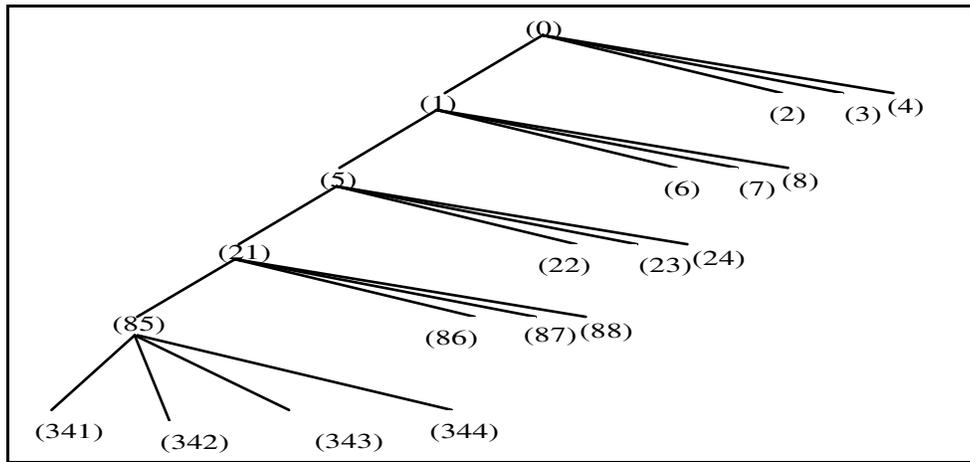
$$X = \text{idwt2}(cA, cH, cV, cD, \text{Lo\_R}, \text{Hi\_R}) \tag{14}$$

IDWT uses the wavelet 'wname' to compute the single-level reconstruction of an Image X, based on approximation matrix (cA) and detailed matrices cH, cV and cD (horizontal, vertical and diagonal respectively). By the equation no (14), we can reconstruct the image using filters Lo\_R (reconstruct low-pass) and Hi\_R (reconstruct high-pass) and 'haar' and 'db2' as the time of reconstruction of an image [33].



**FIGURE 2:** (a) Shown one of the third level decomposition of level 5 (b) Original image used in the decomposition (c) Orthogonal wavelet representation of a sample image.

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one image is broken down into many lower resolution components. This is called the wavelet decomposition tree. In this work, decomposition was done up to level five using Haar and Daubechies (db2) wavelet transformations, as shown in Figure 3. Average of corresponding transform coefficients from visual and thermal images gives the matrix of fused coefficients, which when transformed into the image in spatial domain by inverse wavelet transform produces fused face image. These fused images thus found are passed through PCA for dimensionality reduction, which is described next.



**FIGURE 3:** Wavelet Decomposition Tree.

## 2B. Dimensionality Reduction

Principal component analysis (PCA) is based on the second-order statistics of the input image, which tries to attain an optimal representation that minimizes the reconstruction error in a least-squares sense. Eigenvectors of the covariance matrix of the face images constitute the eigenfaces. The dimensionality of the face feature space is reduced by selecting only the eigenvectors possessing significantly large eigenvalues. Once the new face space is constructed, when a test image arrives, it is projected onto this face space to yield the feature vector—the representation coefficients in the constructed face space. The classifier decides for the identity of the individual, according to a similarity score between the test image's feature vector and the PCA feature vectors of the individuals in the database [27], [30], [36], [37], [34].

## 2C. Artificial Neural Network Using Backpropagation With Momentum

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an “expert” in the category of information it has been given to analyze. The Back propagation learning algorithm is one of the most historical developments in Neural Networks. It has reawakened the scientific and engineering community to the modeling and processing of many quantitative phenomena. This learning algorithm is applied to multilayer feed forward networks consisting of processing elements with continuous differentiable activation functions. Such networks associated with the back propagation learning algorithm are also called back propagation networks [24], [25], [26], [27], [28], [29], [38].

## 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

This work has been simulated using MATLAB 7 in a machine of the configuration 2.13 GHz Intel Xeon Quad Core Processor and 16384.00 MB of Physical Memory. We analyze the performance of our algorithm using the IRIS thermal / visual face database. In this database, all the thermal and visible unregistered face images are taken under variable illuminations, expressions, and poses. The actual size of the images is 320 x 240 pixels (for both visual and thermal). Total 30 classes are present in that database [31]. Some thermal and visual images and their corresponding fused images for Haar and Daubechies (db2) wavelets are shown in Figure 4 and Figure 5 respectively.

To compare results of Haar and Daubechies (db2) wavelets, fusion of visual and thermal images were done separately. For that purpose 200 thermal and 200 visual images were considered. We have increased the size of testing image for both the wavelet. Firstly we have tested our system using 100 fused images using two different wavelet and we increase 10 images per class for 10 different classes i.e. total 100 image. The database is categorized into 10 classes based on changes in illumination and expressions. The classes with illuminations and expression changes are class-1, class-2, class-3, class-4, class-6, class-7, and class-9, whereas class-5, class-8 and class-10 are with changes in expressions only. Both the techniques were applied and experimental results thus obtained are shown in Figure 6.

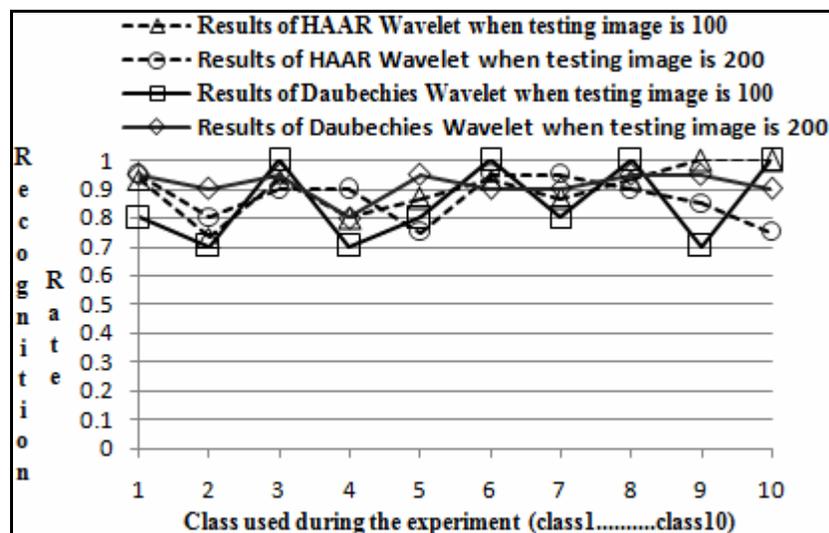
After increasing the number of testing images the recognition rates are increased for both the wavelets. Previously we have average recognition rate is 85% for Haar wavelet and 90.019% for Daubechies wavelet when the number of testing image is 100 (i.e. 10 image per class). Now we have average recognition rate is 87% for Haar and 91.5% for Daubechies wavelet after increase the number of testing image 100 to 200 (i.e. 20 image per class).



**FIGURE 4:** Sample images of IRIS database (a) Thermal images (b) Visual images (c) Corresponding Fused images using Haar Wavelet transform.



**FIGURE 5:** Sample images of IRIS database (a) Thermal images (b) Visual images (c) Corresponding Fused images using Daubechies Wavelet transform.



**FIGURE 6:** Recognition rate of Haar and Daubechies Wavelet.

In the Figure 6, all the recognition rates for Haar as well as Daubechies wavelet transformations have been shown. In this figure “dotted” line indicates results using Haar wavelet (with round and triangle shape marker) and “solid” line indicates results using Daubechies Wavelet (with box and diamond shape marker). From this figure, it can be easily inferred that overall results using “Daubechies” (db2) wavelet produces better result than “Haar” wavelet.

Final results have been computed as average of all the categories and given along with results of other recent methods in Table II. Here, we find results using Daubechies wavelets (db2) shows comparable results.

**TABLE 2:** Comparison between different Fusion Techniques.

Image Fusion Technique		Recognition Rate			
Present method	Haar	85%	When testing image 100	87%	When testing image 200
	Daubechies	90.019%		91.5%	
Simple Spatial Fusion[15]		91.00%			
Fusion of Thermal and Visual [32]		90.00%			
Abs max selection in DWT[15]		90.31%			
Window based absolute maximum selection [15]		90.31%			
Fusion of Visual and LWIR + PCA [6]		87.87%			

#### 4. CONSLUSION

In this paper, a comparative study on two different wavelet transformations, namely, Haar and Daubechies (db2) for image fusion has been presented. After completion of fusion, images were projected into an eigenspace. Those projected fused eigenfaces are classified using a Multilayer Perceptron. Eigenspace is constituted by the images belong to the training set of the classifier, which is a multilayer perceptron. The efficiency of the scheme has been demonstrated on IRIS system has achieved a maximum recognition rate of 100% using both the wavelets i.e. Haar and Daubechies. But on an average 85% and 90.019% (using 100 testing images) and 87% and 91.5% (using 200 testing images) recognition rates were achieved for Haar wavelet and Daubechies wavelet respectively.

#### 5. ACKNOWLEDGMENT

Mr. M. K. Bhowmik is thankful to the project entitled “Development of Techniques for Human Face Based Online Authentication System Phase-I” sponsored by Department of Information Technology under the Ministry of Communications and Information Technology, New Delhi-110003, Government of India Vide No. 12(14)/08-ESD, Dated 27/01/2009 at the Department of Computer Science & Engineering, Tripura University-799130, Tripura (West), India for providing the necessary infrastructural facilities for carrying out this work. The first author would also like to thank Prof. Barin Kumar De, H.O.D. Physics and Dean of Science of Tripura University, for his kind support to carry out this research work.

#### 6. REFERENCES

1. A. Pentland and T. Choudhury. “Face Recognition for Smart Environments”. IEEE Computer. 33(2):50-55, 2000
2. P. Phillips et al., The Feret Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing. May 1998. pp. 295-306

3. L. Wiskott et al., "Face Recognition by Elastic Bunch Graph Matching". Trans. IEEE Pattern Analysis and Machine Intelligence. 19(7):775-779. 1997
4. B. Moghaddam and A. Pentland. "Probabilistic Visual Recognition for Object Recognition". Trans. IEEE Pattern Analysis and Machine Intelligence. 19(7):696-710. 1997
5. P. Penev and J. Atick. "Local Feature Analysis: A General Statistical Theory for Object Representation". Network: Computation in Neural Systems. Mar. 1996, pp. 477-500
6. I. Pavlidis and P. Symosek. "The Imaging Issue in an Automatic Face/Disguise Detecting System". Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (CVBVS 2000)
7. Y. Adini, Y. Moses, and S. Ullman. "Face recognition: The problem of compensating for changes in illumination direction". IEEE Trans. On Pattern Analysis and Machine Intelligence. Vol. 19, No. 7. pp.721-732, 1997
8. A. Ross and A. Jain. "Information fusion in biometrics". Pattern Recognition Letters. Vol. 24, No. 13. pp.2115-2125. 2003
9. T. K. Ho, J. J. Hull, and S. N. Srihari. "Decision combination in multiple classifier systems". IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 16. No. 1. pp.66-75. 1994
10. C. Sanderson and K. K. Paliwal. "Information fusion and person verification using speech and face information". IDIAP Research Report 02-33. Martigny. Switzerland. 2002
11. J. Heo, S. G. Kong, B. R. Abidi and M. A. Abidi. "Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition". Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Vol. 8. pp. 122-122. USA. 2004
12. I. Pavlidis, P. Symosek, B. Fritz. "A Near Infrared Fusion Scheme for Automatic Detection of Humans". U.S. Patent Pending. H16-25929 US. filed in September 3, 1999
13. A. Gyaourova, G. Bebis and I. Pavlidis. "Fusion of infrared and visible images for face recognition". Lecture Notes in Computer Science. 2004.Springer. Berlin
14. A. Selinger, D. Socolinsky. "Appearance-based facial recognition using visible and thermal imagery: a comparative study". Technical Report 02-01. Equinox Corporation. 2002
15. M. Hanif and U. Ali. "Optimized Visual and Thermal Image Fusion for Efficient Face Recognition". IEEE Conference on Information Fusion. 2006
16. G. Bebis, A. Gyaourova, S. Singh and I. Pavlidis. "Face Recognition by Fusing Thermal Infrared and Visible Imagery". Image and Vision Computing. Vol. 24, Issue 7. July 2006. pp. 727-742
17. A. Haar. "Zur Theorie der orthogonalen Funktionensysteme (German)"
18. <http://eom.springer.de/O/o070308.htm>
19. W. G. Gilbert, S. Xiaoping. "Wavelets and Other Orthogonal Systems". 2001
20. T. Paul. "Affine coherent states and the radial Schrodinger equation. Radial harmonic oscillator and hydrogen atom". Preprint

21. I. Daubechies. Ten lectures on wavelets, CBMS-NSF conference series in applied mathematics. SIAM Ed. 1992
22. S. G. Mallat. "A theory for multiresolution signal decomposition: The wavelet representation". IEEE Trans. Pattern Anal. Machine Intell., vol.11. pp. 674-693. July 1989
23. Y. Meyer. Wavelets and operators. Cambridge Univ. Press. 1993
24. M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D.K. Basu and M. Kundu. "Classification of Polar-Thermal Eigenfaces using Multilayer Perceptron for Human Face Recognition". in Proceedings of the 3rd IEEE Conference on Industrial and Information Systems (ICIIS-2008). IIT Kharapur. India. Dec. 8-10. 2008. Page no: 118
25. D. Bhattacharjee, M. K. Bhowmik, M. Nasipuri, D. K. Basu and M. Kundu. "Classification of Fused Face Images using Multilayer Perceptron Neural Network". Proceedings of International Conference on Rough sets. Fuzzy sets and Soft Computing. November 5-7, 2009. organized by Deptt. of Mathematics. Tripura University. pp. 289 – 300
26. M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D.K. Basu and M. Kundu. "Classification of Log Polar Visual Eigenfaces using Multilayer Perceptron". Proceedings of The 2nd International Conference on Soft computing (ICSC-2008). IET. Alwar. Rajasthan. India. Nov. 8-10. 2008. pp:107-123
27. M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu and M. Kundu. "Human Face Recognition using Line Features". proceedings of National Seminar on Recent Advances on Information Technology (RAIT-2009) Feb. 6-7,2009. Indian School of Mines University. Dhanbad. pp: 385
28. M. K. Bhowmik. "Artificial Neural Network as a Soft Computing Tool – A case study". In Proceedings of National Seminar on Fuzzy Math. & its application. Tripura University. Nov. 25 – 26, 2006. pp: 31 – 46
29. M. K. Bhowmik, D. Bhattacharjee and M. Nasipuri. "Topological Change in Artificial Neural Network for Human Face Recognition". Proceedings of National Seminar on Recent Development in Mathematics and its Application. Tripura University. Nov. 14 – 15, 2008. pp: 43 – 49
30. H. K. Ekenel and B. Sankur. "Multiresolution face recognition". Image and Vision Computing. 2005. Vol. 23. pp. 469 – 477
31. <http://www.cse.ohio-state.edu/otcbvs-bench/Data/02/download.html>
32. A. Singh, G. B. Gyaourva and I. Pavlidis. "Infrared and Visible Image Fusion for Face Recognition". Proc. SPIE. vol.5404. pp.585-596. Aug.2004
33. V. S. Petrovic and C. S. Xydeas. "Gradient-Based Multiresolution Image Fusion". IEEE Trans. On image Processing. Vol. 13. No. 2. Feb. 2004.
34. A. Rahim & J. V. Charangatt. "Face Hallucination using Eigen Transformation in Transform Domain". International Journal of Image Processing (IJIP) Vol. 3. Issue 6, pp. 265-282.
35. J. Shen and G. Strang. "Applied and Computational Harmonic Analysis". 5(3). "Asymptotics of Daubechies Filters, Scaling Functions and Wavelets"
36. M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu and M. Kundu. "Classification of fused images using Radial basis function Neural Network for Human Face Recognition".

Proceedings of The World congress on Nature and Biologically Inspired Computing (NaBIC-09). Coimbatore, India. published by IEEE Computer Society. Dec. 9-11, 2009. pp. 19-24

37. M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D.K. Basu and M. Kundu. "Image Pixel Fusion for Human Face Recognition". appeared in International Journal of Recent Trends in Engineering [ISSN 1797-9617]. published by Academy Publishers. Finland. Vol. 2. No. 1, pp. 258–262. Nov. 2009
38. D. Bhattacharjee, M. K. Bhowmik, M. Nasipuri, D. K. Basu & M. Kundu. "A Parallel Framework for Multilayer Perceptron for Human Face Recognition". International Journal of Computer Science and Security (IJCSS), Vol. 3. Issue 6, pp. 491-507.
39. Sathesh & Samuel Manoharan. "A Dual Tree Complex Wavelet Transform Construction and its Application to Image Denoising". International Journal of Image Processing (IJIP) Vol. 3. Issue 6, pp. 293-300.

## Event-Handling Based Smart Video Surveillance System

**Fadhlan Hafiz**

*Faculty of Engineering  
International Islamic University Malaysia  
53100 Kuala Lumpur, Malaysia*

fadhlan\_hafiz@yahoo.co.uk

**A.A. Shafie**

*Faculty of Engineering  
International Islamic University Malaysia  
53100 Kuala Lumpur, Malaysia*

aashafie@iiu.edu.my

**M.H. Ali**

*Faculty of Engineering  
International Islamic University Malaysia  
53100 Kuala Lumpur, Malaysia*

hazratalidu07@yahoo.com

**Othman Khalifa**

*Faculty of Engineering  
International Islamic University Malaysia  
53100 Kuala Lumpur, Malaysia*

khalifa@iiu.edu.my

---

### Abstract

Smart video surveillance is well suited for a broad range of applications. Moving object classification in the field of video surveillance is a key component of smart surveillance software. In this paper, we have proposed reliable software with its large features for people and object classification which works well in challenging real-world constraints, including the presence of shadows, low resolution imagery, occlusion, perspective distortions, arbitrary camera viewpoints, and groups of people. We have discussed a generic model of smart video surveillance systems that can meet requirements of strong commercial applications and also shown the implication of the software for the security purposes which made the whole system as a smart video surveillance. Smart surveillance systems use automatic image understanding techniques to extract information from the surveillance data and handling events and stored data efficiently.

**Keywords:** Smart system, Human detection, Video surveillance, Object classification, Event handling

---

### 1. INTRODUCTION

Many applications of smart camera networks relate to surveillance and security system. Although applications of smart camera networks are fairly broad, they rely on only a few elementary estimation tasks. These tasks are object detection, object localization, target tracking, and object classification. Many surveillance methods are based on a general pipeline-based framework;

moving objects are first detected, then they are classified and tracked over a certain number of frames, and, finally, the resulting paths are used to distinguish “normal” objects from “abnormal” ones. In general, these methods contain a training phase during which a probabilistic model is built using paths followed by “normal” objects [1]. Smart video surveillance systems achieve more than motion detection. The common objectives of smart video surveillance systems are to detect, classify, track, localize and interpret behaviors of objects of interest in the environment [2]. Some countries have implemented cabin video-surveillance systems which require pilot’s involvement in their operation as there is no intelligence built in them. Moreover, after 20 minutes of surveillance, in all such non-automated vigilance systems, the human attention to the video details degenerate into an unacceptable level and the video surveillance becomes meaningless. Thus there is an increasing demand for intelligent video surveillance systems with automated tracking and alerting mechanism [3].

Relevant work in this area include shape-base techniques which exploit features like size, compactness, aspect ratio, and simple shape descriptors obtained from the segmented object [4, 5]. The smart camera delivers a new video quality and better video analysis results, if it is compared to existing solutions. Beside these qualitative arguments and from a system architecture point of view, the smart camera is an important concept in future digital and heterogeneous third generation visual surveillance systems [6]. Similarly, visual object classification is a key component of smart surveillance systems. The ability to automatically recognize objects in images is essential for a variety of surveillance applications, such as the recognition of products in retail stores for loss prevention, automatic identification of vehicle license plates, and many others.

In this paper, we address a simplified two-class object recognition problem: given a moving object in the scene, our goal is to classify the object into either a person (including groups of people) or a vehicle. This is a very important problem in city surveillance, as many existing cameras are pointing to areas where the majority of moving objects are either humans or vehicles. In our system, this classification module generates metadata for higher-level tasks, such as event detection (e.g., cars speeding, people loitering) and search (e.g., finding red cars in the video). We assume static cameras, and thus benefit from background modeling algorithms to detect moving objects [7]. Wolf et al. identified smart camera design as a leading-edge application for embedded systems research [8]. In spite of these simplifications, the classification problem still remains very challenging, as we desire to satisfy the following requirements:

- (a) Real-time processing and low memory consumption
- (b) The system should work for arbitrary camera views
- (c) Correct discrimination under different illumination conditions and strong shadow effects
- (d) Able to distinguish similar objects (such as vehicles and groups of people).

Our approach to address these issues consists of three elements [7]:

- (a) Discriminative features,
- (b) An adaptation process, and
- (c) An interactive interface

We also have defined 3 challenges that need to be overcome:

### ***The multi-scale challenge***

This is one of the biggest challenges of a smart surveillance system. Multi-scale techniques open up a whole new area of research, including camera control, processing video from moving object, resource allocation, and task-based camera management in addition to challenges in performance modeling and evaluation.

**The contextual event detection challenge**

This challenge is mostly on using knowledge of time and deployment conditions to improve video analysis, using geometric models of the environment and other object and activity models to interpret events, and using learning techniques to improve system performance and detect unusual events.

**The large system deployment challenge**

It has several challenges include minimizing the cost of wiring, meeting the need for low-power hardware for battery-operated camera installations, meeting the need for automatic calibration of cameras and automatic fault detection, and developing system management tools.

**2. METHODOLOGY**

**2.1 Camera Selection**

Since the multi-scale challenge incorporates the widest range of technical challenges, we present the generic architecture of a multi-scale tracking system. The architecture presented here provides a view of the interactions between the various components of such a system. We present the concepts that underlie several of the key techniques, including detection of moving objects in video, tracking, and object classification..

Selections of the surveillance features are very important for a smart surveillance software or smart surveillance network. Camera and its components selection depends on the users. We have classified the camera based on the following criteria summarized in Table 1 and our selected hardware in Table 2.

Area		Complexity		Number of Cameras		Environment		Types of Camera	
Small	Large	Simple	Complex	Single	Multiple	Day/Bright	Night/Dark	IP based	Smart

Table 1: Classification of the camera

Components	Description	Quantity
1.Outdoor Box Camera	1/3" Sony Super HAD, 380TVL / 0lx / 3.6mm / 22IR LEDs / 12VDC	3
2.Dome Camera	1/3" Sony Super HAD	1
3. DVR Card	4 Channel 3rd Party DVR Card	1
4. Cable	Lay coaxial cable	50 ft
5.DVR	DVR configuration	1
7.Main Server	High Speed CPU	1

Table 2: The hardware used for our system

**2.2 Camera Placement**

We have developed our software to detect the movements and display features on the monitor as we want to view. A way to use the sun positions to determine the focal length, zenith and azimuth angles of a camera. The idea is to minimize the re-projection error, that is, minimize the distance between the sun labels and the sun position projected onto the image plane. If we have several observations, we can know the solution that gives the best data in a least-squares sense as in the following

$$\min_{f_c, \theta_c, \phi_c} \sum_{i=1}^N ||\mathbf{p}_i - \mathcal{P}(R^{-1}\mathbf{s}_i)||^2, \quad \text{where } \mathbf{p}_i = [ u_i \ v_i ]^T$$

Where,  $P_i$  is the projection operator. Our goal is to find the camera parameters ( $f_c, \theta_c, \Phi_c$ ) that will best align the sun labels  $p_i$  with the rotated sun position. The following equation is used to find those parameters and summarized in Table 3:

$$\min_{f_c, \theta_c, \Phi_c} \sum_{i=1}^N \left( u_i - \frac{-fy'_{s,i}}{x'_{s,i}} \right)^2 + \left( v_i - \frac{fz'_{s,i}}{x'_{s,i}} \right)^2,$$

Name	Symbol	Default Value
Focal Length	$f_c$	1000 px
Number of Image	N	20
Camera Zenith	$\theta_c$	90
Camera Azimuth	$\Phi_c$	0

Table 3: Necessary camera parameters

Camera positioning is very important for clear video image. Our test camera positioning is shown in the following figures. Camera angles are very important parameters to be considered. Depending on the application we can choose either 45 degree or 90 degree angle's camera. The focus point will determine the area under surveillance. Figure 1 and 2 bellow shows the front view and side view of the testing area. In this area we used four different cameras. 1 is dome camera and the rest 3 are box cameras.

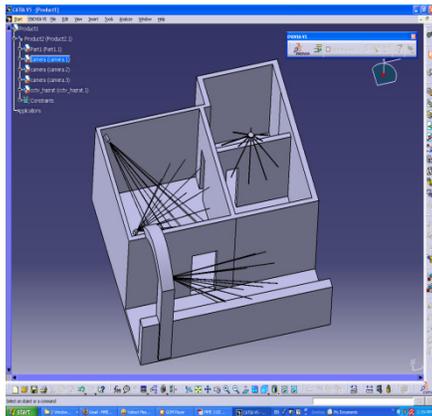


Figure 1 : Front view of the test area

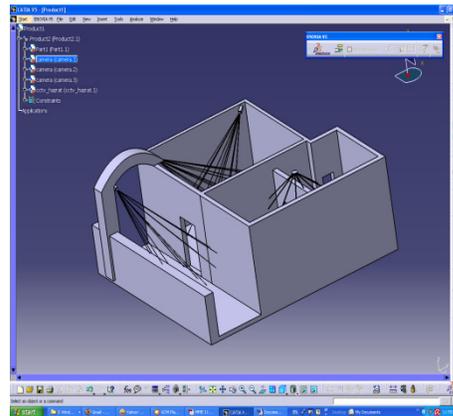


Figure 2 : Side view of the test area

### 2.3 Smart camera video sequences handling

Figure 3 shows the internal structure of our smart surveillance system. From the figure we can see how it works in the practical field. First the video will be recorded and then will be saved in the specific destination in hard drive, next it will use the object detection algorithm, after that it will follow the multi object tracking and classification method. Using this method we can sort the types of the objects detected by the camera. Finally Event classification will be used to index the data in the index storage from where we can retrieve the active object efficiently.

### 2.4 PC and software selection

The PC used for this project is Intel Quad Core Q9550, 2.83GHz with 4GB of memory and the software used for the system development are:

- Windows Vista.
- Microsoft Visual Studio 2008
- SQL Server Database
- AForge Image Processing Library (open source)

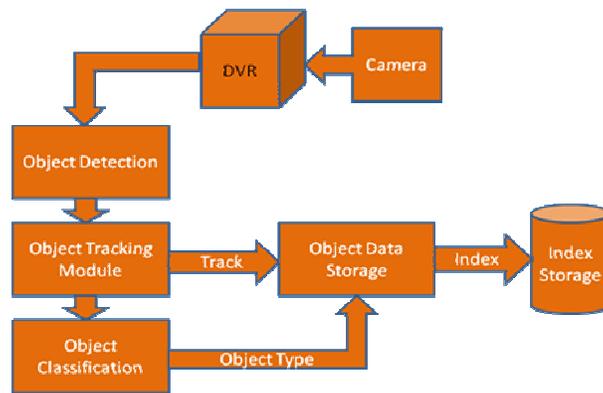


Figure 3: Overall internal structure of smart surveillance system

### 3. DETECTION, EVENT CLASSIFICATION AND DATA STORAGE

#### 3.1 Detection algorithm

The overall human detection technique proposed in this system is discussed in details in [9] and it can be classified into two parts, (1) Image Pre-processing and (2) Segmentation. Image pre-processing includes frame processing, foreground segmentation, and binarization. While Segmentation includes Shadow removal, morphological operation, noise removal, and size filtering. The overall steps employed are summarized in Figure 4 and also illustrated with example in Figure 5.

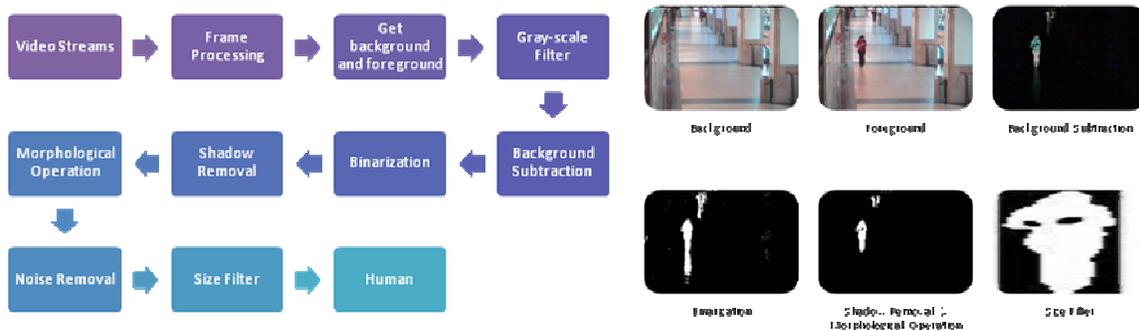


Figure 4: Human detection process flow Graphical overview

Figure 5

#### 3.2 Event classification

This classification system classifies the events that occur within a space that may be monitored by one or more cameras. The classification and associated information data processing are carried out in real-time. The following is descriptions of few elements in event classification:

- Event: An event is defined as any detection, movement or event that occurs within the camera's secured area.
- Event ID: This is a unique number which identifies a specific event. Main index to Video ID and Human ID. Used in *Indexing and Data Retrieval*
- Time: Time at which the event occurs.
- Location: Location where the event occurs
- Video ID: This is a unique number which identifies a specific video recorded active camera and is the index to the physical video storage information.
- Human ID: This is a unique number which identifies a specific human who appears within camera focus area. Index to the image, appearance information, and tracking data

### 3.3 Data Storage System

The storage system in this surveillance system comprises of two components: (a) physical storage and (b) database. Video data from the cameras are saved into the hard disk and this falls into physical storage category. The video files are saved in frames, and not in usual video format like *.wmv* or *.avi* format. This is because the system, which is developed in .NET framework, cannot save the video data while analyzing the video streams for detection purpose. This feat is applicable in COM but not in .NET. As a solution, we copy the image frame, one at a time, and save the images on the hard disk just before the detection class processes the images. These images are saved in JPEG format. All the data associated to that video such as the physical location of the images, the time stamp, duration, and frame speed (fps), are saved in the video database and each video is assigned a unique ID as the key for that information. This information is saved in the database using SQL Server Database.

### 3.4 Data Indexing and Retrieval

This is the core and the brain of this surveillance system. Those IDs which were assigned earlier in real-time, such as Event ID, Video ID and Human ID are used to index all the data in the database and for automated data retrieval when requested by the user. Every ID in this system is linked to each other so that user can access all necessary information like human appearance, time, track, and video data in one comprehensive and inter-related search. The flow of *Data Indexing and Retrieval* is illustrated in Figure 6. Once process objects are identified, the data extracted from objects are stored together with the unique ID which indexed earlier into database. Then the data retrieval will handle the request for data using the assigned ID as the key for retrieval of data from database and then retrieve all the associated data.

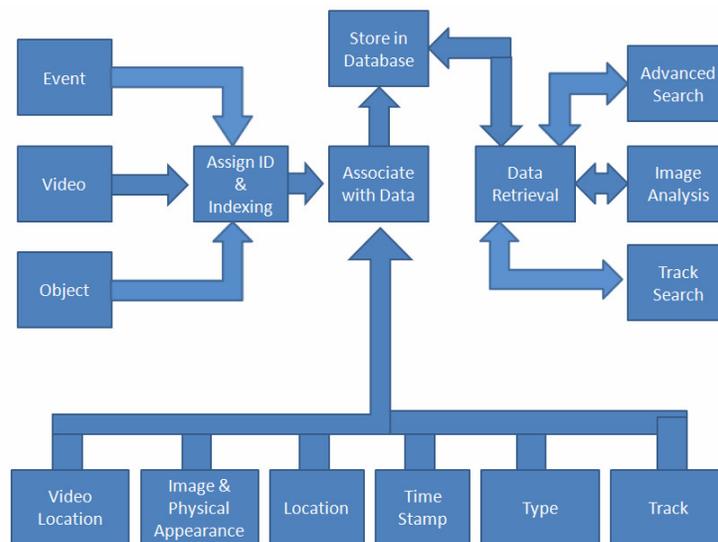


Figure 6 : Data Indexing and Retrieval process flow

## 4. SYSTEM OVERVIEW

Our developed system gives more advance result compared to the existing camera system. It has varieties of displaying options which will easily help to analyze data in any time in the region of the cameras. The functions can be divided into 5 categories as follows:

### 1. Surveillance Mode

In this mode, user can connect to all available cameras, and change the surveillance settings for each camera, such as motion detection, human detection, security settings and secured parameters. This mode also enables administrative control such as system lock, password management and profile selection. It summarizes all the running components and the result of each component such as number of events, number of human detected, camera frames per second, and intruder detection. It also has full screen mode. This mode is illustrated in Figure 7 and Figure 8.

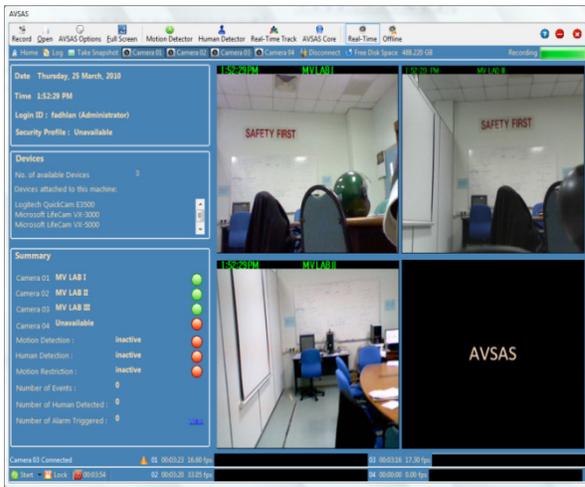


Figure 7: Surveillance in windowed mode



Figure 8: Surveillance in full screen

### 2. Core Databases

All events and videos are saved into the hard disk and their details and locations are indexed into the event and video database. This database will keep track of any object such as human, detected by the surveillance systems and link the events to corresponding video. Human and video database are shown in Figure 9 and 10 respectively. This will enable the user to easily search for any event in interest such as human at specific time and place automatically without the hassle of looking through terabytes of video data.

### 3. Advanced Search

*Advanced Search* is a sophisticated search engine for searching any event, human or object at any time and place. The resulting event from this search can be used to directly open the corresponding video in playback mode or to view the track of the requested human. Each event bears a unique ID and unique time stamp so that a user can directly access the video of the event using the *Data Indexing and Retrieval* system discussed earlier. *Advanced Search* with human images result is illustrated in Figure 11

### 4. Playback Mode

*Playback Mode* as in Figure 12 can carry out several video analysis such as automated video searching according to time constraints. It has playback feature from normal surveillance system

such as play, pause, stop, fast forward, and rewind with adjustable magnitude. The user can open up to maximum 4 video at a time where each video window have the same time stamp but corresponds to different camera locations.

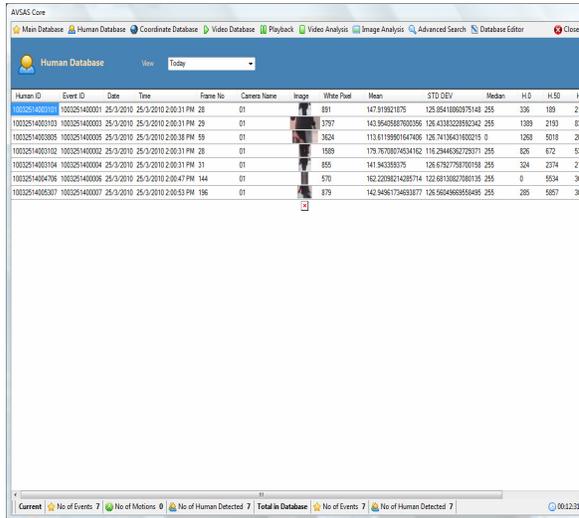


Figure 9: Human Database

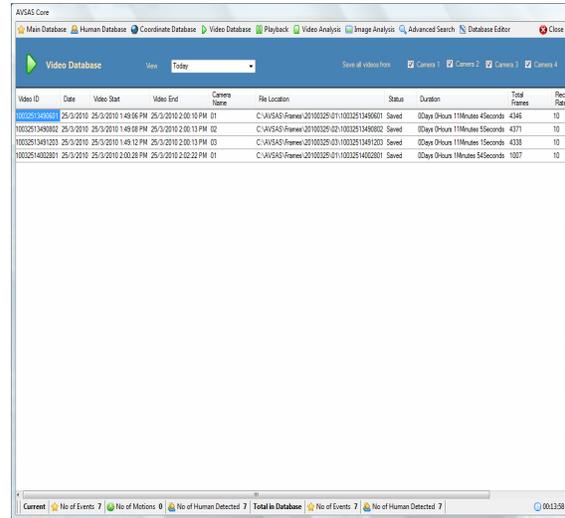


Figure 10: Video Database

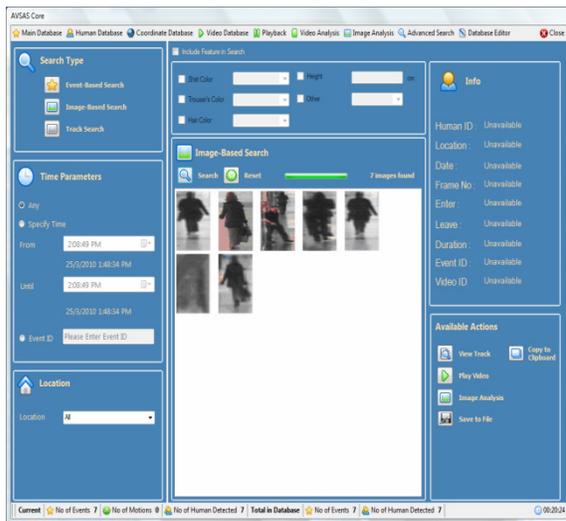


Figure 11: Advanced Search

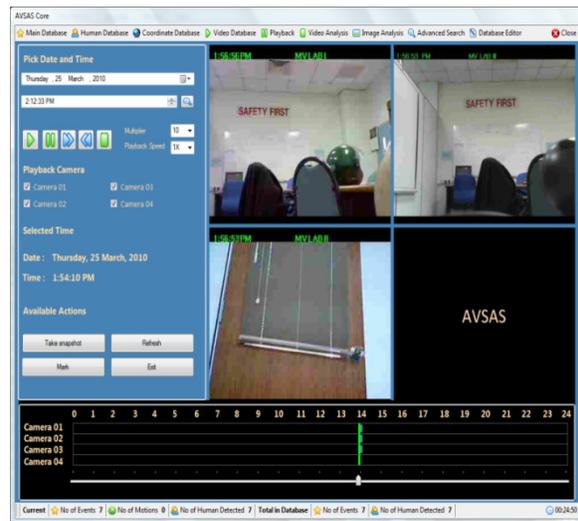


Figure 12: Playback Mode

## 5. Video and Track Analysis

*Video Analysis* is linked to *Advanced Search* and it handles the playback of the video containing the event searched in *Advanced Search*. Any video opened in this mode will list all the events contained in the video as indexed by the system beforehand. User can click any event listed to go directly to the video at the time it happens. The video controls of this mode are much less the same with *Playback Mode*. This mode is illustrated in Figure 13. While *Track Analysis* enables user to view previously recorded track of any human that appears in the video. This mode is illustrated in Figure 14.

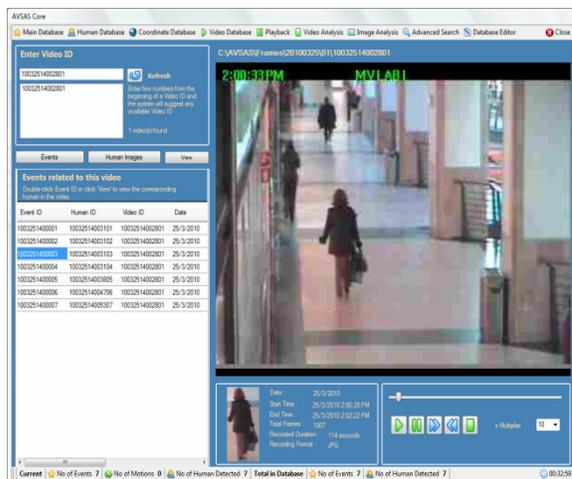


Figure 13: Video Analysis

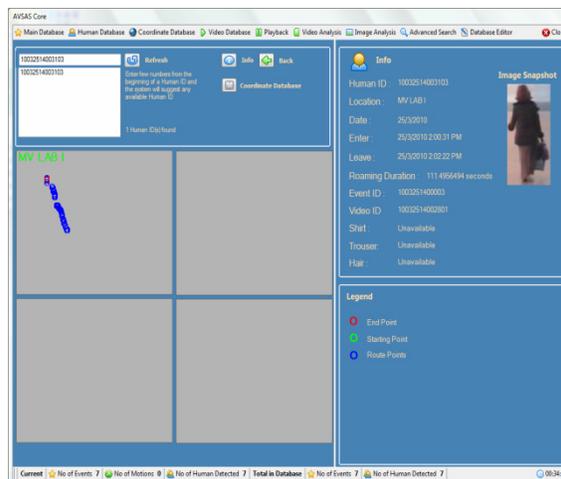


Figure 14: Track Analysis

## 5. RESULT AND DISCUSSION

Our developed software works very efficiently with the real time video surveillance system. The software is a great tool to classify and track the movement of any object under the camera's secured area. The software has details display mode like time, place, human or non-human, how many object, trajectory path of the moving object, video retrieval using *Data Indexing and Retrieval*, playing past videos from the hard drive and so on. The details of those features have been discussed in details in the system overview section. This software is based on the following key video analysis technologies:

- **Human and Object Detection:** This software can detect moving human and/or objects in a video sequence generated by a static camera. The detection techniques are invariant to changes in natural lighting, reasonable changes in the weather, distraction movements and camera shake. Several algorithms are available in this software including adaptive background subtraction with healing which assumes a stationary background and treats all changes in the scene as objects of interest and salient motion detection [10] which assumes that a scene will have many different types of motion, of which some types are of interest from a surveillance perspective. The result of human detection is illustrated in Figure 15
- **Human and Object Tracking:** This software can track the position of multiple objects as they move around a space that is monitored by a static camera as illustrated in Figure 16.
- **Object Classification:** This software uses various properties of an object especially human including shape, size and movement to assign an event and object type label to the objects. Our system fulfills the following criteria for *Advanced Search*.

Searching capability and smart indexing of data made this software handles data efficiently and ease the management of large video data. The searching itself comprises of combinations of logics as follows:

- Search by *Event Type* retrieves all event matches the requested type
- Search by *Time* retrieves all events that occurred during a specified time interval.
- Search by *Location* retrieves all objects within a specified area in a camera.

- Search by *Image* retrieves all images of human or objects that has appeared within the camera's secured area.
- Joint Search combines one or more of the above criteria as specified by the user



Figure 15: Result of human detection carried out by this system

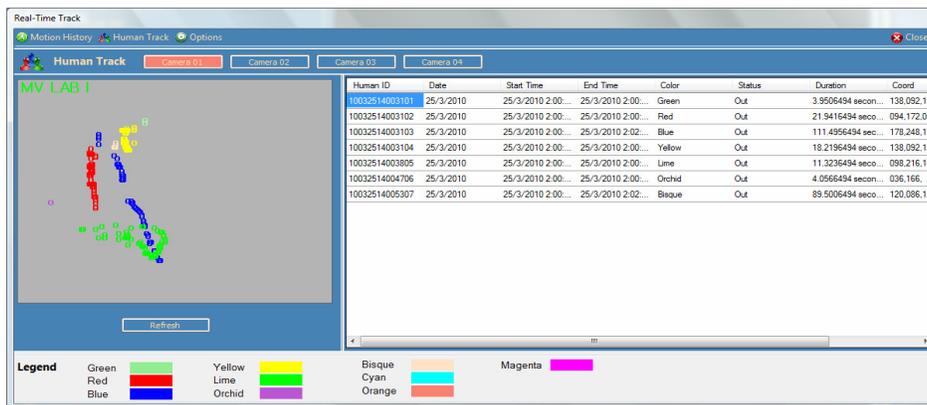


Figure 16: Multiple human tracking

## 6. CONCLUSION AND FUTURE WORKS

We have presented our smart video surveillance software for the surveillance purposes. Also we have introduced a generic smart video surveillance systems model, which relates the computer vision algorithms. All these methods have been linked with the smart surveillance system. From the practical point of view we found the developed software is more effective compared to the traditional surveillance system as well as it has a details display mode which helps us to track the moving object in an easier way and smart data handling for indexing of massive video data.

**Future Works:** The *event classification* and *Data Indexing and Retrieval* system would be improved further to enable wider search criteria to be implemented such as human/object specific appearance search and also to enhance the detection subsystem to be able to detect more objects. The system also will be improved to be able to classify human behavior from human body posture, movement speed and frequency of appearances. Moreover, the partial body occlusion is the main drawback from using shape-based human detection in this detection class as the system cannot determine correctly the boundary of human body in case of occlusion. But we will overcome this problem using head detection to correctly determine the number of human presents in the scene and therefore can locate the bounding boxes accurately based on human-shape model. The code for the software will also be revised to implement *multi-threading* for better performance with multi-core processors.

## 7. REFERENCES

1. W. Hu, T. Tab, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Trans. Syst. Man Cybern -- Part C: App. and Reviews*, vol. 34, no. 3, pp. 334–352, 2004.
2. D. M. Gavrilu, "The Visual Analysis of Human Movement: A survey," *Computer Vision and Image Understanding*, vol. 73, 1999, pp. 82-98.
3. C. S. Regazzoni, V. Ramesh, and G. L. Foresti, "Distributed Embedded Smart Cameras for Surveillance Applications", *Proceedings of the IEEE*, 39(2), Oct 2006.
4. L. Brown, "View Independent Vehicle/Person Classification", *VSSN'04*, October 15, 2004, New York, USA.
5. B. Bose and E. Grimson, "Improving object classification in far-field video", *CVPR*, 2004.
6. Case Study by Seagate, "Smart Surveillance Systems Tap Reliability and Performance of Seagate SV35 Series Drives." *CS509*, October 2006.
7. L. Chen et al, "An Integrated System for Moving Object Classification in Surveillance Videos" *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*.2008.
8. W. Wolf, B. Ozer, and T. Lv, "Smart Cameras as Embedded Systems", *IEEE Computer Science Journal* , 35(9):48–53, Sep 2002.
9. Fadhlan Hafiz et.al, "Human Detection For Real-Time Video Surveillance", In *Proceedings of International Conference on Engineering Technology 2010*, Kuala Lumpur, Malaysia, 2009
10. Tian, Y.I. Hampapur, "A Robust Salient Motion Detection with Complex Background for Real-Time Video Surveillance", In: *IEEE Computer Society Workshop on Motion and Video Computing*, Breckenridge, Colorado, January 5 and 6, 2005

## Script Identification of Text Words from a Tri Lingual Document Using Voting Technique

**M.C. Padma**

padmapes@gmail.com

*Dept. of Computer Science & Engineering,  
PES College of Engineering,  
Mandya-571401, Karnataka, India*

**P. A. Vijaya**

pavmkv@gmail.com

*Dept. of E. & C. Engineering,  
Malnad College of Engineering,  
Hassan-573201, Karnataka, India*

---

### Abstract

In a multi script environment, majority of the documents may contain text information printed in more than one script/language forms. For automatic processing of such documents through Optical Character Recognition (OCR), it is necessary to identify different script regions of the document. In this context, this paper proposes to develop a model to identify and separate text words of Kannada, Hindi and English scripts from a printed tri-lingual document. The proposed method is trained to learn thoroughly the distinct features of each script and uses the simple voting technique for classification. Experimentation conducted involved 1500 text words for learning and 1200 text words for testing. Extensive experimentation has been carried out on both manually created data set and scanned data set. The results are very encouraging and prove the efficacy of the proposed model. The average success rate is found to be 99% for manually created data set and 98.5% for data set constructed from scanned document images.

**Keywords:** Multi-lingual document processing, Script Identification, Feature Extraction, Binary Tree Classifier.

---

### 1. INTRODUCTION

Automatic script identification has been a challenging research problem in a multi script environment over the last few years. In recent years, the growing use of physical documents has made to progress towards the creation of electronic documents to facilitate easy communication and storage of documents. However, the usage of physical documents is still prevalent in most of the communications. For instance, the fax machine remains a very important means of communication worldwide. Also, the fact that paper is a very comfortable and secured medium to deal with, ensures that the demand for physical documents continues for many more years to come. So, there is a great demand for software, which automatically extracts, analyses and

stores information from physical documents for later retrieval. All these tasks fall under the general heading of document image analysis, which has been a fast growing area of research in recent years.

One important task of document image analysis is automatic reading of text information from the document image. The tool Optical Character Recognition (OCR) performs this, which is broadly defined as the process of reading the optically scanned text by the machine. Almost all existing works on OCR make an important implicit assumption that the script type of the document to be processed is known beforehand. In an automated multilingual environment, such document processing systems relying on OCR would clearly need human intervention to select the appropriate OCR package, which is certainly inefficient, undesirable and impractical. If a document has multilingual segments, then both analysis and recognition problems become more severely challenging, as it requires the identification of the languages before the analysis of the content could be made [10]. So, a pre-processor to the OCR system is necessary to identify the script type of the document, so that specific OCR tool can be selected. The ability to reliably identify the script type using the least amount of textual data is essential when dealing with document pages that contain text words of different scripts. An automatic script identification scheme is useful to (i) sort document images, (ii) to select specific Optical Character Recognition (OCR) systems and (iii) to search online archives of document image for those containing a particular script/language.

India is a multi-script multi-lingual country and hence most of the document including official ones, may contain text information printed in more than one script/language forms. For such multi script documents, it is necessary to pre-determine the language type of the document, before employing a particular OCR on them. With this context, in this paper, it is proposed to work on the prioritized requirements of a particular region- Karnataka, a state in India. According to the three-language policy adopted by most of the Indian states, the documents produced in Karnataka are composed of texts in Kannada- the regional language, Hindi – the National language and English. Such trilingual documents (documents having text in three languages) are found in majority of the private and Government sectors, railways, airlines, banks, post-offices of Karnataka state. For automatic processing of such tri-lingual documents through the respective OCRs, a pre-processor is necessary which could identify the language type of the texts words. In addition to these three scripts, the document may contain English numerals also to represent some numerical information. So, in this paper, it is proposed to develop a model to identify and separate text words of Kannada, Hindi and English scripts and English numerals. In this paper, the terms script and language could be interchangeably used as the three languages - Kannada, Hindi and English belong to three different scripts.

This paper is organized as follows. The Section 2 briefs about the previous work carried out in this area. The database constructed for testing the proposed model is presented in Section 3. Section 4 briefs about the necessary preprocessing steps. In Section 5, complete description of the proposed model is explained in detail. The details of the experiments conducted and the states of results obtained are presented in section 6. Conclusions are given in section 7.

## **2. LITERATURE SURVEY**

Automatic script identification is a challenging research problem in a multi script environment over the last few years. Major work on Indian script identification is by Pal, Choudhuri and their team [1, 3, 5]. Pal and Choudhuri [1] have proposed an automatic technique of separating the text lines from 12 Indian scripts (English, Devanagari, Bangla, Gujarati, Tamil, Kashmiri, Malayalam, Oriya, Punjabi, Telugu and Urdu) using ten triplets formed by grouping English and Devanagari with any one of the other scripts. This method works only when the triplet type of the document is known. Script identification technique explored by Pal [3] uses a binary tree classifier for 12 Indian scripts using a large set of features. The method suggested in [3] segments the input image up to character level for feature extraction and hence complexity increases. Lijun Zhou et. al. [9] has

developed a method for Bangla and English script identification based on the analysis of connected component profiles. Santanu Choudhuri, et al. [4] has proposed a method for identification of Indian languages by combining Gabor filter based technique and direction distance histogram classifier considering Hindi, English, Malayalam, Bengali, Telugu and Urdu. Gopal Datt Joshi, et. al. [6] have presented a script identification technique for 10 Indian scripts using a set of features extracted from log-Gabor filters. Ramachandra Manthalkar et.al. [19] have proposed a method based on rotation-invariant texture features using multichannel Gabor filter for identifying seven Indian languages namely Bengali, Kannada, Malayalam, Oriya, Telugu and Marathi. Hiremath et al. [20] have proposed a novel approach for script identification of South Indian scripts using wavelet based co-occurrence histogram features. Though global approaches are faster, they are applicable and well suited only when the whole document or a paragraph or a text line are in one and only one script. But, in majority of the documents one text line itself may contain texts in different languages. For such documents, it is necessary to identify the script type at word level.

Sufficient work has also been carried out on non-Indian languages [2, 17, 18]. Tan [2] has developed a rotation invariant texture feature extraction method for automatic script identification for six languages: Chinese, Greek, English, Russian, Persian and Malayalam. Lijun Zhou et. Al. [9] has developed a method for Bangla and English script identification based on the analysis of connected component profiles. Peake and Tan [10] have proposed a method for automatic script and language identification from document images using multiple channel (Gabor) filters and gray level co-occurrence matrices for seven languages: Chinese, English, Greek, Korean, Malayalam, Persian and Russian. Wood et al. [14] have proposed projection profile method to determine Roman, Russian, Arabic, Korean and Chinese characters. Hochberg et al. [15] have presented a method for automatically identifying script from a binary document image using cluster-based text symbol templates. Andrew Bhush [17] has presented a texture-based approach for automatic script identification. Spitz has [18] proposed method to discriminate between the Chinese based scripts and the Latin based scripts.

Some considerable amount of work has been carried out on specifically the three languages - Kannada, Hindi and English. Basavaraj Patil et. al. [7] have proposed a neural network based system for script identification of Kannada, Hindi and English languages. Vipin Gupta et. al. [13] have presented a novel approach to automatically identify Kannada, Hindi and English languages using a set of features- cavity analysis, end point analysis, corner point analysis, line based analysis and Kannada base character analysis. Word level script identification in bilingual documents through discriminating features has been developed by Dhandra et. al. [8]. Padma et. al. [11] have presented a method based on visual discriminating features for identification of Kannada, Hindi and English text lines. Though a great amount of work has been carried out on identification of the three languages Kannada, Hindi and English, very few works are reported in literature at word level. Also, the great demand for automatic processing of tri-lingual documents shows that much more work needs to be carried out on word level identification. So, this paper focuses on word wise identification of Kannada, Hindi and English scripts.

### **3. DATA COLLECTION**

Standard database of documents of Indian languages is currently not available. In this paper, it is assumed that the input data set contains text words of Kannada, Hindi and English scripts and English numerals. For the experimentation of the proposed model, three sets of database were constructed, out of which one database was used for learning and the other two databases were constructed to test the system. The text words of Kannada and English scripts, and English numerals were created using the Microsoft word software. These text words were imported to the Micro Soft Paint program and saved as black and white bitmap (BMP) images. The font type of Times New Roman, Arial, Bookman Old Style and Tahoma were used for English language. The font type of Vijaya, Kasturi and Sirigannada were used for Kannada language. The font size of 14, 20 and 26 were used for both Kannada and English text words. However, the performance is

independent of font size. The text words of Hindi language were constructed by clipping only the text portion of the document downloaded from the Internet. So, the data set constructed using Microsoft word software and by clipping the text portion from the downloaded documents is called manually created data set. Thus the data set of 500 text words from each of the four classes (Kannada, Hindi, English and English Numerals) was constructed to train the proposed system.

To test the proposed model, two different data sets were constructed. One dataset of size 300 text words was constructed manually similar to the data set constructed for learning and the other data set was constructed from the scanned document images. The printed documents like newspapers and magazines were scanned through an optical scanner to obtain the document image. The scanner used in this research work for obtaining the digitized images is HP Scan Jet 5200c series. The scanning is performed in normal 100% view size at 300 dpi resolution. The test document image of size 600x600 pixels were considered such that each text line would contain text words in mixture of the three languages. Manually constructed dataset is considered as good quality dataset and the data set constructed from the scanned document images are considered poor quality data set. The test data set was constructed such that 200 text words were incorporated from each of the three scripts - Kannada, Hindi and English, and English numerals.

#### **4. PREPROCESSING**

Any script identification method used for identifying the script type of a document, requires conditioned image input of the document, which implies that the document should be noise free, skew free and so on. In this paper, the preprocessing techniques such as noise removal and skew correction are not necessary for the manually constructed data sets. However, for the datasets that were constructed from the scanned document images, preprocessing steps such as removal of non-text regions, skew-correction, noise removal and binarization is necessary. In the proposed model, text portion of the document image was separated from the non-text region manually. Skew detection and correction was performed using the existing technique proposed by Shivakumar [16]. Binarization can be described as the process of converting a gray-scale image into one, which contains only two distinct tones, that is black and white. In this work, a global thresholding approach is used to binarize the scanned gray scale images where black pixels having the value 0's correspond to object and white pixels having value 1's correspond to background.

The document image is segmented into several text lines using the valleys of the horizontal projection profiles computed by a row-wise sum of black pixels. The position between two consecutive horizontal projections where the histogram height is least denotes the boundary of a text line. Using these boundary lines, document image is segmented into several text lines. Each text line is further segmented into several text words using the valleys of the vertical projection profile computed by a column-wise sum of black pixels. From the experimentation, it is observed that the distance between two words is greater than two times the distance between two characters in a word. So, the threshold value for inter word gap is decided as two times the inter character gap. Using this inter word gap, each text line is segmented into several text words. Then, a bounding box is fixed for the segmented text word by finding the leftmost, rightmost, topmost and bottommost black pixels. Thus, the image of the bounded text word is prepared ready for further processing such as feature extraction.

#### **5. OVERVIEW OF THE PROPOSED MODEL**

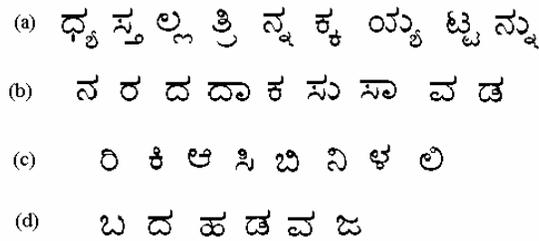
The proposed model is inspired by a simple observation that every script/language defines a finite set of text patterns, each having distinct visual discriminating features. Hence, the new model is designed by using the distinct features of the scripts under consideration. Scripts are made up of different shaped patterns to produce different character sets. Individual text patterns of one script are collected together to form meaningful text information in the form of a text word, a text line or a paragraph. The collection of the text patterns of the one script exhibits distinct visual

appearance and hence it is necessary to thoroughly study the discriminating features of each script that are strong enough to distinguish from other scripts. To arrive at the distinct features of each script under consideration, the complete character set of those scripts should be thoroughly studied. Sometimes, a text word may even contain only two characters thus making the feature extraction process too complex. As a result, a large number of features have to be considered to develop word level script identification model. The properties of the three scripts - Kannada, Hindi and English, and English Numerals are described below.

#### a. Some Discriminating Features of Kannada Script

Modern Kannada character set has 47 basic characters, out of which the first 13 are vowels and the remaining 34 characters are consonants [10]. Some books report 14 vowels and 36 consonants. By adding vowels to each consonant, modified consonants are obtained. A consonant or a modified consonant is combined with another consonant to form a compound character. As a result, Kannada text words consists of combination of vowels, consonants, modified consonant and/or compound characters. The compound characters have descendants called 'vathaksharas' found at their bottom portions. Some examples of Kannada compound characters with descendants are given in Figure 1. The presence of these descendants is one of the discriminating features of Kannada script, which is not present in the other two scripts - Hindi and English and hence, it could be used as a feature named bottom-component to identify the text word as a Kannada script.

It could be observed that most of the Kannada characters have either horizontal lines or hole-like structures present at the top portion of the characters. Also, it could be observed that majority of Kannada characters have upward curves present at their bottom portion. Some characters have double-upward curves found at their bottom portions. In addition, left curve and right curve are also present at the left and right portion of some characters. Thus, the presence of the structures such as – horizontal lines, hole-like structures, bottom-up-curves, descendants, left-curves and right-curves could be used as the supporting features to identify Kannada scripts. Some examples of Kannada characters with the above said features are given in Figure 1. The probability of presence of these features is thoroughly studied from a large collection of documents. The density of the occurrence of these features is thoroughly studied and the features with maximum density are considered in the proposed model.



**FIGURE 1.** Some characters of Kannada script (a) Characters with descendants, (b) Characters with horizontal lines (c) Characters with holes at the top portion and (d) Characters with double-upward curves

#### b. Some Discriminating Features of Hindi Script

It could be noted that many characters of Hindi script have a horizontal line at the upper part called sirorekha [1], which is generally called a headline. It could be seen that, when two or more characters are combined to form a word, the character headline segments mostly join one another and generates one long headline at the top portion of each text word. These long horizontal lines are present at the top portion of the characters. The presence of such horizontal lines is used as supporting features for identifying Hindi script. Another strong feature that could

be noticed in Hindi script is the presence of vertical lines. Some examples of Hindi text words are given in Figure 2.

सिलसिला बना इसी तरह दुर्ग

FIGURE 2. Some text words of Hindi script

### c. Some Discriminating Features of English Script

English character set has 26 alphabets in both upper and lower cases. One of the most distinct and inherent characteristics of most of the English characters is the existence of vertical line-like structures. It could be observed that the upward-curve and downward-curve shaped structures are present at the bottom and top portion of majority of English characters respectively. So, it was inspired to use these distinct characteristics as supporting features in the proposed script identification model.

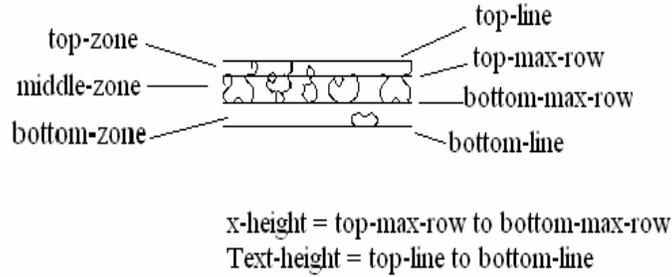
### d. Properties of English Numerals

It could be observed that majority of the documents contain numerical information in addition to the text information. This numeral information is represented using the English numeral symbols '0' to '9' only, even when the document is printed in the scripts prevailed in Karnataka. So, it is necessary to identify and separate these English numerals from the text written in the three scripts. One important feature of numerals is that the height of all the numeral symbols is the same. Another feature of the English numerals is that the width of all the symbols (except the symbol '1') is the same. These features could be considered as supporting features in classifying the test word as an English numeral.

As the text words of Hindi scripts are visually more distinct when compared to the visual appearance of English and Kannada text words and English numerals, one or two features are enough to separate Hindi text words from others. So, a test sample is initially checked whether it is a Hindi text word or not. If it not a Hindi text word, then classifying the test sample into any of the remaining three classes English, Kannada and English numerals is a complex task, as some of the text words almost possess similar features. So, arriving at the features that is strong enough to distinguish among the three classes requires detailed study of the characters of those three classes.

### e. Text word Partitioning

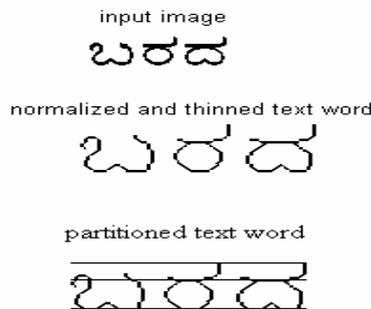
By thoroughly observing the structural outline of the characters of the three scripts, it is observed that the distinct features are present at some specific portion of the characters. So, in this paper, the discriminating features are well projected by partitioning the text line using the four lines that are obtained from the top-profile and the bottom-profile of each text line. The top-profile (bottom-profile) of a text line represents a set of black pixels obtained by scanning each column of the text line from top (bottom) until it reaches a first black pixel. Thus, a component of width N gets N such pixels. The row at which the first black pixel lies in the top-profile (bottom-profile) is called top-line (bottom-line). The row number having the maximum number of black pixels in the top-profile (bottom-profile) is called the attribute top-max-row (bottom-max-row). Using these four lines – top-line, bottom-line, top-max-row and bottom-max-row as the reference lines, the features are extracted from each text line of the respective script. A sample partitioned Kannada text word is shown in Figure 3. The attribute 'x-height' represents the difference between top-max-row and bottom-max-row and the attribute 'text-height' represents the difference between top-line and bottom-line.



**FIGURE 3.** Partitioned Kannada Text Word.

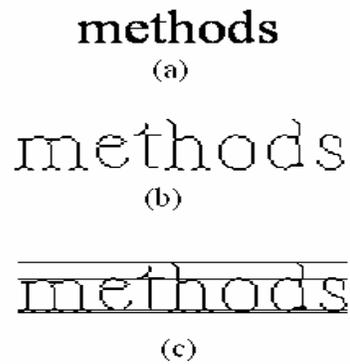
Different text words are partitioned in different ways, as different shaped characters are present in text words. A sample Kannada text word that is partitioned into three zones namely top-zone, middle-zone and bottom-zone is shown in Figure 3. A partition with at least three pixels height (fixed through experimentation) is considered as a top-zone or bottom-zone. A text word can be partitioned into three zones only when the four reference lines namely top-line, top-max-row, bottom-max-row and bottom-line are obtained. However, for some text words where top-line and top-max-row occur at the same location, top-zone is not obtained. Similarly, for some text words if bottom-max-row and bottom-line occur at the same location, then bottom-zone is not obtained. This is because of the absence of ascendants or descendants. Ascendants are the portion of the characters that are protruded above the top-max-row and descendants are the portions of the characters that are protruded below the bottom-max-row. In Kannada script the presence of 'vothaksharas' could be considered as a descendant. So, for a Kannada text word with descendant, the two reference lines – bottom-max-row and bottom-line are present and the space between the bottom-max-row and bottom-line could be called as the bottom-zone. The partitioning of the typical Kannada text word with the descendant is shown in Figure 3.

However, if the text word without the descendant is partitioned, then the bottom-zone is not obtained, since the bottom-max-row and the bottom-line occur at the same row. Such a text word in Kannada script without the descendant is shown in Figure 4. Similarly, a text word without ascendants does not possess top-zone.

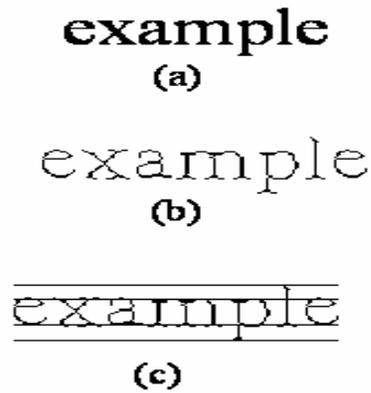


**FIGURE 4.** Kannada Text Word Without bottom-zone.

English text words are partitioned in a similar way as that of Kannada text word partitioning. This is because, some English characters like 'b, d, f, h, k, l and t' have ascendants and some characters like 'g, j, p, q and y' have descendants. So, if the characters of the text word have ascendants, then the top-zone is obtained and if the characters of the text word have descendants then the bottom-zone is obtained. For some other characters like 'a, c, e, m, n, o, r, s, u, v, w, x, z', there are no ascendants and descendants. For the text words having these characters, top-zone and also bottom-zone are not obtained. So, only middle zone is obtained for such text words. Different partitioned English text words are shown in Figures 5 and 6.

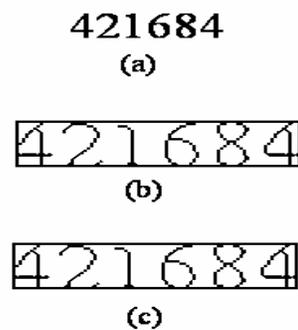


**FIGURE 5.** Partitioned English Text Word Without Descendant (a) Input Word (b) Preprocessed Word (c) Partitioned Word without bottom-zone.



**FIGURE 6.** Partitioned English Text Word With Descendant (a) Input Word (b) Preprocessed Word (c) Partitioned Word with top-zone and bottom-zone

It is observed that all English numerals are equal in height. So, the partitioning of a text word containing only English numerals results in middle-zone only. A sample image of English numeral and its partitioned image are shown in Figure 7.



**FIGURE 7.** Partitioned English Numeral (a) Input Word (b) Preprocessed Word with Bounding Box (c) Partitioned Word without top-zone and bottom-zone

#### f. Feature Extraction

The distinct features useful for identifying the three scripts – Kannada, Hindi and English are shown in Table 1. The entry 'Y' in the Table 1 means that the feature in the corresponding row is used for identifying the script in the corresponding column. Thus, seven features for Kannada, three features for Hindi and three features for English are used. It is observed in Table 1 that the features used for identifying English numerals are not listed. The method of identifying English numerals is explained in the later Section.

**TABLE 1.** Features of Kannada, Hindi and English languages.

	<b>Features</b>	<b>Kannada</b>	<b>Hindi</b>	<b>English</b>
F1	Bottom-components	Y	--	--
F2	Bottom-max-row-no	--	Y	--
F3	Top-horizontal-line	Y	Y	--
F4	Vertical-lines	--	Y	Y
F5	Top-holes	Y	--	--
F6	Top-down-curves	--	--	Y
F7	Bottom-up-curves	Y	--	Y
F8	Bottom-holes	Y	--	--
F9	Left-curve	Y	--	--
F10	Right-curve	Y	--	--

The method of extracting the distinct features, which are used in the proposed model, is explained below:

##### **Feature 1: Bottom-component**

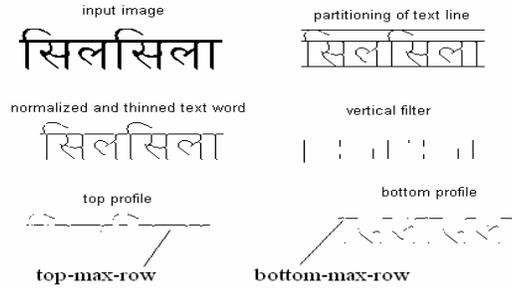
The presence of vathaksharas or descendants found at the bottom portion of Kannada script could be used as a feature called bottom-component. The feature named 'bottom-component' is extracted from the bottom-portion of the input text line. Bottom-portion is computed as follows:

Bottom-portion =  $f(x,y)$  where  $x$ =bottom-max-row to  $m$  and  $y=1$  to  $n$  ; where  $f(x,y)$  represent the matrix of the preprocessed input image of size  $(m \times n)$ .

Through experimentation, it is estimated that the number of pixels of a descendant is greater than 8 pixels and hence the threshold value for a connected component is fixed as 8 pixels. Any connected component whose number of pixels is greater than 8 pixels is considered as the feature bottom-component. Such bottom-components extracted from Kannada script are shown in Figure 11.

##### **Feature 2: Bottom-max-row-no**

It is observed through experimentation that for Kannada and English script, the two attributes top-max-row and bottom-max-row occur at a distance of  $x$ -height as shown in the partitioned Kannada text word shown in Figure 3 and partitioned English text words shown in Figure 5 and 6. But, for a partitioned Hindi text word, the  $x$ -height is 0 or 1 pixel, as the top-max-row and bottom-max-row occur at the same location. This is because when the bottom profile of Hindi script is computed the pixels of the headline happen to be the pixels of bottom profile. So, the top-max-row and bottom-max-row occur at the same location in a Hindi text word. This is the distinct property of Hindi script that is not so in the other two scripts English and Kannada. Hence, if the bottom-max-row is equal to the top-max-row, then the value of the attribute bottom-max-row could be used as a strong feature named 'bottom-max-row-no' to separate Hindi script from the other two scripts. A typical Hindi text word with the feature 'bottom-max-row-no' is shown in Figure 8.



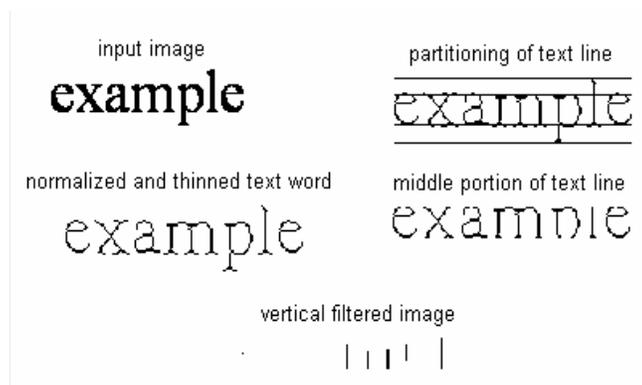
**FIGURE 8.** Hindi text word with bottom-max-row and vertical lines

**Feature 3: Top-horizontal-line**

It could be noted that the horizontal line like structures are present at the top-max-row of Kannada and Hindi scripts. The connected components present at the top-max-row of the text word are analyzed. If the number of pixels of these connected components is greater than the 75% of the x-height, then such components are used as the feature top-horizontal-line. The probability of presence of this feature is calculated from the complete Kannada character set. Also, the distribution of this feature is analyzed using 500 text words from all the three languages Kannada, Hindi and English. From the experimental analysis, it is observed that the presence of top-horizontal-line is more in Kannada and Hindi script and it is almost absent in the case of English script. So, using the feature named top-horizontal-line, Kannada and Hindi scripts could be separated from English script. If the length of the horizontal line (length of the horizontal line is measured with the number of pixels of that component) is greater than two times the x-height, then Hindi text word can be separated from Kannada word. So, using the length of the feature top-horizontal-line, Hindi can be well separated from Kannada script. The feature top-horizontal-line is shown in the output images of Hindi and Kannada text words in Figures 8 and 11 respectively.

**Feature 4: Vertical lines**

It is noticed that the Hindi and English scripts have vertical line segments. To extract these vertical lines, the middle-zone of the text line is extracted as below:  
 Middle-zone =  $g(x,y)$  where  $x = \text{top-max-row to bottom-max-row}$  and  $y = 1 \text{ to } n$   
 where  $g(x,y)$  is the input matrix size  $(m,n)$ . By convolving a vertical line filter over the image of the middle-zone, vertical lines are extracted. Typical vertical lines extracted from English script are shown in Figure 9. The presence of these vertical lines more in Hindi and English script, whereas it is absent in Kannada script. Hence, the feature vertical lines are used to identify Hindi and English script.



**FIGURE 9.** English text word with vertical lines

**Feature 5: Top-holes**

Hole is a connected component having a set of white pixels enclosed by a set of black pixels (black pixels having the value 0's correspond to object and white pixels having value 1's correspond to background). By thoroughly observing the words of Kannada scripts, it is noticed that hole-like structures are found at the top portion. To compute the hole-like structures, the attribute top-pipe is obtained from the matrix of the pre-processed image as follows:

Top-pipe  $(x1, y1) = f(x, y)$  where  $x = \text{top-max-row} - t$  to  $\text{top-max-row} + t$  and  $y = 1$  to  $n$  where  $f(x,y)$  and  $n$  represents the input image and number of columns of the input image. The variable 't' is used as a threshold value and  $t = \text{round}(x\text{-height}/3)$ , where the term 'x-height' represents the difference between top-max-row and bottom-max-row.

Presence of holes at the top-pipe is used as the feature top-holes and it is used to identify the text word as Kannada script as this feature is not present in the other two anticipated languages. Hole-like structures extracted from the sample Kannada script text word is shown in Figure 11.

**Features 6 & 7: Top-down-curves and Bottom-up-curves**

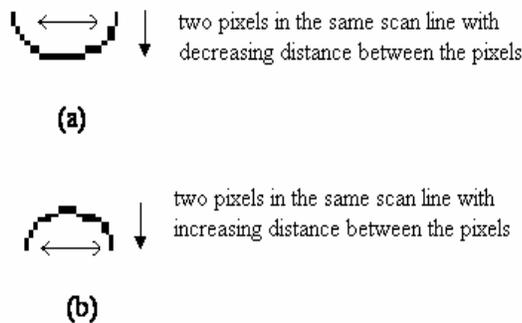
By thoroughly observing the structural shape of the two scripts – Kannada and English, it is observed that the upward and downward shaped components are present at the region of top-max-row and bottom-max-row. This inspired us to extract the two attributes top-pipe and bottom-pipe as follows:

Top-pipe =  $g(x,y)$  where  $x = \text{top-max-row} - t$  to  $\text{top-max-row} + t$  and  $y = 1$  to  $n$  and

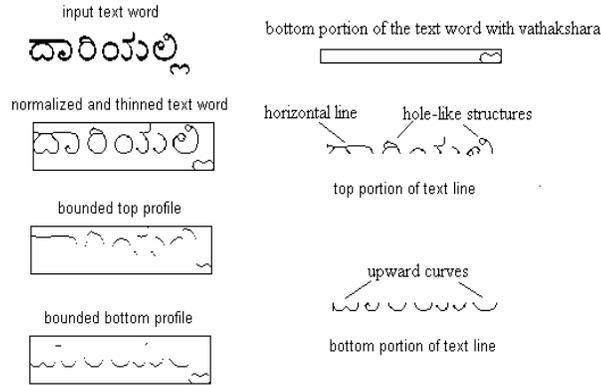
Bottom-pipe =  $g(x,y)$  where  $x = \text{bottom-max-row} - t$  to  $\text{bottom-max-row} + t$  and  $y = 1$  to  $n$

where  $g(x,y)$  and  $n$  represents the input image and number of columns of the input image. The variable 't' is used as a threshold value and  $t = \text{round}(x\text{-height}/3)$ , where the term 'x-height' represents the difference between top-max-row and bottom-max-row..

Detecting the curve shaped portion from a character is the key for extracting the features named top-down-curves and bottom-up-curves. The presence of a curve is obtained by verifying the variation between two pixels of a connected component that appear on the same scan line for the complete scan of the component. The increasing variations of the two pixels for the entire scan of the component results in top-down-curves and decreasing variations of the two pixels for the entire scan of the component results in bottom-down-curves. Components having the shape upward curve and downward curve are shown in Figure 10.



**FIGURE 10.** (a) upward curves (b) downward curves



**FIGURE11.** Output image of Kannada text word

**Feature 8: Bottom-holes**

By thoroughly observing the text words of Kannada scripts, it is observed that some characters have hole-like structures at their bottom portion. To compute the hole-like structures, the attribute bottom-pipe is obtained from the matrix of the pre-processed image as follows:

Bottom-pipe  $(x1, y1) = f(x, y)$  where  $x = \text{bottom-max-row} - t$  to  $\text{bottom-max-row} + t$  and  $y = 1$  to  $n$  where  $f(x,y)$  and  $n$  represents the input image and number of columns of the input image. The variable 't' is used as a threshold value and  $t = \text{round}(x\text{-height}/3)$ , where the term 'x-height' represents the difference between top-max-row and bottom-max-row..

Presence of holes at the bottom-pipe is used as the feature bottom-holes and it is used to identify the text word as Kannada script as this feature is not present in the other two anticipated scripts.

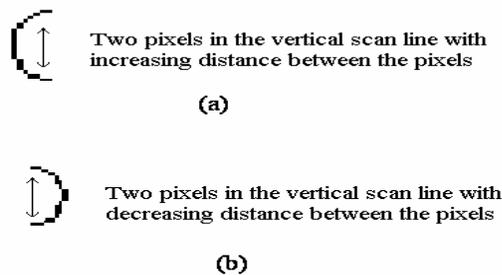
**Feature 9 and 10: Left-curve and Right-curve**

The structural shape of the two scripts – Kannada and English is observed thoroughly and noticed that the left-curve and right-curve shaped components are present at the middle-zone of a partitioned text word. This inspired us to extract the middle-zone as follows:

Middle-zone =  $g(x, y)$

where  $x = (\text{top-max-row} + t)$  to  $(\text{bottom-max-row} - t)$  and  $y = 1$  to  $n$ ,  $g(x, y)$  represents the input image and 'n' is the number of columns of the input image. The variable 't' is used as a threshold value determined through experimentation and  $t = 2$ .

Detecting the curve shaped portion from a character is the key for extracting the features named top-down-curves and bottom-up-curves. The presence of a left curve is obtained by verifying the distance between two pixels of a connected component that appear on the same vertical scan line of a component. The increasing variations of the two pixels for the entire vertical scan of the component results in the left-curve and decreasing variations of the two pixels for the entire vertical scan of the component results in right-curve. Components having the shape left-curve and right-curve are shown in Figure 12.



**FIGURE 12.** (a) Left-curve (b) Right-curve

#### **g. Simple voting technique for classification**

A simple voting technique is an approach used for classifying a test sample based on the maximum number of votes obtained for a class type. In this chapter, a simple voting technique is used to classify the given test sample into any of the four classes- Kannada, English, Hindi and English Numerals. The overview of the working principle of the simple voting technique is explained below:

Here, a simple voting technique is used for classifying the test word. For each class type, a particular accumulator, which is initialized to zero, is used to count the number of features present in a test word. For the present problem, there are four script classes – Kannada, English, Hindi and English Numerals. So, four accumulators namely KA, HA, EA and NA are used for the four script classes - Kannada, Hindi, English and English Numerals respectively. In this method, the test word is segmented into several blocks through vertical projection profiles. A block may contain one or more than one character with one or more connected components. So, in this method, the term character or block may be used interchangeably. After segmenting the text word, the number of blocks present in a test word is calculated. Each segmented block is tested for the presence of the features given in Table 1. For example, if the feature F3 is present in a block, then the accumulator KA is incremented, as only Kannada characters possess this feature. Then the presence of the feature F4 is tested and if this feature is present only English accumulator is incremented. This process is repeated for all the blocks of the given test sample. At the end of this procedure, all the four accumulators are stored with some value. The highest value is found in one of the accumulator and the label of highest value shows the class label of the test sample.

The features are extracted from the partitioned image in the order given in Table 1. Later, a simple voting technique is used to classify the test word into the respective script type. The voting technique used in the word level script identification model works in three stages as explained below:

##### **Stage 1:**

If the feature 'bottom-component' is present in the test image, then the test sample is classified as of type Kannada script, as the 'bottom-component' is the only distinct feature present among the four classes. If the feature 'bottom-component' is not present in the test image, then the presence of the next distinct feature 'bottom-max-row-no' is tested. If it is present, then the test word is classified as Hindi script. If the two features – 'bottom-component' and 'bottom-max-row-no' is not present in the test image, then it is not possible to classify the test word at this stage. So, those test words that are not classified in this stage are given to the next stages, which use the concept of simple voting technique. Thus, a simple voting technique is used only if the test image does not possess the first two features Bottom-component and bottom-max-row-no.

From the partitioned image, the value of the top-line and top-max-row are compared. Similarly, the value of the bottom-line and bottom-max-row are compared. For some characters that do not possess any ascendants and descendants, the top-line and top-max-row may occur at the same position and similarly the bottom-line and bottom-max-row may occur at the same position. Thus, a partitioned image may contain only one, two or all the three zones. If the test word is partitioned into only one zone, it means that test word contains the characters that do not possess any ascendants and descendants of Kannada and English scripts. So, the test image may be an English numeral or a Kannada word or an English word without ascendants and descendants. The test words that have two or three zones are fed to the stage 2 and those test words that have only one zone are fed to the stage 3 for further processing.

##### **Stage 2:**

The partitioned text words that have two or all the three zones are given as input to this stage. The given test word is segmented into several blocks/characters using the vertical projection profiles as presented in our earlier paper [11]. The test words that are given to this stage could be of type either Kannada or English or Hindi script type and definitely not of type English Numerals. Because, partitioned English numerals have only one zone i.e., middle-zone and they

do not have top-zone and bottom-zone. Once the test word is decided as not of English numerals, the features from F3 to F10 given in Table 1 are extracted from the characters of the segmented text word. Based on the presence of features in the characters of the text word, the corresponding accumulator gets incremented. Finally, the test words are classified as the script type of the accumulator that has maximum value.

### Stage 3:

The partitioned text word that has only one zone is given to the stage 3. So, the test word may be an English numeral, or a Kannada or an English text word that does not possess ascendants and descendants. Classifying the text words that come under these conditions is quite complex. The characters of Kannada text word that has only one zone generally have hole like structures or downward curves at the top portion of the characters and upward curves or double upward curves found at the bottom portion of the characters. If a character has any of these features, then the accumulator assigned for Kannada is incremented.

Similarly, if there are vertical lines like structures of length x-height found in the characters of test word, the accumulator assigned for English is incremented. Thus, all the features F3 through F10 are tested for each of the characters of the test word. For n number of characters in the test word, if the content of the accumulator with maximum value less than  $n/2$ , then, it means that the test word is not classified into either Kannada or English script type. Then the next step is to compute the width of each character. If the width of 90% of the characters is equal to one another, then that test word is classified as English Numeral.

### Algorithm Stage 1:

**Input: Document Image containing text words of Kannada, Hindi, English and English Numerals.**

**Output: Script type of each text word.**

1. Preprocess the input document image.
2. Segment the document image into several text lines.
3. Repeat for each text line
4. { Segment the text line into words
5. Repeat for each text word
6. { Partition the text word
7. If (bottom-component)  
{ Then Classify the text word as "Kannada" script and Return}
8. If (Bottom-max-row)  
{ Then Classify the text word as "Hindi" script and Return.}
9. If (only one zone) then call stage 3 Else call stage 2. } }

### Algorithm Stage 2:

1. Segment the text word into characters
2. Initialize the four accumulators -KA, HA, EA, NA to zero.
3. Repeat for each character
4. { Repeat for each feature F3 through F10  
{ If (feature present)  
Then increment the corresponding accumulator } }
5. Find the accumulator with maximum value.
6. Classify the script type of the text word as the corresponding accumulator.
7. Return

### Algorithm Stage 3:

1. Segment the text word into characters/blocks
2. Initialize the four accumulators -KA, HA, EA, NA to zero.
3. Repeat for each character
4. { Repeat for each feature F3 through F10  
{ If (feature present)

- ```
        Then increment the corresponding accumulator } }
5. Find the accumulator content with maximum value.
6. If (accumulator value >= n/2)
    { Then Classify the script type of the text word as the corresponding accumulator.
      Else If (width of 90% of the characters is same)
        Then Classify the script type as English Numerals
        Else Reject
7. Return
```

## 6. RESULTS AND DISCUSSION

The proposed algorithm has been tested on a test data set of 300 document images containing about 500 text words from each script. The test data set is constructed such that the English text words contain characters that possess ascendants (for example b, d, f, h, k, l, t) and descendants (for example g, j, p, q, y). The English text word without any ascendants and descendants (for example words like 'cow', 'man', 'scanner') are also considered in the test data set. The performance of classification is encouraging when tested with all kinds of words having the characters with and without ascendants and descendants. Similarly, the test data set of Kannada and Hindi scripts were constructed such that all characters of the two scripts are included in the test words. The algorithm is tested for text words containing two to six characters. The success rate is sustained even for the text words with only two characters. This is because all the features present in one or the other character are used in the proposed model. The test data set consisted of English Numerals also. English Numerals with two to eight digits were included with all the combinations of the nine symbols. Satisfactory success rate was achieved even in classifying the English Numerals. The failure in classifying the English numerals occurs only when the test word contains the number '1', as the width of this symbol is smaller and the width of all the remaining symbols is same.

The proposed algorithm has been implemented using Matlab R2007b. The average time taken to identify the script type of the text word is 0.1846 seconds on a Pentium-IV with 1024 MB RAM based machine running at 1.60 GHz. A sample manually constructed test document containing text words of all the four classes- Kannada, Hindi, English and English numerals are given in Figure 13.

multilingual    हर्षोल्लास  
2458 विमल    increasingly  
Phno 3403896    different  
ಕನ್ನಡದ    lettering  
विमल before    गंगोत्री  
presenting    ಮೈಸೂರು

**FIGURE 13.** Manually created test image containing Kannada, Hindi, English and Numeral words.

The algorithm is tested for various font types and the results are given in Table 2. The proposed method is independent of font type and size. Since the features are considered to be at specific region of the partitioned text word, the variation in the font size does not affect the performance of

the algorithm. The success rate of percentage of recognition of all the three scripts is given in Table 3. From the experimentations on the test data set, the overall accuracy of the system has turned out to be 98.8%. From the Table 3, it could be observed that the 100% accuracy is obtained for Hindi text word. This is because of the distinct feature of Hindi script. The performance of the proposed algorithm falls down for English text words printed in italics. This is one limitation. However, for the Kannada text words printed in italics, the performance is sustained. The performance of the proposed model was evaluated from the scanned document images also. The overall accuracy of the system reduces to 98.5% due to noise and skew-error in the scanned document images. However, if the scanned document images undergo suitable preprocessing techniques, the performance can be improved.

**TABLE 2.** Percentage of Recognition on manually created data set for different font styles

| Script type      | Font Style                    | Number of samples | Correct recognition | Recognition rate |
|------------------|-------------------------------|-------------------|---------------------|------------------|
| Kannada          | Sirigannada                   | 160               | 157                 | 98.13%           |
|                  | Kasturi                       | 170               | 167                 | 98.23%           |
|                  | Vijaya                        | 170               | 168                 | 98.82%           |
| Hindi            | Vijaya                        | 500               | 500                 | 100%             |
| English          | Times New Roman               | 100               | 97                  | 97%              |
|                  | Arial                         | 100               | 98                  | 98%              |
|                  | Tahoma                        | 100               | 98                  | 98%              |
|                  | Bookman Old Style             | 100               | 99                  | 99%              |
|                  | Verdana                       | 100               | 98                  | 98%              |
|                  | English Text: Upper Case only | 50                | 50                  | 100%             |
| English Numerals | Times New Roman               | 100               | 97                  | 96%              |
|                  | Arial                         | 100               | 96                  | 95%              |
|                  | Tahoma                        | 100               | 97                  | 96%              |
|                  | Bookman Old Style             | 100               | 97                  | 96%              |
|                  | Verdana                       | 100               | 96                  | 95%              |

**TABLE 3.** Percentage of Classification of the four classes

|                  | Dataset 1 (manually created data set) |               |          | Dataset 2 (scanned data set) |               |          |
|------------------|---------------------------------------|---------------|----------|------------------------------|---------------|----------|
|                  | Classified                            | Misclassified | Rejected | Classified                   | Misclassified | Rejected |
| Kannada          | 98.4%                                 | 0.8%          | 1.8%     | 97.8%                        | 0.8%          | 1.4%     |
| Hindi            | 100%                                  | 0%            | 0%       | 100%                         | 0%            | 0%       |
| English          | 98.3%                                 | 0.5%          | 1.2%     | 97.6%                        | 0.8%          | 1.6%     |
| English Numerals | 95.6%                                 | 3.2%          | 1.8%     | 95.8%                        | 2.7%          | 2.5%     |

## 7. CONCLUSION

In this paper, a new method to identify and separate text words of the Kannada, Hindi and English scripts and also English numerals is presented. Experimental results show performance of the proposed model. The performance of the proposed algorithm is encouraging when the proposed algorithm is tested using manually created data set. However, the performance slightly comes down when the algorithm is tested on scanned document images due to noise and skew-error. Our future work is to identify the numeral information printed in the Kannada and Hindi

scripts and also to reach higher rate of success (100%). Further, it is planned to identify the scripts from a degraded document images.

## References

- [1] U.Pal, B.B.Choudhuri, "*Script Line Separation From Indian Multi-Script Documents*", 5th Int. Conference on Document Analysis and Recognition (IEEE Comput. Soc. Press), 406-409, (1999).
- [2] T.N.Tan, "*Rotation Invariant Texture Features and their use in Automatic Script Identification*", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 7, pp. 751-756, (1998).
- [3] U. Pal, S. Sinha and B. B. Chaudhuri, "*Multi-Script Line identification from Indian Documents*", In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003) 0-7695-1960-1/03 © 2003 IEEE, vol.2, pp.880-884, (2003).
- [4] Santanu Choudhury, Gaurav Harit, Shekar Madnani, R.B. Shet, "*Identification of Scripts of Indian Languages by Combining Trainable Classifiers*", ICVGIP, Dec.20-22, Bangalore, India, (2000).
- [5] S. Chaudhury, R. Sheth, "*Trainable script identification strategies for Indian languages*", In Proc. 5th Int. Conf. on Document Analysis and Recognition (IEEE Comput. Soc. Press), pp. 657-660, 1999.
- [6] Gopal Datt Joshi, Saurabh Garg and Jayanthi Sivaswamy, "*Script Identification from Indian Documents*", LNCS 3872, pp. 255-267, DAS (2006).
- [7] S.Basavaraj Patil and N V Subbareddy, "*Neural network based system for script identification in Indian documents*", Sadhana Vol. 27, Part 1, pp. 83-97. © Printed in India, (2002).
- [8] B.V. Dhandra, Mallikarjun Hangarge, Ravindra Hegadi and V.S. Malemath, "*Word Level Script Identification in Bilingual Documents through Discriminating Features*", IEEE - ICSCN 2007, MIT Campus, Anna University, Chennai, India. pp.630-635. (2007).
- [9] Lijun Zhou, Yue Lu and Chew Lim Tan, "*Bangla/English Script Identification Based on Analysis of Connected Component Profiles*", in proc. 7th DAS, pp. 243-254, (2006).
- [10] G. S. Peake and T. N. Tan, "*Script and Language Identification from Document Images*", Proc. Workshop Document Image Analysis, vol. 1, pp. 10-17, 1997.
- [11] M. C. Padma and P.Nagabhushan, "*Identification and separation of text words of Karnataka, Hindi and English languages through discriminating features*", in proc. of Second National Conference on Document Analysis and Recognition, Karnataka, India, pp. 252-260, (2003).
- [12] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins, "*Digital Image Processing using MATLAB*", Pearson Education, (2004).
- [13] Vipin Gupta, G.N. Rathna, K.R. Ramakrishnan, "*A Novel Approach to Automatic Identification of Kannada, English and Hindi Words from a Trilingual Document*", Int. conf. on Signal and Image Processing, Hubli, pp. 561-566, (2006).
- [14] S. L. Wood, X. Yao, K. Krishnamurthy and L. Dang, "*Language identification for printed text independent of segmentation*", Proc. Int. Conf. on Image Processing, pp. 428-431, 0-8186-7310-9/95, 1995 IEEE.
- [15] J. Hochberg, L. Kerns, P. Kelly and T. Thomas, "*Automatic script identification from images using cluster based templates*", IEEE Trans. Pattern Anal. Machine Intell. Vol. 19, No. 2, pp. 176-181, 1997.
- [16] Shivakumar, Nagabhushan, Hemanthkumar, Manjunath, 2006, "*Skew Estimation by Improved Boundary Growing for Text Documents in South Indian Languages*", VIVEK-International Journal of Artificial Intelligence, Vol. 16, No. 2, pp 15-21.
- [17] Andrew Busch, Wageeh W. Boles and Sridha Sridharan, "*Texture for Script Identification*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 11, pp. 1720-1732, Nov. 2005.

- [18] A. L. Spitz, "*Determination of script and language content of document images*", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19, No.3, pp. 235–245, 1997.
- [19] Ramachandra Manthalkar and P.K. Biswas, "*An Automatic Script Identification Scheme for Indian Languages*", IEEE Tran. on Pattern Analysis And Machine Intelligence, vol.19, no.2, pp.160-164, Feb.1997.
- [20] Hiremath P S and S Shivashankar, "*Wavelet Based Co-occurrence Histogram Features for Texture Classification with an Application to Script Identification in a Document Image*", Pattern Recognition Letters 29, 2008, pp 1182-1189.

## Wavelet Packet Based Texture Features for Automatic Script Identification

**M.C. Padma**

padmapes@gmail.com

*Dept. of Computer Science & Engineering,  
PES College of Engineering,  
Mandya-571401, Karnataka, India*

**P. A. Vijaya**

pavmkv@gmail.com

*Dept. of E. & C. Engineering,  
Malnad College of Engineering,  
Hassan-573201, Karnataka, India*

---

### Abstract

In a multi script environment, an archive of documents printed in different scripts is in practice. For automatic processing of such documents through Optical Character Recognition (OCR), it is necessary to identify the script type of the document. In this paper, a novel texture-based approach is presented to identify the script type of the collection of documents printed in ten Indian scripts - Bangla, Devanagari, Roman (English), Gujarati, Malayalam, Oriya, Tamil, Telugu, Kannada and Urdu. The document images are decomposed through the Wavelet Packet Decomposition using the Haar basis function up to level two. Gray level co-occurrence matrix is constructed for the coefficient sub bands of the wavelet transform. The Haralick texture features are extracted from the co-occurrence matrix and then used in the identification of the script of a machine printed document. Experimentation conducted involved 3000 text images for learning and 2500 text images for testing. Script classification performance is analyzed using the K-nearest neighbor classifier. The average success rate is found to be 98.24%.

**Keywords:** Document Processing, Wavelet Packet Transform, Feature Extraction, Script Identification.

---

### 1. INTRODUCTION

The progress of information technology and the wide reach of the Internet are drastically changing all fields of activity in modern days. As a result, a very large number of people would be required to interact more frequently with computer systems. To make the man-machine interaction more effective in such situations, it is desirable to have systems capable of handling inputs in the form of printed documents. If the computers have to efficiently process the scanned images of printed documents, the techniques need to be more sophisticated. Even though computers are used widely in almost all the fields, undoubtedly paper documents occupy a very

important place for a longer period. Also, a large proportion of all kinds of business writing communication exist in physical form for various purposes. For example, to fax a document, to produce a document in the court, etc. Therefore, software to automatically extract, analyze and store information from the existing paper form is very much needed for preservation and access whenever necessary. All these processes fall under the category of document image analysis, which has received significance as a major research problem in the modern days.

Script identification is an important problem in the field of document image processing, with its applications to sort document images, as pre processor to select specific OCRs, to search online archives of document images for those containing a particular language, to design a multi-script OCR system and to enable automatic text retrieval based on script type of the underlying document. Automatic script identification has been a challenging research problem in a multilingual environment over the last few years. All existing works on automatic language identification are classified into either local approach or global approach. Ample work has been reported in literature using local approaches [1, 3, 7-10]. The local features are extracted from the water reservoir principle [1, 3], morphological features [8], profile, cavities, corner points, end point connectivity [13], top and bottom profile based features [11, 12]. In local approaches, the features are extracted from a list of connected components such as line, word and character, which are obtained only after segmenting the underlying document image. So, the success rate of classification depends on the effectiveness of the pre-processing steps namely, accurate Line, Word and Character segmentation. It sounds paradoxical as LWC segmentation can be better performed, only when the script class of the document is known. Even when the script classes are known from the training data, testing requires the performance of LWC segmentation prior to script identification. But, it is difficult to find a common segmentation method that best suits for all the script classes. Due to this limitation, local approaches cannot meet the criterion as a generalized scheme.

In contrast, global approaches employ analysis of regions comprising of at least two text lines and hence fine segmentation of the underlying document into line, word and character, is not necessary. Consequently, the script classification task is simplified and performed faster with the global approach than the local approach. So, global schemes can be best suited for a generalized approach to the script identification problem. Adequate amount of work has been reported in literature using global approaches [2, 4, 6]. Santanu Choudhuri, et al. [4] have proposed a method for identification of Indian languages by combining Gabor filter based technique and direction distance histogram classifier considering Hindi, English, Malayalam, Bengali, Telugu and Urdu. Gopal Datt Joshi, et. al. [6] have presented a script identification technique for 10 Indian scripts using a set of features extracted from log-Gabor filters. Dhanya et al. [14] have used Linear Support Vector Machine (LSVM), K-Nearest Neighbour (K-NN) and Neural Network (NN) classifiers on Gabor-based and zoning features to classify Tamil and English scripts. Recently, Hiremath et al. [15] have proposed a novel approach for script identification of South Indian scripts using wavelet based co-occurrence histogram features. Ramachandra Manthalkar et.al. [16] have proposed a method based on rotation-invariant texture features using multichannel Gabor filter for identifying seven Indian languages namely Bengali, Kannada, Malayalam, Oriya, Telugu and Marathi. They [16] have used multichannel Gabor filters to acquire rotation invariant texture features. From their experiment, they observed that rotation invariant features provide good results for script identification. Srinivas Rao Kunte et al. [17] have suggested a neural approach in on-line script recognition for Telugu language employing wavelet features. Peeta Basa Pati et al. [18] have presented a technique using Gabor filters for document analysis of Indian bilingual documents.

Sufficient amount of work has also been carried out on non-Indian languages [2, 22-30]. One of the first attempts in automatic script and language recognition is due to Spitz and his coworkers [23]. Assuming connected components of individual characters have been extracted from the document image, Spitz first locates upward concavities in the connected components. He then discriminates Asian languages (Japanese, Chinese, and Korean) against European languages (English, French, German, and Russian) based on the vertical distributions of such concavities. The three Asian languages are differentiated from each other by comparing the statistics of the optical densities (the number of black pixels per unit area) of the connected components, whereas the European languages are discriminated by means of the most frequent occurring

word shape tokens also derived from the connected components [23]. Tan [2] has developed a rotation invariant texture feature extraction method for automatic script identification for six languages: Chinese, Greek, English, Russian, Persian and Malayalam. Peake and Tan [19] have proposed a method for automatic script and language identification from document images using multiple channel (Gabor) filters and gray level co-occurrence matrices for seven languages: Chinese, English, Greek, Korean, Malayalam, Persian and Russian.

Hochberg et al. [25] have presented a method for automatically identifying script from a binary document image using cluster-based text symbol templates. The system develops a set of representative symbols (templates) for each script by clustering textual symbols from a set of training documents and represents each cluster by its centroid. "Textual symbols" include discrete characters in scripts such as Cyrillic, as well as adjoined characters, character fragments and whole words in connected scripts such as Arabic. To identify a new document's script, the system compares a subset of symbols from the document to each script's templates, screening out rare or unreliable templates and choosing the script whose templates provide the best match. Later, they have extended their work on thirteen scripts - Arabic, Armenian, Burmese, Chinese, Cyrillic, Devanagari, Ethiopic, Greek, Hebrew, Japanese, Korean, Roman, and Thai. Chew Lim Tan et al. [24] presents a technique of identification of English, Chinese, Malay and Tamil in image documents using features like bounding boxes of character cells and upward concavities. Andrew Busch et al. [22] have exploited the concept of texture features for script identification of English, Chinese, Greek, Cyrillic, Hebrew, Hindi, Japanese and Persian. Wood et al. [24] have proposed projection profile method to determine Roman, Russian, Arabic, Korean and Chinese characters. Zhu et al. have presented an unconstrained language identification technique using a shape codebook concept [26]. Later, Zhu et al. have extended their work on language identification of handwritten document images [27]. Lu et al. have explored the challenging problem of script and language identification on degraded and distorted document images [28, 29]. Stefan Jaeger et al [30] have proposed a multiple classifier system for script identification at word level using Gabor filter analysis of textures. Their system identifies Latin and non-Latin words in bilingual printed documents. The classifier system comprises of four different architectures based on nearest neighbors, weighted Euclidean distances, Gaussian mixture models, and support vector machines. Global approaches make use of the texture-based features. These texture features can be extracted from a portion of a text region that may comprise of several text lines.

Texture could be defined in simple form as "repetitive occurrence of the same pattern". Texture could be defined as something consisting of mutually related elements. Another definition of texture claims that, "an image region has a constant texture if a set of its local properties in that region is constant, slowly changing or approximately periodic". Texture classification is a fundamental issue in image analysis and computer vision. It has been a focus of research for nearly three decades. Briefly stated, there are a finite number of texture classes  $C_i$ ,  $i = 1, 2, 3, n$ . A number of training samples of each class are available. Based on the information extracted from the training samples, a decision rule is designed, which classifies a given sample of unknown class into one of the  $n$  classes [2]. Image texture is defined as a function of the spatial variation in pixel intensities. The texture classification is fundamental to many applications such as automated visual inspection, biomedical image processing, content-based image retrieval and remote sensing. One application of image texture is the recognition of image regions using texture properties. From the literature survey, it is observed that sufficient work has been carried out using texture features [11-14, 18, 20]. Existing methods on Indian script identification use the texture features extracted from the co-occurrence matrix, wavelet based co-occurrence histogram [15] and Gabor filters [17, 18]. Hiremath and Shivakumar [31] have considered Haralick features for texture classification using wavelet packet decomposition. Very few works are reported on script identification particularly using wavelet transform based features. In this paper, the texture features useful for script identification are extracted from the co-occurrence matrix constructed from the sub band coefficients of the wavelet packets transforms. As such, no work has been reported that uses the wavelet packet based texture features for script identification.

The rest of the paper is organized as follows. The Section 2 briefs about the wavelet packet transform. The database constructed for testing the proposed model is presented in Section 3. In

Section 4, complete description of the proposed model is explained in detail. The experimental results obtained are presented in section 5. Conclusions are given in section 6.

## 2. WAVELET PACKET TRANSFORM

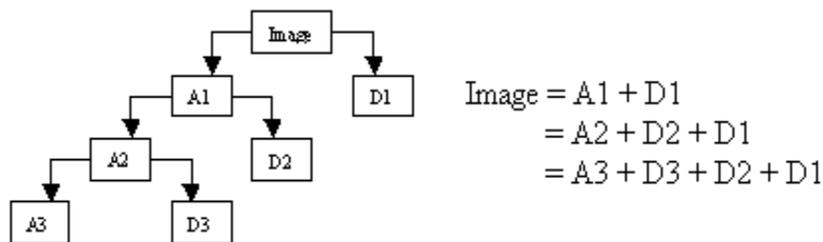
Research interest in wavelets and their applications has grown tremendously over the past few years. It has been shown that wavelet-based methods continue to be powerful mathematical tools and offer computational advantage over other methods for texture classification. The different wavelet transform functions filter out different range of frequencies (i.e. sub bands). Thus, wavelet is a powerful tool, which decomposes the image into low frequency and high frequency sub band images.

The Continuous Wavelet Transform (CWT) is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function  $\psi$  given in Equation (1).

$$C(scale, position) = \int_{-\infty}^{\infty} f(t)\psi(scale, position, t)dt \tag{1}$$

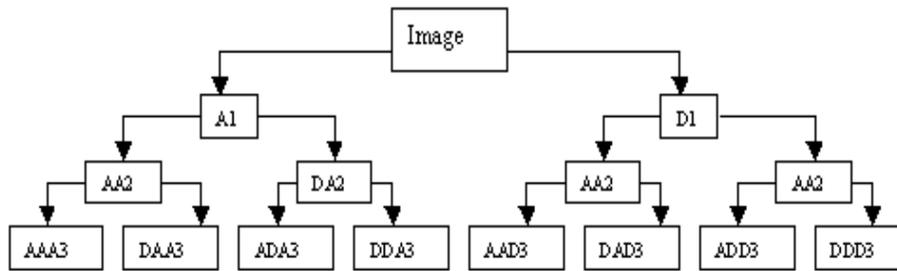
The results of the CWT are many wavelet coefficients  $C$ , which are functions of scale and position. The wavelet transform decomposes a signal into a series of shifted and scaled versions of the mother wavelet function. Due to time frequency localization properties, discrete wavelet and wavelet packet transforms have proven to be appropriate starting point for classification tasks. In the 2-D case, the wavelet transform is usually performed by applying a separable filter bank to the image. Typically, a low filter and a band pass filter are used. The convolution with the low pass filter results in the approximation image and the convolutions with the band pass filter in specific directions result in the detail images.

In the simple wavelet decomposition, only the approximation coefficients are split iteratively into a vector of approximation coefficients, and a vector of detail coefficients are split at a coarser scale. That means, for  $n$ -level decomposition,  $n+1$  possible ways of decomposition are obtained as shown in Figure 1. The successive details are never reanalyzed in the case of simple wavelet decomposition.



**FIGURE 1.** Wavelet Decomposition Tree

The concept of wavelet packets was introduced by Coifman et.al. [23]. In the case of wavelet packets, each detail coefficient vector is also decomposed as in approximation vectors. The recursive splitting of both approximate and detail sub images will produce a binary tree structure as shown in Figure 2.



**FIGURE 2.** Wavelet Packet Decomposition Tree

Generally, in the case of wavelet transforms, the features are extracted from only the approximate sub band coefficients. But, in the case of wavelet packet transforms, the features are extracted from both approximate and detail sub band coefficients, as the detail sub band images can also be decomposed recursively. The features derived from a detail images uniquely characterize a texture. The combined transformed coefficients of the approximate and detail images give efficient features and hence could be used as essential features for texture analysis and classification. In this paper, the features are extracted from the sub bands of the transformed images for script identification and the complete description of the proposed model is given in Section 4.

### 3. DATA COLLECTION

Standard database of documents of Indian languages is currently not available. For the proposed model, the data set was constructed from the scanned document images. The printed documents like books, newspapers, journals and magazines were scanned through an optical scanner to obtain the document image. The scanner used for obtaining the digitized images is HP Scan Jet 5200c series. The scanning is performed in normal 100% view size at 300 dpi resolution. The image size of 256x256 pixels was considered. The training data set of 300 images and test data set of 250 images were used from each of the ten scripts.

### 4. OVERVIEW OF THE PROPOSED MODEL

Scripts are made up of different shaped patterns to produce different character sets. Individual text patterns of one script are collected together to form meaningful text information in the form of a text word, a text line or a paragraph. The collection of the text patterns of the one script exhibits distinct visual appearance. So, a uniform block of texts, regardless of the content, may be considered as distinct texture patterns (a block of text as single entity) [2]. This observation implies that one may devise a suitable texture classification algorithm to perform identification of text language. Hence, the proposed model is inspired by this simple observation that every language script defines a finite set of text patterns, each having a distinct visual appearance. In this model, the texture-based features are extracted from the sub band coefficients of wavelet packet transforms.

#### 4.1 Preprocessing of Input Images

In general, text portion of the scanned document images are not good candidates for the extraction of texture features. The varying degrees of contrast in gray-scale images, the presence of skew and noise could all potentially affect such features, leading to higher classification error rates. In addition, the large areas of white space, unequal word and line spacing, and variable height of the text lines due to different font sizes can also have a significant effect on the texture

features. In order to reduce the impact of these factors, the text blocks from which texture features are to be extracted must undergo a significant amount of preprocessing. In this paper, preprocessing steps such as removal of non-text regions, skew-correction, noise removal and binarization is necessary. In the proposed model, text portion of the document image was separated from the non-text region manually. Skew detection and correction was performed using the technique proposed by Shivakumar [21]. Binarization can be described as the process of converting a gray-scale image into one, which contains only two distinct tones, that is black and white. In this work, a global thresholding approach is used to binarize the scanned gray scale images, where black pixels having the value 0's correspond to object and white pixels having value 1's correspond to background. It is necessary to thin the document image as the texts may be printed in varying thickness. In this paper, the thinning process is achieved by using the morphological operations.

#### *4.1.1 To Construct Uniform Block of Text*

In general, the scanned image may not necessarily have uniform height of the text line Height of the text line is the difference between the topmost black pixel to the bottommost black pixel of a text line and it is obtained through horizontal projection profile. To obtain text lines of uniform height, the necessary threshold values are computed by calculating the mean and the standard deviation of the text line heights. All those text lines, which fall outside the threshold values (fixed through experimentation), are removed. This process is repeated by calculating the new mean and standard deviation of the remaining text block, until no remaining text lines are removed.

#### *4.1.2 Inter-line Spacing Normalization*

In a block of text, the white space between the text lines may vary and hence, it is necessary to normalize these white spaces. The width of the white spaces between the text lines is obtained by computing the distance between the horizontal runs of black pixels of the two text lines. The white space width of the text lines greater than eight pixels is reduced to eight pixels, which is fixed by experimental study. Thus, the spacing between the text lines is normalized to have uniform inter-line spacing.

#### *4.1.3 Inter-word Spacing Normalization*

Generally, some text lines might have varying inter-word spacing. So, it is necessary to normalize the inter-word spacing to a maximum of 5 pixels. Normalization of the inter-word spacing is achieved by projecting the pixels of each text line vertically; counting the number of white pixels from left to right and reducing the number of white pixels greater than 5 pixels to 5. Inter-word gaps smaller than 5 pixels are allowed to remain, as that does not affect to get text blocks.

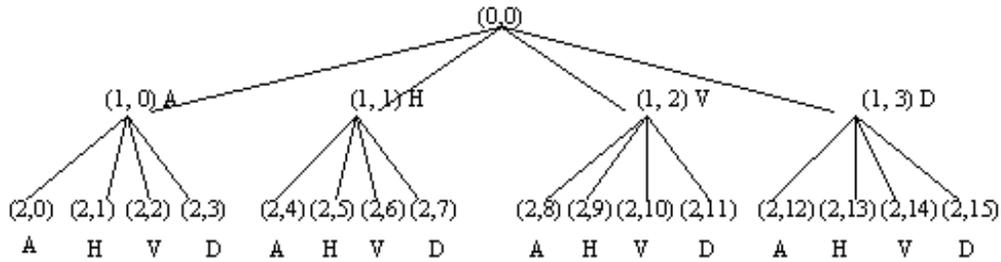
#### *4.1.4 Padding*

A text line that is smaller than the average width of a text line generally appears in a block of text. This is due to the presence of headings or the text line that happens to be the last line of a paragraph. So, it is necessary to obtain a collection of text lines of equal length by padding the text lines of smaller length. Padding of text lines is done by copying and replicating the text content of the line so that the length of the text line would be increased to the average length of text line.

Then, the text block of size 256X256 pixels is extracted from the preprocessed text block. From each block of preprocessed text image, the texture features are extracted for the purpose of script identification.

## **4.2 Texture Feature Extraction**

In this work, the known input images are decomposed through the Wavelet Packet Decomposition using the Haar (Daubechies 1) basis function to get the four sub band images namely Approximation (A) and three detail coefficients - Horizontal (H), Vertical (V) and the Diagonal (D). The Haar wavelet transformation is chosen because the resulting wavelet bands are strongly correlated with the orientation elements in the GLCM computation. The second reason is that the total pixel entries for Haar wavelet transform are always minimum. Through experimentation the Haar basis function up to the level two is found to be best, yielding distinct features and hence Haar basis function up to level two is used in this method. This result in a total of 20 sub bands, four sub bands at the first level and sixteen sub bands (four for each sub band) in the next level as shown in Figure 3.



**FIGURE 3.** Wavelet Packet Tree up to Level - 2. (A–Approximation, H–Horizontal, V–Vertical and D–Diagonal)

It is not necessary to consider all the twenty sub bands for feature extraction. This is because, the four types of sub bands – approximation, horizontal, vertical and diagonal obtained from the wavelet transforms retain specific type of information by filtering out other information. Therefore, when the horizontal, vertical and diagonal sub bands are decomposed further into four bands, all the four sub bands may not be necessarily used as some information in the second level is lost. So, it is necessary to consider only the relevant sub bands at the second level. For example, in the sub band (1, 1) of level one which gives only horizontal detail coefficients, it is sufficient to consider only the approximation and horizontal detail coefficients of its second level. The vertical and diagonal sub bands of (1, 1) are not considered as they exhibit less or poor information. Thus, the sub bands that exhibit the similar type of coefficients from the two levels are selected. In the proposed method, the sub bands of the two levels are combined to form four groups as given below.

- Group 1: Approximation sub bands: (1, 0), (2, 0) = (A, AA)
- Group 2: Horizontal sub bands: (1, 1), (2, 1), (2, 4), (2, 5) = (H, AH, HA, HH)
- Group 3: Vertical sub bands: (1, 2), (2, 2), (2, 8), (2, 10) = (V, AV, VA, VV)
- Group 4: Diagonal sub bands: (1, 3), (2, 3), (2, 12), (2, 15) = (D, AD, DA, DD)

Thus, only fourteen sub bands - two approximate sub band, four horizontal sub band, four vertical sub bands and four diagonal sub bands are selected out of the twenty sub bands.

Many researchers have used the coefficient values of the approximate and detail sub band images as a texture feature vector [15]. According to Andrew Busch [22], a better representation of natural textured images can be obtained by applying a nonlinear function to the coefficients of the wavelet transform. In this paper, the nonlinear transform function proposed by Andrew Busch [22] is applied on the wavelet packet coefficients. So, the wavelet packet coefficients are quantized using the quantization function derived by Andrew Busch [22]. Then, gray level co-occurrence matrices are constructed for the quantized wavelet sub bands. The description of the gray level co-occurrence matrix is briefed out in the next section.

#### 4.2.1. Gray Level Co-occurrence Matrices (GLCMs)

Gray Level Co-occurrence Matrix (GLCM) has been proven to be a very powerful tool for texture image segmentation. Gray-level co-occurrence matrices (GLCMs) are used to represent the pair wise joint statistics of the pixels of an image and have been used for many years as a means of characterizing texture [22]. GLCM is a two dimensional measure of texture, which show how often each gray occurs at a pixel located at a fixed geometric position relative to each other pixel, as a function of its gray level. GLCMs are in general very expensive to compute due to the requirement that the size of each matrix is  $N \times N$ , where  $N$  is the number of gray levels in the image. So, it is necessary to reduce the number of discrete gray levels of the input image in order to obtain co-occurrence matrix of smaller size. So, if the gray levels are divided into fewer ranges, the size of the matrix would be reduced, thus leading to less noisy entries in the matrix.

In this paper, the gray levels of the quantized sub bands are divided into fewer ranges to obtain a new transformed sub band which results in reduced size of the co-occurrence matrix. Then, from

the new transformed sub bands, GLCMs are constructed using the values  $d=1$ , where  $d$  represents the linear distance in pixels. The value of  $\theta$  is fixed based on the type of the sub band. For the approximate sub bands i.e., group1 ((1, 0), (2, 0)), GLCMs are constructed with the value  $\theta = \{00, 450, 900, 1350\}$ . The value of  $\theta$  is taken as 00 for horizontal sub bands (group2), 900 for vertical sub bands (group3) and, 450 and 1350 for diagonal sub bands (group4). Thus, totally, twenty four GLCM (eight GLCM for group1, four GLCM for group2, four GLCM for group3 and eight GLCM for group4) are constructed.

Haralick [32] has proposed the textural features that can be extracted from the co-occurrence matrix. In this paper, Haralick texture features [32] such as inertia, total energy, entropy, contrast, local homogeneity, cluster shade, cluster prominence, and information measure of correlation are extracted from the gray level co-occurrence matrices obtained from the coefficients of the sub bands. These texture features are given in Table 1. These features are known as the wavelet packet co-occurrence features.

**TABLE 1.** Wavelet Packet Co-occurrence Features Extracted from a Co-occurrence Matrix  $C(i, j)$ .

|                                     |                                                                                                          |
|-------------------------------------|----------------------------------------------------------------------------------------------------------|
| Inertia:                            | $F1 = \sum_{i,j=0}^n (i - j)^2 C(i, j)$                                                                  |
| Total Energy:                       | $F2 = \sum_{i,j=0}^n C^2(i, j)$                                                                          |
| Entropy:                            | $F3 = - \sum_{i,j=0}^n C(i, j) \log C(i, j)$                                                             |
| Contrast:                           | $F4 = - \sum_{i,j=0}^n C(i, j)  i - j ^k, k \in \mathbb{Z}$                                              |
| Local Homogeneity:                  | $F5 = \sum_{i,j=0}^n \frac{1}{1 + (i - j)^2} C(i, j)$                                                    |
| Cluster Shade:                      | $F6 = \sum_{i,j=0}^n (i - M_x + j - M_y)^3 C(i, j)$                                                      |
| Cluster Prominence:                 | $F7 = \sum_{i,j=0}^n (i - M_x + j - M_y)^4 C(i, j)$                                                      |
| Information Measure of Correlation: | $F8 = \frac{- \sum_{i,j=0}^n C(i, j) \log C(i, j) - H_{x,y}}{\max(H_x, H_y)}$                            |
| where                               | $M_x = \sum_{i,j=0}^n iC(i, j) \quad \text{and} \quad M_y = \sum_{i,j=0}^n jC(i, j)$                     |
|                                     | $H_{x,y} = - \sum_{i,j=0}^n C(i, j) \log \left( \sum_{j=0}^n C(i, j) \cdot \sum_{i=0}^n C(i, j) \right)$ |
|                                     | $H_x = - \sum_{i=0}^n \left\{ \sum_{j=0}^n P(i, j) \cdot \log \sum_{j=0}^n P(i, j) \right\}$             |
|                                     | $H_y = - \sum_{j=0}^n \left\{ \sum_{i=0}^n P(i, j) \cdot \log \sum_{i=0}^n P(i, j) \right\}$             |

The eight Haralick texture features are extracted from the twenty four GLCMs resulting in a total of 192 features. In order to reduce the dimension of the features, the mean values of the eight features are computed individually from each of the four groups. That means, eight features of any sub band is obtained by taking the average of the eight features computed from the GLCMs of that sub band. For example, from the four GLCMs of the sub band (1, 0), average of the eight features are computed. Thus, eight features of the sub band (1, 0) and eight features of sub band (2, 0) of group 1, results in 16 texture features. But, the average value of each of the eight features is computed individually from the sub bands of the corresponding groups. For example, the average value of the feature – ‘Inertia’ for the Group-1 is computed from the inertia of sub band (1, 0) and the inertia of sub band (2, 0). Similarly, the average value of each of the eight features is computed from sub bands of the individual groups, resulting in 32 features (8 features each from four groups), thus reducing the dimensionality of the features from 192 to 32. As the features are extracted from the two levels of wavelet packet transforms and then the average of the feature values are computed, the features could be considered as optimal features. These optimal features are strong enough to discriminate the ten script classes considered in this paper. The 32 optimal features are obtained from a training data set of 300 images from each of the ten Indian scripts - Bangla, Devanagari, Roman (English), Gujarati, Malayalam, Oriya, Tamil, Telugu, Kannada and Urdu. These features are stored in a feature library and used as texture features later in the testing stage.

### 4.3 Classification

In the proposed model, K -nearest neighbor classifier is used to classify the test samples. The features are extracted from the test image X using the proposed feature extraction algorithm and then compared with corresponding feature values stored in the feature library using the Euclidean distance formula given in equation (2),

$$D(M) = \sqrt{\sum_{j=1}^N [f_j(x) - f_j(M)]^2} \quad (2)$$

where N is the number of features in the feature vector f,  $f_j(x)$  represents the jth feature of the test sample X and  $f_j(M)$  represents the jth feature of Mth class in the feature library. Then, the test sample X is classified using the k-nearest neighbor (K-NN) classifier. In the K -NN classifier, a test sample is classified by a majority vote of its k neighbors, where k is a positive integer, typically small. If K =1, then the sample is just assigned the class of its nearest neighbor. It is better to choose K to be an odd number to avoid tied votes. So, in this method, the K -nearest neighbors are determined and the test image is classified as the script type of the majority of these K-nearest neighbors. The testing algorithm employed in this paper for script identification consists of the following steps.

#### Algorithm Testing ()

Input: Text portion of the document image containing one script only.

Output: Script type of the test document.

1. Preprocess the input document image.
2. Analyze the test image using 2-d Wavelet Packet Transform with Haar wavelet up to level 2 and obtain the wavelet packet tree.
3. Quantize the wavelet coefficients using the quantization function derived by Andrew Busch [24].
4. Select the sub bands from the wavelet packet tree as given below:
  - Group 1: Approximation sub bands: (1, 0), (2, 0) = (A, AA)
  - Group 2: Horizontal sub bands: (1, 1), (2, 1), (2, 4), (2, 5) = (H, AH, HA, HH)
  - Group 3: Vertical sub bands: (1, 2), (2, 2), (2, 8), (2, 10) = (V, AV, VA, VV)
  - Group 4: Diagonal sub bands: (1, 3), (2, 3), (2, 12), (2, 15) = (D, AD, DA, DD)
5. Construct the Gray Level Co-occurrence Matrix of each sub band.

6. Extract the eight Haralick texture features given in Table 1 from the selected sub bands.
7. Compute the average value for each of the eight features from the four groups as explained in the previous section.
8. Classify the script type of the test image by comparing the feature values of the test image with the feature values stored in the knowledge base using K-nearest neighbor classifier.

The experiment is conducted for varying number of neighbors like  $K = 3, 5$  and  $7$ . The performance of classification was best when the value of  $K = 3$ . Thus the script type of the test image is classified by comparing the feature values of the test image with the feature values stored in the feature matrix using K-nearest neighbor classifier.

### 5. EXPERIMENTAL RESULTS

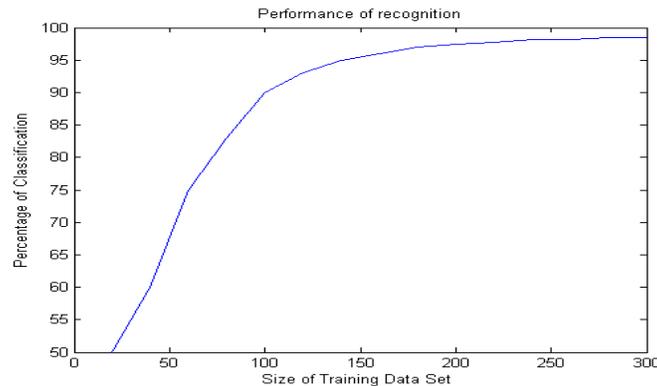
To test the proposed model test data set was constructed from the scanned images of newspapers, magazines, journals and books at 300 dpi grayscale. Around 250 images were chosen from each of ten Indian scripts namely Bangla, Devanagari, Roman (English), Gujarati, Malayalam, Oriya, Tamil, Telugu, Kannada and Urdu. Only the text portion of the scanned images was selected. However, images that would contain multiple columns and variable inter-line and inter-word spacing were also considered as the preprocessing steps are used to prepare them for further processing. Sample images of Bangla, Hindi, Kannada, English, Gujarati, Malayalam, Oriya, Telugu, Urdu and Tamil scripts are shown in Figure 4. The proposed algorithm is implemented using Matlab R2007b. The average time taken to identify the script type of the document is 0.2843 seconds on a Pentium-IV with 1024 MB RAM based machine running at 1.60 GHz.



FIGURE 4. Sample images of Bangla, Hindi, Kannada, English, Gujarati, Malayalam, Oriya, Telugu, Urdu and Tamil scripts.

### 5.1 Optimal Size of Training Data Set

It is necessary to determine the optimal size of the training data set to obtain best performance of the proposed model. In the proposed model, the test samples are tested with varying number of training data set. The overall performance of recognition verses size of the training data set is shown in Figure 5. From the Figure 5, it is observed that the proposed method attains best performance of 98.24% with an optimal training data set of 300 samples.



**FIGURE 5.** The overall performance of recognition verses size of the training data set.

The confusion matrix of the proposed method for classifying twelve Indian scripts through extensive experimentation is given in Table 2. The average classification accuracy of the proposed wavelet based method is 98.24%. The experimental results demonstrate the effectiveness of the proposed texture features.

Test images having some special characters like numerals, punctuation marks and italicized text were also considered. However, the presence of these symbols does not affect the recognition rate as they are rarely seen in a meaningful text region. A small amount of page skew was inevitably introduced during the scanning process. This skew was compensated by using the method outlined in [21].

**TABLE 2.** Percentage of Recognition of 10 Indian scripts (Ban=Bangla, Dev=Devanagari, Rom=Roman (English), Guj=Gujarati, Mal=Malayalam, Ori=Oriya, Tam=Tamil, Tel=Telugu, Kan=Kannada and Urd=Urdu)

| Input Type                   | Script | Classified Script Type |      |      |      |      |      |      |      |      |      |          |
|------------------------------|--------|------------------------|------|------|------|------|------|------|------|------|------|----------|
|                              |        | Ban                    | Dev  | Rom  | Guj  | Ori  | Mal  | Tam  | Tel  | Kan  | Urd  | Rejected |
| Ban                          |        | 246                    | 3    | -    | 1    | -    | -    | -    | -    | -    | -    | -        |
| Dev                          |        | 2                      | 248  | -    | -    | -    | -    | -    | -    | -    | -    | -        |
| Rom                          |        | -                      | -    | 245  | -    | 2    | -    | 2    | -    | -    | -    | 1        |
| Guj                          |        | -                      | -    | 2    | 246  | 1    | -    | -    | -    | -    | -    | 1        |
| Ori                          |        | -                      | -    | 2    | 1    | 245  | -    | 1    | -    | -    | -    | 1        |
| Mal                          |        | -                      | -    | 1    | -    | -    | 244  | 2    | -    | 2    | -    | 1        |
| Tam                          |        | -                      | -    | 3    | -    | 2    | -    | 245  | -    | -    | -    | -        |
| Tel                          |        | -                      | -    | -    | -    | -    | 2    | 1    | 244  | 3    | -    | -        |
| Kan                          |        | -                      | -    | -    | -    | -    | 2    | -    | 2    | 246  | -    | -        |
| Urd                          |        | -                      | -    | -    | -    | -    | -    | -    | -    | -    | 247  | 3        |
| Percentage of classification |        | 98.4                   | 99.2 | 98.0 | 98.4 | 98.0 | 97.6 | 98.0 | 97.6 | 98.4 | 98.8 | -        |
| Error                        |        | 1.6                    | 0.8  | 2.0  | 1.6  | 2.4  | 2.8  | 2.0  | 2.8  | 1.6  | 1.2  | -        |

## 6. CONSLUSION

In this paper, a global approach for script identification that uses the wavelet based texture features is presented. The texture features are extracted from the GLCMs constructed from a set of wavelet packet sub band coefficients. The experimental results demonstrate that the new approach can better perform in classifying ten Indian scripts. The performance of the proposed model shows that the global approach could be used to solve a practical problem of automatic script identification.

## 7. REFERENCES

1. U.Pal, B.B.Choudhuri, : Script Line Separation From Indian Multi-Script Documents, 5th Int. Conference on Document Analysis and Recognition(IEEE Comput. Soc. Press), 406-409, (1999).
2. T.N.Tan, : Rotation Invariant Texture Features and their use in Automatic Script Identification, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 7, pp. 751-756, (1998).
3. U. Pal, S. Sinha and B. B. Chaudhuri : Multi-Script Line identification from Indian Documents, In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003) 0-7695-1960-1/03 © 2003 IEEE, vol.2, pp.880-884, (2003).
4. Santanu Choudhury, Gaurav Harit, Shekar Madnani, R.B. Shet, : Identification of Scripts of Indian Languages by Combining Trainable Classifiers, ICVGIP, Dec.20-22, Bangalore, India, (2000).
5. S. Chaudhury, R. Sheth, "Trainable script identification strategies for Indian languages", In Proc. 5th Int. Conf. on Document Analysis and Recognition (IEEE Comput. Soc. Press), pp. 657-660, 1999.
6. Gopal Datt Joshi, Saurabh Garg and Jayanthi Sivaswamy, :Script Identification from Indian Documents, LNCS 3872, pp. 255-267, DAS (2006).
7. S.Basavaraj Patil and N V Subbareddy, : Neural network based system for script identification in Indian documents", Sadhana Vol. 27, Part 1, pp. 83-97. © Printed in India, (2002).
8. B.V. Dhandra, Mallikarjun Hangarge, Ravindra Hegadi and V.S. Malemath, : Word Level Script Identification in Bilingual Documents through Discriminating Features, IEEE - ICSCN 2007, MIT Campus, Anna University, Chennai, India. pp.630-635. (2007).
9. U. Pal and B. B. Chaudhuri, "Automatic separation of Roman, Devnagari and Telugu script lines", Advances in Pattern Recognition and Digital techniques, pp. 447-451, 1999.
10. Lijun Zhou, Yue Lu and Chew Lim Tan, : Bangla/English Script Identification Based on Analysis of Connected Component Profiles, in proc. 7th DAS, pp. 243-254, (2006).
11. M. C. Padma and P.Nagabhushan, : Identification and separation of text words of Karnataka, Hindi and English languages through discriminating features, in proc. of Second National Conference on Document Analysis and Recognition, Karnataka, India, pp. 252-260, (2003).
12. M. C. Padma and P.A.Vijaya, : Language Identification of Kannada, Hindi and English Text Words Through Visual Discriminating Features, International Journal of Computational Intelligence Systems (IJCIS), Volume 1, Issue 2, pp. 116-126, (2008).
13. Vipin Gupta, G.N. Rathna, K.R. Ramakrishnan, : A Novel Approach to Automatic Identification of Kannada, English and Hindi Words from a Trilingual Document, Int. conf. on Signal and Image Processing, Hubli, pp. 561-566, (2006).
14. D. Dhanya, A. G. Ramakrishnan, and P. B. Pati, "Script identification in printed bilingual documents", Sadhana, vol. 27, pp. 73-82, 2002.
15. Hiremath P S and S Shivashankar, "Wavelet Based Co-occurrence Histogram Features for Texture Classification with an Application to Script Identification in a Document Image", Pattern Recognition Letters 29, 2008, pp 1182-1189.
16. Ramachandra Manthalkar and P.K. Biswas, "An Automatic Script Identification Scheme for Indian Languages", IEEE Tran. on Pattern Analysis And Machine Intelligence, vol.19, no.2, pp.160-164, Feb.1997.

17. R. Sanjeev Kunte and R.D. Sudhakar Samuel, "On Separation of Kannada and English Words from a Bilingual Document Employing Gabor Features and Radial Basis Function Neural Network", Proc. of ICCR, pp. 640-644, 2005.
18. Peeta Basa Pati, S. Sabari Raju, Nishikanta Pati and A. G. Ramakrishnan, "Gabor filters for Document analysis in Indian Bilingual Documents", 0-7803-8243-9/04/ IEEE, ICISIP, pp. 123-126, 2004.
19. G. S. Peake and T. N. Tan, "Script and Language Identification from Document Images", Proc. Workshop Document Image Analysis, vol. 1, pp. 10-17, 1997.
20. Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins,; Digital Image Processing using MATLAB, Pearson Education, (2004).
21. Shivakumar, Nagabhushan, Hemanthkumar, Manjunath, 2006, "Skew Estimation by Improved Boundary Growing for Text Documents in South Indian Languages", VIVEK-International Journal of Artificial Intelligence, Vol. 16, No. 2, pp 15-21.
22. Andrew Busch, Wageeh W. Boles and Sridha Sridharan, "Texture for Script Identification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 11, pp. 1720-1732, Nov. 2005.
23. A. L. Spitz, "Determination of script and language content of document images", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19, No.3, pp. 235-245, 1997.
24. Chew Lim Tan, Peck Yoke Leong, Shoujie He, "Language Identification in Multilingual Documents", International symposium on intelligent multimedia and distance education, 1999.
25. J. Hochberg, L. Kerns, P. Kelly and T. Thomas, "Automatic script identification from images using cluster based templates", IEEE Trans. Pattern Anal. Machine Intell. Vol. 19, No. 2, pp. 176-181, 1997.
26. Guangyu Zhu, Xiaodong Yu, Yi Li and David Doermann, "Unconstrained Language Identification Using A Shape Codebook", The 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), pp. 13-18, 2008.
27. Guangyu Zhu, Xiaodong Yu, Yi Li and David Doermann, "Language Identification for Handwritten Document Images Using A Shape Codebook", Pattern Recognition, 42, pp. 3184-3191, December 2009.
28. Lu S and C.L. Tan, "Script and Language Identification in Degraded and Distorted Document Images," Proc. 21st Nat'l Conf. Artificial Intelligence, pp. 769-774, 2006.
29. Lu Shijian and Chew Lim Tan, "Script and Language Identification in Noisy and Degraded Document Images", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, No. 1, pp. 14-24, 2008.
30. Stefan Jaeger, Huanfeng Ma and David Doermann, "Identifying Script on Word-Level with Informational Confidence", 8th Int. Conf. on Document Analysis and Recognition, pages 416-420, August 2005.
31. Hiremath P S and S Shivashankar, "Texture classification using Wavelet Packet Decomposition", ICGSTs GVIP Journal, 6(2), 2006, pp. 77-80.
32. R.M.Haralick, K. Shanmugam and I.Dinstein, "Textural features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, Vol.3, pp. 610-621, 1973.

# Image Denoising Using Earth Mover's Distance and Local Histograms

**Nezamoddin N. Kachouie**

nezam.nk@gmail.com

*Department of Systems Design Engineering,  
University of Waterloo, Waterloo, ON, Canada*

*Present Affiliation: Harvard-MIT Health Sciences and Technology  
Harvard Medical School, Cambridge, MA, USA*

---

## Abstract

In this paper an adaptive range and domain filtering is presented. In the proposed method local histograms are computed to tune the range and domain extensions of bilateral filter. Noise histogram is estimated to measure the noise level at each pixel in the noisy image. The extensions of range and domain filters are determined based on pixel noise level. Experimental results show that the proposed method effectively removes the noise while preserves the details. The proposed method performs better than bilateral filter and restored test images have higher signal to noise ratio than those obtained by applying popular Bayesshrink wavelet denoising method.

**Keywords:** Denoising, Bilateral filtering, Local histogram, Earth mover's distance.

---

## 1. INTRODUCTION

Noise elimination is an important concern in image processing and computer vision. Images obtained from the real world are corrupted with noise. The image noise might decrease to some negligible levels under ideal conditions such that denoising is not necessary, but in general to recover the image the corrupting noise must be removed for practical purposes. Noise makes ambiguities in the underlying signal relative to its observed form by perturbations which are not related to the scene under study. The goal of denoising is to remove the noise and to retain the important signal features as much as possible. Linear filters, which consist of convolving the image with a constant matrix to obtain a linear combination of neighborhood values, have been widely used for noise elimination in the presence of additive noise. However they can produce a blurred and smoothed image with poor feature localization and incomplete noise suppression.

Gaussian filters are typical linear filters that have been widely used for image denoising. Gaussian filters assume that image signals have smooth spatial variations and pixels in a neighborhood have close values, so noise will be suppressed while signal will be preserved by averaging pixel values over a local neighborhood. The assumption of slow spatial signal variations works well in smooth regions; however it fails and undesirably blurs the signal where spatial variations are high such as edges.

To overcome this shortcoming and prevent undesirable blurring in regions with high spatial signal variations, a number of filters in spatial and spatial-frequency domain are proposed. The most popular ones in spatial domain are anisotropic diffusion [1-3], bilateral filtering and its extensions [4-8]. Diffusion based methods iteratively solve partial differential equations and average the

signal over spatial neighborhood whose extension is determined based on local signal variations. Bilateral filtering also known as range and domain filtering is a non-linear filter which performs weighted averaging in both range and domain.

Bilateral filtering was introduced by Tomasi and Manduchi [4] to smooth noisy images while preserve edges using neighboring pixels. Bilateral filtering is a local, nonlinear, and noniterative technique which considers both gray level (color) similarities and geometric closeness of the neighboring pixels. In a traditional domain filter, weight of the pixels decays by distance from the center of the filter. Low pass filters assume that spatial variations is slow over the image, so by weighted averaging of pixel values in a neighborhood, noise will be averaged away while the signal will be preserved. However, this assumption fails at edges where the spatial variations are not smooth and application of the low pass filter blurs the edges. Bilateral filter overcomes this by filtering the image in both range and domain. Pixels in a neighborhood are considered close either based on their spatial location (domain), or based on their pixel values (range). Therefore bilateral filter averages pixel values based on weights that decay by both distance and pixel dissimilarity.

There are several extensions to improve bilateral filtering [5-8]. In [6], a training-based bilateral filtering is proposed where a general degradation model is considered for degraded images. Then a restoration algorithm is developed to restore the degraded images with unknown degradation process. Therefore, the success of restoration process depends on the general definition of degradation model. In [7,8], different methods to speed up the bilateral filtering have been proposed.

Nonlinear filters in spatial-frequency domain have also been proposed to preserve detail signal and suppress the noise. The most popular ones are wavelet based denoising techniques [9-12]. In wavelet based denoising methods, the noise is estimated and wavelet coefficients are thresholded to separate signal and noise. Various approaches to nonlinear wavelet-based denoising have been introduced among them Bayesshrink wavelet denoising is developed in the Bayesian framework and has been widely used for image denoising [10-11].

In this paper, an adaptive technique is proposed to tune the extensions of range and domain filters. In the proposed method, the distance of the local histogram from the estimated noise histogram is measured using earth mover's distance. The measured distance at each spatial location is then used for adaptive tuning of bilateral filter. The proposed method provides promising results and effectively removes the noise while preserves the signal characteristics. The proposed method is presented in the next section followed by results and conclusions.

## 2. The Proposed Method

Let pure signal  $S$  (here an image) be distorted by additive noise  $n$ . We can write

$$I = S + n \cdot 1 \quad (1)$$

where  $I$  is the noisy signal. The goal of denoising is separating signal  $S$  and noise  $n$  by estimating  $n$  such that  $S$  can be extracted from  $I$ :

$$\hat{S} = I - \hat{n} \cdot 1 \quad (2)$$

This can be done by applying a filter  $h$  to the signal  $I$

$$\hat{S} = h * I \quad (3)$$

where traditionally  $h$  is defined as a local filter assigning higher weights to neighboring pixels which are spatially closer to the central pixel  $x_c$  of the neighborhood. A popular and simple case of  $h$  is Gaussian filter

$$h = h_d(x, \mu_d, \sigma_d) = \exp \left\{ -\frac{1}{2} \left( \frac{\|x - \mu_d\|}{\sigma_d} \right)^2 \right\} \quad (4)$$

where  $\mu_d = x_c$  is the central pixel of the neighborhood such that  $d(x, \mu_d) = \|x - \mu_d\|$  is the Euclidean distance between  $x_c$  and a neighboring pixel  $x$ . Gaussian domain filtering by using a Gaussian filter averages away noise and preserves the signal in smooth regions, however in the same way it averages away and blurs signal details such as edges. A popular solution to solve this problem is employing bilateral filter [4].

### Bilateral Filter

Bilateral filter combines range and domain filtering

$$h(x, \mu_d, \sigma_d, \mu_r, \sigma_r) = h_d(x, \mu_d, \sigma_d) h_r(x, \mu_r, \sigma_r) \quad (5)$$

where the range filter averages the signal values in a neighborhood by assigning the weights based on the similarity of the neighboring pixels and the central pixel:

$$h_r(x, \mu_r, \sigma_r) = \exp \left\{ -\frac{1}{2} \left( \frac{\|I(x) - \mu_r\|}{\sigma_r} \right)^2 \right\} \quad (6)$$

where  $\mu_r = I(\mu_d) = I(x_c)$  is the intensity value of the central pixel of the neighborhood such that  $r(I(x), \mu_r) = \|I(x) - \mu_r\|$  is the absolute intensity difference of the central and a neighboring pixel  $x$ .

Bilateral filtering overcomes the shortcomings of linear domain filtering by combining the linear domain filter with a nonlinear range filter. As a result bilateral filter preserves signal details such as edges while suppresses noise, however it considers fix parameters ( $\sigma_d, \sigma_r$ ) for extensions of both domain and range filters. The performance of bilateral filter can be improved by adaptively tuning the filter parameters over the image based on the spatial noise level.

### Adaptive Range and Domain Filtering

In the proposed method, to improve the performance of bilateral filtering, spatial noise level is locally estimated to determine the filter parameters ( $\sigma_d, \sigma_r$ ).

To estimate the local spatial noise level  $n_l$ , the image noise histogram  $n_g$  is estimated and compared with the local signal. To compare two probability density functions (PDFs), a number of nonparametric models have been used including minimizing the comparison  $\chi^2$  function between two PDFs.

The  $\chi^2$  distance between histograms of two delta functions  $\delta(x_1)$  and  $\delta(x_2)$  where  $x_1 \neq x_2$  is the same regardless of the distance between  $x_1$  and  $x_2$ . This is not generally suitable for many image processing applications where different smooth regions could be represented with disjoint  $\delta$  functions.

The earth mover's distance (EMD) or the Wasserstein distance is a mathematical measure to compare distributions (histograms). EMD was first introduced by Gaspard Monge in 1781, it was later used as a distance measure for intensity images [13]. The EMD between two distributions is

the least work that is required to move one distribution to another such that two distributions completely cover each other.

Let  $H_a$  and  $H_b$  be two normalized histograms with cumulative distributions  $C_a$  and  $C_b$  respectively. EMD between  $H_a$  and  $H_b$  is defined by

$$E(H_a, H_b) = \int_0^1 |C_a(k) - C_b(k)| dk \quad (7)$$

Local histogram for each pixel  $x$  in image  $I$  is computed over the neighborhood  $w$  consisting pixel  $x$  and its neighboring pixels. The EMD is then computed to compare the normalized local histogram  $H_x$  and image noise histogram  $n_g$

$$E(H_x, n_g) = \int_0^1 |C_x(k) - C_n(k)| dk \quad (8)$$

where  $C_x$  and  $C_n$  are cumulative distributions of  $H_x$  and  $n_g$  respectively. The extensions of domain and range filters ( $\sigma_d$  and  $\sigma_r$ ) at each pixel  $x$  are set using  $E(H_x, n_g)$ . The domain filter extension at pixel  $x$  is defined as

$$\sigma_d = \frac{(1 - E(H_x, n_g) + \epsilon_d)}{\sigma} \quad (9)$$

and we have

$$h_d(x, \mu_d, \sigma_d) = \exp \left\{ -\frac{1}{2} \left( \frac{\|x - \mu_d\|}{\left( \frac{1 - \int_0^1 |C_x(k) - C_n(k)| dk + \epsilon_d}{\sigma} \right)} \right)^2 \right\} \quad (10)$$

where  $E(H_x, n_g)$  is normalized EMD between noise and pixel histograms,  $\sigma$  is the filter extension parameter, and  $\epsilon_d$  is considered to avoid domain filter extension  $\sigma_d$  to be set to zero. The range filter extension also is tuned based on  $E(H_x, n_g)$

$$\sigma_r = (1 - E(H_x, n_g) + \epsilon_r) \cdot \sigma \quad (11)$$

and we have

$$h_r(x, \mu_r, \sigma_r) = \exp \left\{ -\frac{1}{2} \left( \frac{\|I(x) - \mu_r\|}{\left( \frac{1 - \int_0^1 |C_x(k) - C_n(k)| dk + \epsilon_r}{\sigma} \right)} \right)^2 \right\} \quad (12)$$

where  $\epsilon_r$  is considered to avoid range filter extension  $\sigma_r$  to be set to zero.

| Image     | Bilateral | Bayesshrink | The Proposed Method |
|-----------|-----------|-------------|---------------------|
| Lena      | 28.70     | 29.06       | 30.09               |
| Boat      | 27.86     | 27.87       | 28.30               |
| Cameraman | 29.56     | 28.63       | 30.03               |
| Barbara   | 27.31     | 26.75       | 27.39               |
| Goldhill  | 27.81     | 28.11       | 28.56               |

**TABLE 1:** Comparison of the proposed method with bilateral and Bayesshrink wavelet filtering methods.

Clearly there is a tradeoff here to choose the domain filter extension  $\sigma_d$ : as the filter extension  $\sigma_d$  expands the number of neighborhood elements grows, allowing for greater noise reduction in the computation but at the same time causing greater spatial blurring by fusion of values from more distant locations. Moreover, the range filter essentially compresses the image histogram by fusion of pixel values and is set by  $\sigma_r$ .

In the proposed method the maximum of  $\sigma_d$  and  $\sigma_r$  are set by  $\sigma$  based on equations (9) and (11). As  $0 \leq (1 - E) \leq 1$ , for pixels which are contaminated with high noise, the distance between noise and the pixel histograms  $E$  is small, therefore  $(1 - E)$  will be large. Considering  $\sigma$  is fixed,  $\sigma_d$  will be large allowing the neighborhood to be extended for greater noise reduction while  $\sigma_r$  will be large based on (11) to allow significant histogram compression.

On the other hand for pixels which are contaminated with low noise,  $E$  is large, therefore  $(1 - E)$  is small, and in turn  $\sigma_d$  will be small avoiding the neighborhood to be extended which in turn it allows less blurring. Considering that the pixel either is not contaminated with noise or is contaminated with low noise,  $\sigma_d$  will be small. Pixels have close values in small neighborhood, therefore  $\sigma_r$  will be small avoiding significant histogram compression.

**Noise Histogram Estimation**

To estimate the noise histogram ( $n_g$ ), the local variance is first computed. Considering the local neighborhood  $w$ , the local variance of pixel  $x$  is defined as:

$$\sigma_L^2(x) = \frac{\sum_w (I(x_w^i) - \mu_L(x))^2}{N_w} \tag{13}$$

where  $N_w$  is the number of pixels in the neighborhood  $w$  and

$$\mu_L(x) = \frac{\sum_w I(x_w^i)}{N_w} \tag{14}$$

Noise histogram  $n_g$  is estimated by computing the histogram of the local variance image. Further, we set  $\sigma = \sigma_n$  where the noise power  $\sigma_n^2$  is estimated by obtaining the mean of local variance image:

$$\sigma_n^2 = \frac{1}{M \cdot N} \sum_M \sum_N \sigma_L^2(x_{mn}) \tag{15}$$

Finally, the local histogram  $H_x$  is computed for each pixel  $x$  and EMD is used to measure the distance between noise histogram  $n_g$  and local histogram  $H_x$ . The schematic of noise histogram estimation is depicted in Fig. 1.

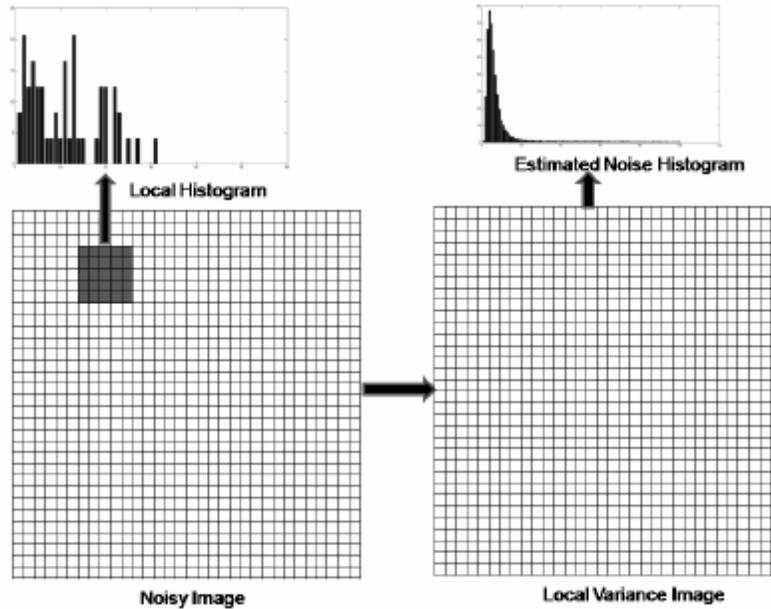


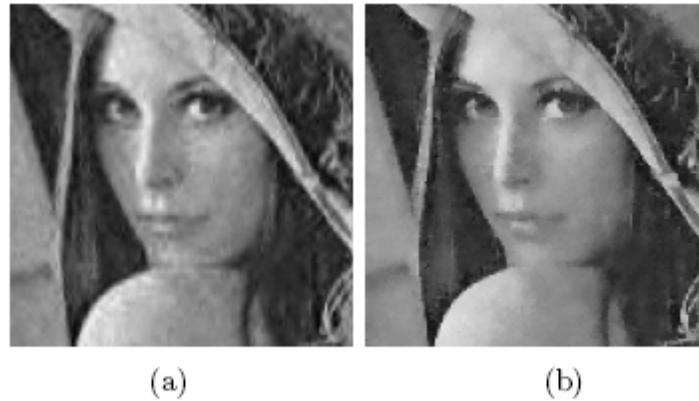
Figure 1: Noise histogram estimation

### 3. Results and CONSLUSION

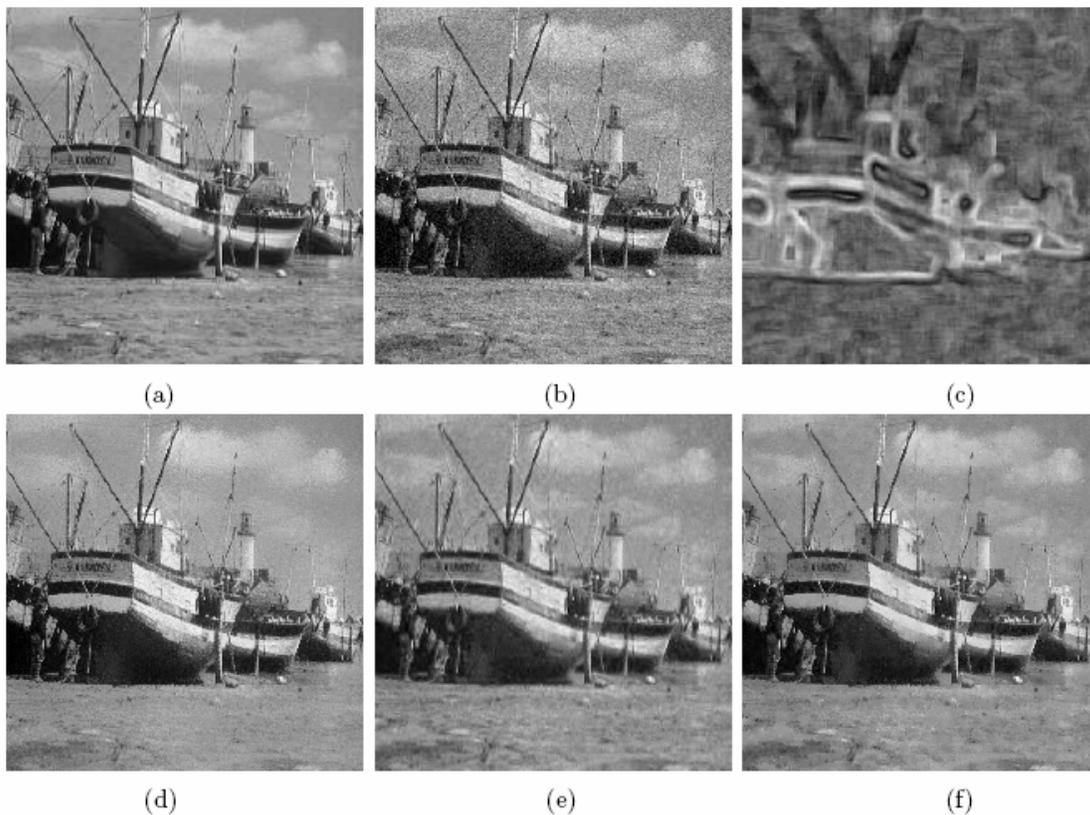
To test the proposed method five test images were used. Test images were corrupted by additive Gaussian noise with standard deviation of 15 and 25. The proposed method, the original bilateral filter, and a popular wavelet denoising method so called Bayesshrink wavelet denoising were applied to the corrupted test images. The recovered images applying aforementioned three methods were compared both based on PSNR and visual quality. The results are summarized in Tab. 1.

As we can observe in Tab. 1 for additive Gaussian noise with standard deviation of 15, the proposed method performs better than the original bilateral filtering method. It gains higher PSNR than both the original bilateral filtering and Bayesshrink wavelet denoising methods for all of the test images. The recovered images applying the proposed method have also better visual quality.

The recovered images applying Bayesshrink wavelet and the proposed method are depicted in Fig. 2. The proposed method performs better and the restored image gains higher PSNR (Tab. 1). It has also a better visual quality than that of Bayesshrink method which can be observed by a closer look. Fig. 3 shows the Boat noisy image and EMD computed for the noisy image. The denoised Boat image using the original bilateral filtering, Bayesshrink wavelet, and the proposed method are depicted in this figure.



**Figure 2:** Restored Lena test image: (a) Bayesshrink wavelet. (b) The proposed method.



**Figure 3:** Boat test image: (a) Original image. (b) Noisy. (c) EMD computed for (b). (d) Bilateral. (e) Bayesshrink wavelet. (f) The proposed method.

Fig. 4 shows the application of bilateral filter and the proposed method to the Cameraman test image where the image is corrupted with additive Gaussian noise with  $\sigma_n = 25$ . The application of Bilateral filtering and the proposed method to Goldhill test image which is corrupted with additive Gaussian noise with  $\sigma_n = 15$  is depicted in Fig. 5. Fig. 6 shows the comparison of the Bayesshrink, bilateral, and the proposed method where they are applied to the Lena test image corrupted with additive Gaussian noise with  $\sigma_n = 15$ . As we can observe in Fig. 4-6, the proposed

method performs better and produces smoother results while the details are better preserved in comparison with bilateral filter and the Bayesshrink. It also gains higher PSNR (Tab. 1).

In this paper an adaptive range and domain filtering method based on local histograms was introduced. The noise histogram is estimated and the extensions of range and domain filters are tuned at each spatial location by measuring the distance between the pixel's and noise histograms using earth mover's distance. The proposed method was applied to several test images and its performance was compared with the original bilateral filtering and Bayesshrink wavelet denoising methods. The experimental results obtained by the proposed method showed the improvement of the visual image quality and increase of PSNR in comparison with the bilateral filtering and Bayesshrink wavelet.



**Figure 4:** Cameraman test image: (a) Original image. (b) Noisy image (PSNR = 20.65). (c) Bilateral filtering (PSNR = 24.70). (d) The proposed method (PSNR = 26.00).



**Figure 5:** Goldhill test image: (a) Original image. (b) Noisy image (PSNR = 24.71). (c) Bilateral filtering (PSNR = 27.81). (d) The proposed method (PSNR = 28.56).



**Figure 6:** Lena test image: (a) Original image. (b) Bayesshrink (PSNR = 29.06). (c) Bilateral filtering (PSNR = 28.70). (d) The proposed method (PSNR = 30.09).

#### 4. REFERENCES

1. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", IEEE Tran. on PAMI, 12(7), pp. 629-639, 1990.
2. G. Sapiro and D. L. Ringach, "Anisotropic diffusion of color images", in Proc. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, 2657, pp. 471-482, 1996.
3. M. Ceccarelli, V. D. Simone, and A. Murli, "Well-posed anisotropic diffusion for image denoising", IEE Proc. on VISIP, 149(4), pp. 244-252, 2002.
4. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images", in Proceedings of Intl Conference on Computer Vision (ICCV), pp. 836-846, 1998.

5. J. Xie and P. A. Heng, "Color image diffusion using adaptive bilateral filter", in 27th Annual EMBS International Conference, pp. 3433-3436, 2005.
6. B. Zhang and J. P. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise Removal", IEEE Tran. on Image Processing, 17(5), pp. 664-678, 2008.
7. S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing Approach", MIT technical report, MIT-CSAIL-TR-2006-073, 2006.
8. M. Elad, "On the bilateral filter and ways to improve it", IEEE Tran. on Image Processing, 10(11), pp. 1141-1151, 2002.
9. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", Biometrika, 81(1), pp. 425-455, Sept 1994.
10. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and Compression", IEEE Tran. on Image Processing, 9(9), pp. 1532-1546, 2000.
11. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image designing", IEEE Tran. on Image Processing, 9(9), pp. 1522-1531, 2000.
12. M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation", IEEE Transactions on Image Processing, 12(1), pp. 16-28, 2003.
13. S. Peleg, M. Werman, and H. Rom, "A unified approach to the change of resolution: Space and gray-level", IEEE Tran. on PAMI, 11(7), pp. 739-742, 1989.

## Robust Digital Image-Adaptive Watermarking Using BSS Based Extraction Technique

**Sangeeta Jadhav**

*Asst Prof./E & TC  
AIT, Pune University  
Pune (MS), 411015, India*

djsangeeta@rediffmail.com

**Dr Anjali Bhalchandra**

*HOD & Prof/E & TC  
BAM University  
Aurangabad (MS), 431005, India*

asbhalchandra@yahoo.com

---

### Abstract

In a digital watermarking scheme, it is not convenient to carry the original image all the time in order to detect the owner's signature from the watermarked image. Moreover, for those applications that require different watermarks for different copies, it is preferred to utilize some kind of watermark-independent algorithm for extraction process i.e. dewatermarking. Watermark embedding is performed in the blue channel, as it is less sensitive to human visual system. This paper proposes a new color image watermarking method, which adopts Blind Source Separation (BSS) technique for watermark extraction. Single level Discrete Wavelet Transform (DWT) is used for embedding. The novelty of our scheme lies in determining the mixing matrix for BSS model during embedding. The determination of mixing matrix using Quasi-Newton's (BFGS) technique is based on texture analysis which uses energy content of the image. This makes our method image adaptive to embed the watermark into original image so as not to bring about a perceptible change in the marked image. An effort is also made to check feasibility of proposed method in device dependent color spaces viz. YIQ, YCbCr and HSI. BSS based on joint diagonalization of the time delayed covariance matrices algorithm is used for the extraction of watermark. The proposed method, undergoing different experiments, has shown its robustness against many attacks including rotation, low pass filtering, salt n paper noise addition and compression. The robustness evaluation is also carried out with respect to the spatial domain embedding.

**Keywords:** - DWT, BSS, BFGS, Mixing matrix, Attacks, Dewatermarking.

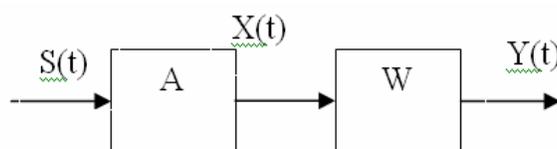
---

## 1. INTRODUCTION

With the development of network and multimedia techniques, data can now be distributed much faster and easier. Unfortunately, engineers still see immense technical challenges in discouraging unauthorized copying and distributing of electronic documents [1, 2]. Different kinds of

handwritten signatures, seals or watermarks have been used since ancient times as a way to identify the source or creator of document or picture. However, in digital world, digital technology for manipulating images has made it difficult to distinguish the visual truth. One potential solution for claiming the ownership is to use digital watermarks. A digital watermark is a transparent, invisible information pattern that is inserted into a suitable component of the data source by using a specific computer algorithm. In nature, the process of watermark embedding is the same as some special kind of patterns or under-written images are added into the host image, we can consider it as a mixture of host image and watermark, thus without host image, the watermark detection is equal to blind source separation in the receiver. Blind digital watermarking does not need the original images or video frames in the detection stage, thus it is the only feasible way to do watermarking in many multimedia applications, such as data monitoring or tracking on the internet, notification of copyright in playing DVD's. In particular some watermarking schemes require access to the 'published' watermarked signal that is the original signal just after adding the watermark. These schemes are referred as semi-blind watermarking schemes. Private watermarking [3] and non-blind-watermarking mean the same: the original cover signal is required during the detection process. The watermarked image is viewed as linear mixture of sources [4] i.e. original image and watermark and then we attempt to recover sources from their linear mixtures without resorting to any prior knowledge by using Blind Source Separation theory. Independent Component Analysis (ICA) is probably the most powerful and widely-used method for performing Blind Source Separation [15].

To present the basic principle of this new watermarking technique based on BSS, the paper is restricted to watermarking and dewatermarking with the simplest BSS model. The BSS model used to embed the watermark in the blue channel is shown below.



**FIGURE 1:** BSS model

The simplest BSS model assumes the existence of 'n' independent components i.e. the source signals  $S_1, S_2, S_3, \dots, S_n$   $[S(t)]$ , and the same number of linear and instantaneous mixtures  $X_1, X_2, \dots, X_n$   $[X(t)]$  of these sources. In vector matrix notation form the mixing matrix model can be represented as -

$$x = A * s \quad (1)$$

Where A is square (n x n) mixing matrix. W is separating matrix or demixing matrix and  $Y_1, Y_2, \dots, Y_n$   $[Y(t)]$  are estimated output sources which should be identical to sources represented by S(t).

Image watermarking techniques proposed so far can be categorized based on the domain used for watermarking embedding domain. The first class includes the spatial domain methods [9]. These embed the watermark by directly modifying the pixel values of the original image. The second contains Transform domain techniques, Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) [11], Discrete Cosine Transform (DCT) [10]. The third class is the feature domain technique, where region, boundary and object characteristics are taken into account [16]. The first class includes the works of Adib et al [4] have proposed to use the blue channel as the embedding medium. In [5] the authors have benefited from a new decomposition of the color images by the use of hyper complex numbers, namely the Quaternion and they achieved their watermarking/data-hiding operation on the component of the quaternion Fourier Transform. In the recent past, significant attention has been drawn to Blind source Separation by Independent Component Analysis [7,8] and has received increasing care in different image data applications such as image data compression, recognition, analysis etc. The technique of BSS

has been extended to the field of watermarking images [6, 12]. In [4], several assumptions are made regarding values of the mixing coefficients, distribution of the watermark as well as the mixing process. The proposed BSS based method is more flexible in the sense that the system finds out best suited mixing matrix using Quasi-Newton (BFGS) algorithm to keep the watermark hidden in the selected image. Watermark embedding in wavelet domain with adaptive mixing matrix makes the method quite difficult to extract unknown watermark using the existing simple methods, for example using [12].

The objective of this paper is to introduce an efficient digital image watermarking scheme based on BSS theory adopting watermark embedding in wavelet domain, which is more robust to the dewatermarking attacks as compared to the methods [10] in spatial domain embedding.

In the present work the effort has been also put to check the feasibility of proposed method in device dependent color spaces namely HSI, YIQ and YCbCr. In device dependent color spaces, color produced depends on parameters used as well as the equipment used for the display.

The RGB color space is highly correlated except of the blue channel because of its low sensitivity to human perception. The same set of embedding and detecting procedure is applied to all the color spaces so as to achieve the best comparison among them. The simulation results are shown for blue channel of RGB color space.

A BSS/ICA algorithm based on Joint diagonalization of the time delayed covariance matrices [13] is used for the extraction of watermark.

The paper is structured as follows: section 2 describes the proposed watermarking method including the watermark embedding using DWT and estimation of mixing matrix. The watermark extraction using BSS/ICA algorithm is also discussed in this section. The simulation results are illustrated in section 3. The robustness testing w. r. t. spatial domain embedding is analysed in section 4. Finally section 5 mentions conclusions and future work.

## 2. WATERMARKING SYSTEM

In the generic watermark embedding scheme, the inputs to the system are the original image and the watermark. To assure the identifiability of BSS model, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources [15].

### 2.1 The Watermark Embedding Scheme

In this paper the effective watermark embedding consists of mainly three phases. In first phase the blue channels of the host image and watermark image are extracted. The size of host image selected is  $512 \times 512$  ( $M \times M$ ) and size of watermark image is  $64 \times 64$  ( $N \times N$ ) so that  $M \gg N$ . In order to determine the sub-image of interest, the host image is divided into  $128 \times 128$  blocks and a sliding square window containing  $N_b$  number of such blocks in both the horizontal and vertical directions (a tentative sub-image) is considered. It has been shown that the energy of textured portion of image is high. Based on the energy content of the image, the two blue channel sub-images of size  $128 \times 128$ , one representing the smooth portion and other the textured one are taken out. In high textured area the visibility is low; therefore a textured sub-image is selected to embed the watermark. In second phase a single level DWT using haar wavelet function is applied to this textured sub-image and only the lowest frequency band (LL1 of size  $64 \times 64$ ) is selected for embedding the watermark (size  $64 \times 64$ ). To have as many mixtures as sources, the mixing matrix  $A$  is selected to be a square matrix (order  $2 \times 2$ ). The mixing operator 'A' has to be appropriately chosen such that the human vision can not determine that the message (watermark) is contained inside a host image. A Quasi-Newton (BFGS) algorithm [14] is used to estimate the mixing matrix  $A$  to keep the watermark hidden.

### 2.2 Estimation Of Mixing Matrix (Statistical Model)

In the proposed method the sources namely original image and watermark are known. The concept of correlation cancellation is used to estimate the mixing matrix  $A$  [13].

Consider two zero mean vector signals  $x(k) \in \mathbb{R}^m$  and  $s(k) \in \mathbb{R}^n$  that are related by the linear transformation

$$x(k) = As(k) + e(k)$$

where  $A \in \mathbb{R}^{m \times n}$  is unknown full rank mixing matrix and  $e(k) \in \mathbb{R}^m$  is a vector of zero mean error, interference or noise depending on application. Generally vectors  $s(k)$  and  $x(k)$  are correlated i.e.  $R_{xs} = E\{x s^T\} \neq 0$  but the error or noise  $e$  is uncorrelated with  $s$ , hence our objective is to find out the matrix  $A$  such that the new pair of vectors

$e = x - As$  and  $s$  are no longer correlated with each other and can be expressed in terms of equation as-

$$R_{es} = E\{e s^T\} = E\{(x - As)s^T\} = 0$$

The cross correlation matrix can be written as

$$R_{es} = E\{x s^T - A s s^T\} = R_{xs} - A R_{ss}$$

Hence the optimal mixing matrix can be expressed as

$$A_{opt} = R_{xs} R_{ss}^{-1} = E\{x s^T\} (E\{s s^T\})^{-1}$$

The same result is obtained by minimizing the mean square error cost function

$$\begin{aligned} J(e) &= \frac{1}{2} E\{e^T e\} = E\{(x - As)^T (x - As)\} \\ &= \frac{1}{2} (E\{x^T x\} - E\{s^T A^T x\} - E\{s^T A^T x\} - E\{x^T A s\} + E\{s^T A^T A s\}) \end{aligned}$$

By computing the gradient of the cost function  $J(e)$  w.r.t.  $A$  we obtain

$$\frac{\partial J(e)}{\partial A} = -E\{x s^T\} + A E\{s s^T\}$$

Hence applying the Quasi-Newton or BFGS approach, we obtain adaptive algorithm for the estimation of the mixing matrix to keep the watermark hidden in the host or original image.

$$\Delta A(k) = -V * \frac{\partial J(e)}{\partial A(k)} \tag{2}$$

$J(e)$  could be energy, entropy, homogeneity or inertia of the original image. In this paper energy of the original image is used.  $V$  is a system constant.  $\Delta A(k)$  = controlled rate of change of  $J$  w.r.t. mixing matrix  $A$ . Assuming the optimum value of  $A$  is achieved when gradient is zero.

$$A = \begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix} \tag{3}$$

Equation (2) becomes

$$\frac{\partial J(e)}{\partial A(k)} = \frac{\partial J}{\partial a11} + \frac{\partial J}{\partial a12} + \frac{\partial J}{\partial a21} + \frac{\partial J}{\partial a22}$$

But if we can simplify coefficients by certain assumptions

i.e. let  $a11 = a12 = 1$   
 $a12 = 1 - t$   
 $a22 = 1 + t$

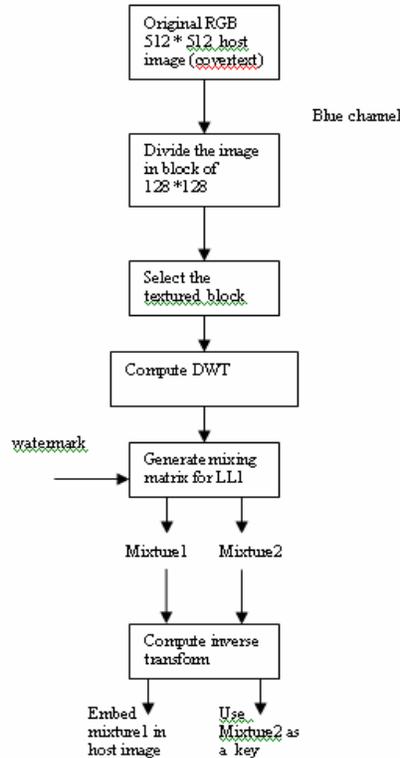
$$A = \begin{bmatrix} 1 & 1 - t \\ 1 & 1 + t \end{bmatrix}$$

Then instead of  $A$  we just have to check for 't' so the equation gets reduce to

$$\frac{\partial J(e)}{\partial A(k)} = \frac{dJ(e)}{dt} \quad [\text{since } dA=dt]$$

In BFGS algorithm the value of system constant  $V=1$  initially and as the process grows it gets updated and the value of 'A' gets converge easily.

In third phase, one of the compound sub-images (watermarked sub-image) is encrusted into the corresponding blocks of the earlier chosen region (by the BFGS algorithm) for the embedding, in the original image called watermarked image and is open to the public. The remaining watermarked textured sub-image is kept secret by the copyright owner. It will constitute the secret key corresponding to the location at which the watermark is fused with the original (host) image.



**FIGURE 2:** Flowchart showing the Watermark Embedding Process

As shown in Figure 2- Mixture1 and Mixture2 has following relationship- Equation (1) i.e.  $x=A*s$  can be written in matrix form as

$$\begin{bmatrix} \text{Mixture1} \\ \text{Mixture2} \end{bmatrix} = A * \begin{bmatrix} \text{Source1} \\ \text{Source2} \end{bmatrix}$$

From Equation (3)

$$\begin{aligned} \text{Mixture1} &= a_{11} * \text{Source1} + a_{12} * \text{Source2} \\ \text{Mixture2} &= a_{21} * \text{Source1} + a_{22} * \text{Source2} \end{aligned}$$

Thus a watermark embedding process is summarized in the following steps.

Step-1 Take the host and watermark color images, respectively of size (MxM) and (NxN) with  $M \gg N$ . Select their blue channels.

Step-2 Select textured regions block based on energy metric. Take one level DWT and use LL1 for further processing.

Step-3 Obtain the mixing matrix using Quasi-Newton (BFGS) algorithm [14] in order to keep watermark hidden in textured sub-image to form the watermarked mixtures. Take inverse wavelet of watermarked mixtures.

Step-4 One of the compound sub-images (watermarked sub-image) is encrusted into the corresponding blocks of the earlier chosen region of high energy in the original image. The other secret watermarked mixture in blue channels must be kept for a prospective use in the watermark extraction process.

### 2.3 Watermark Extraction

#### A) PCA whitening –watermark detection

Standard Principal Component Analysis (PCA) is often used for whitening process [12], since it can compress information optimally in the mean-squared error sense, while filtering possible noise simultaneously. The PCA whitening matrix is given by

$$V = D^{-1/2} U^T$$

Where D is the diagonal matrix of data covariance matrix  $E[X_i X_i^T]$  and U is its eigenmatrix, and  $E[.]$  denotes the expectation operator.

If the rank of D is equal to two for watermarked image, meaning that there are totally two image sources. On the other hand, if the image is unwatermarked image the rank D will be reduced to one.

After pre-whitening process, the sources are recovered by iteratively estimating the unmixing matrix W through a joint diagonalization of the time delayed covariance matrices algorithm [13]

As shown in Figure 3, the extraction process can be summarized in the following steps.

Step-1 Extract the marked block from the tampered watermarked image by using the first part of the key which is the position key.

Step-2 Obtain the blue channels of the extracted blocks.

Step-3 Apply a PCA whitening process on associated blue channel.

Step-4 Post processing has to be done on the whitened blue channel. Joint diagonalization algorithm is used to recover both host image and watermark.

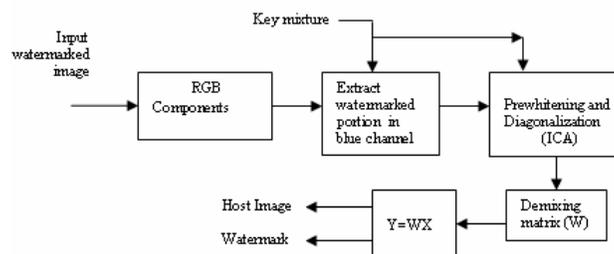


FIGURE 3: Watermark Extraction Using BSS

### 3 SIMULATION RESULTS

#### 3.1 Feasibility Of Proposed Method In Blue Channel Of RGB Color Space

Simulation experiments are conducted to demonstrate the feasibility and robustness of proposed BSS based watermark extraction method. Some results of DWT embedding are given below



FIGURE 4: Original Image (a) and Watermark (b)

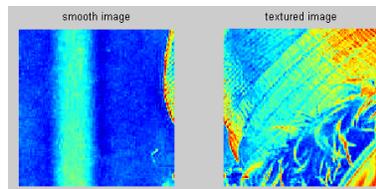


FIGURE 5: Smooth and Textured Portions of Original Image

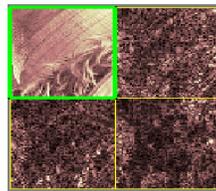


FIGURE 6: DWT of Textured Sub-image

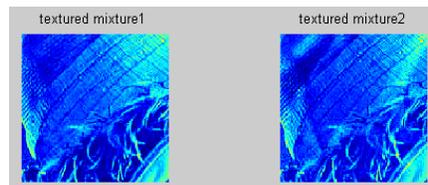
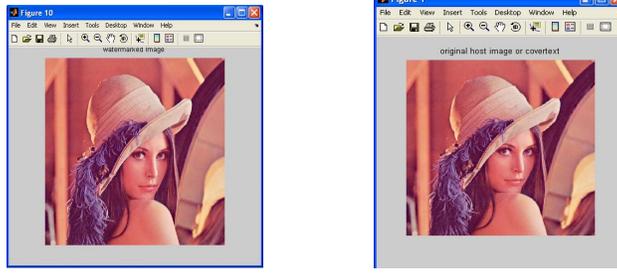
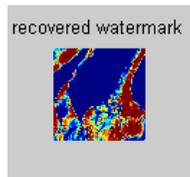


FIGURE 7 :Watermarked Mixture Sub-images



(a) (b)

**FIGURE 8 :**Watermarked Image (a) Original Host Image or Covertext (b)



**FIGURE 9:** Recovered Watermark

### 3.2 Feasibility Of Proposed Method In Device Dependent Color Spaces

The proposed method of watermarking is tested over device dependent color spaces mentioned in Table 1. The Table 1 shows the value of PSNR and Correlation Coefficient computed by Equation (5) and (6) for recovered watermark using spatial domain embedding. The value of mixing matrix generated using BFGS method is also mentioned in the table.

| Sr No. | Color Space    | PSNR  | Correlation Coefficient | Mixing Matrix         |
|--------|----------------|-------|-------------------------|-----------------------|
| 1      | YIQ<br>[...]   | 73.11 | 0.86                    | 1 0.31<br>1 -0.1      |
| 2      | YCbCr          | 26.07 | 0.80                    | 1 0.0023              |
|        | (I) Y          |       |                         | 1 -0.001              |
|        | (II) Cb        | 25.6  | -0.84                   | 1 0.0013<br>1 -0.0013 |
| 3      | (III) Cr       | 25.85 | 0.96                    | 1 0.001<br>1 -0.0022  |
|        | HSI/HSV        | Inf   | 0.53                    | 1 0.864               |
|        | (I) H          |       |                         | 1 -0.883              |
| (II) S | 54.98          |       |                         | 0.61                  |
|        | (III) I(Value) | 53.54 | 0.285                   | 1 0.474<br>1 -0.139   |

**TABLE 1:** Embedding in Spatial Domain Without Attack

In Table 2, the performance parameters PSNR and Correlation Coefficient computed for recovered watermark using BSS extraction technique; with DWT domain embedding is shown. The value of mixing matrix is also shown in the table.

| Sr No. | Color Space          | PSNR  | Correlation Coefficient | Mixing Matrix |
|--------|----------------------|-------|-------------------------|---------------|
| 1      | YIQ<br>(I) [r, g, b] | 68.34 | 0.87                    | 1 0.0017      |
|        | 1 -0.34              |       |                         |               |
| 2      | YCbCr<br>(I) Y       | 26.28 | 0.89                    | 1 0.001       |
|        | (II) Cb              | 26.03 | 0.79                    | 1 0.0013      |
|        | (III) Cr             | 25.80 | 0.96                    | 1 0.0013      |
|        | 1 -0.0035            |       |                         |               |
| 3      | HSI/HSV<br>(I) H     | 66.54 | 0.73                    | 1 0.128       |
|        | (II) S               | 55.25 | 0.612                   | 1 -0.907      |
|        | (III) I(Value)       | 53.56 | 0.29                    | 1 0.0012      |
|        |                      |       |                         | 1 -0.26       |

TABLE 2: Embedding in DWT Domain Without Attack

#### 4 ROBUSTNESS TESTING WITH RESPECT TO SPATIAL DOMAIN EMBEDDING

The watermarking system should be robust against data distortions introduced through standard data processing and attacks. It should be virtually impossible for unauthorized users to remove it and practically the image quality must be degraded before the watermark is lost. There are many attacks against which image watermarking system could be judged. The attacks include average filtering, rotation (+90°), median filtering, Salt n Paper noise and so on. These various attacks are applied to the watermarked images to evaluate whether the proposed dewatermarking system can recover the embedded watermark, thus measuring the robustness of the watermarking system to these types of attacks.

Mean Square Error (MSE), PSNR (Peak signal to Noise Ratio) and NC (Normalized Cross-Correlation) are used to estimate the quality of extracted watermark.

The equations used are defined as below-

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [r(i, j) - r^*(i, j)]^2 \tag{4}$$

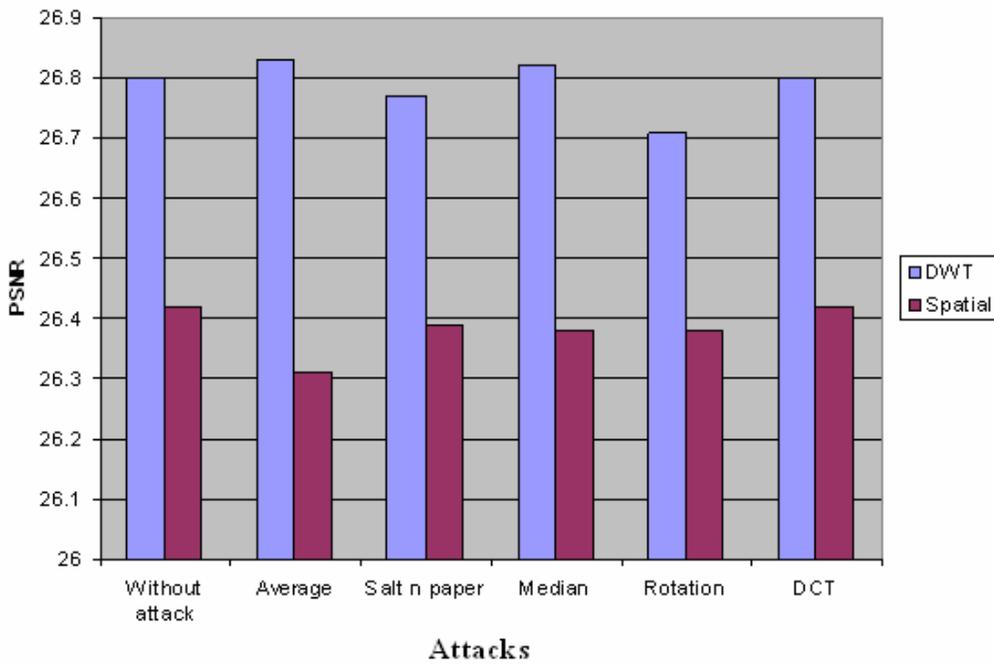
Where r(i,j) represents pixel at location (i,j) of the original watermark and r\*(i,j) represents the pixel at location (i,j) of recovered watermark. M,N denotes the size of the pixel.

$$PSNR = 10 \log_{10} \left[ \frac{255^2}{MSE} \right] \tag{5}$$

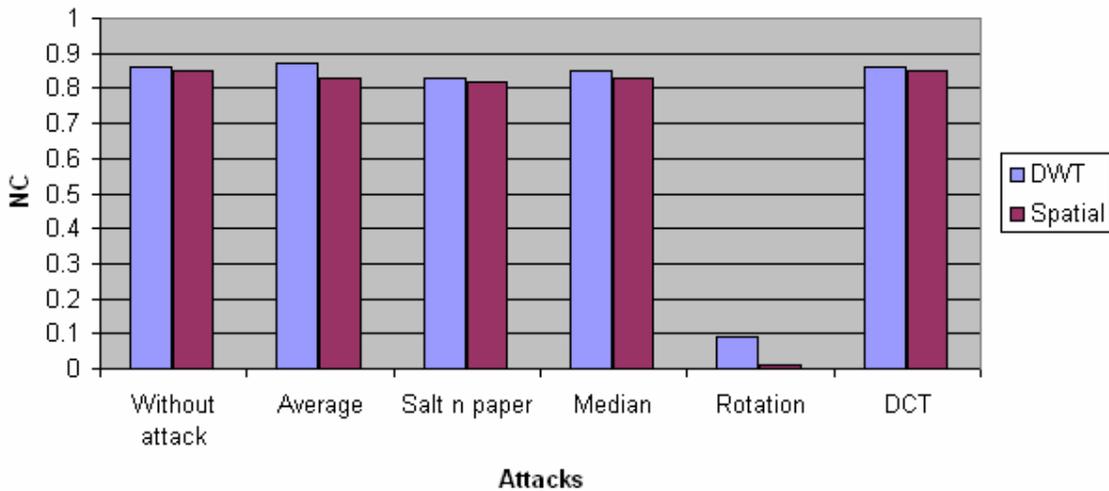
$$NC = \frac{\sum_{m=1}^M \sum_{n=1}^N W * W'}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N W^2} \times \sqrt{\sum_{m=1}^M \sum_{n=1}^N W'^2}} \tag{6}$$

Where W is original watermark and W' is recovered watermark with zero mean value each.

As shown in Figure 10, the PSNR in dB is calculated by using Equation 5 and compared for both the types of embedding viz. Spatial domain and DWT domain embedding. It is observed that the PSNR obtained in DWT embedding is high under various attack conditions. In Figure 11, the correlation coefficients (NC) comparison is shown.



**FIGURE 10:** PSNR( dB) comparison for DWT based and spatial watermark embedding



**FIGURE 11:** Normalized Correlation Coefficient (NC) Comparison

JPEG Quality variation is tested for DWT based and spatial based embedding and plotted against PSNR of recovered watermark.

As shown in graph, it is observed that the PSNR values are more in case of DWT based embedding as compared to the Spatial based embedding proving that DWT based embedding is more robust against the watermarking attacks.

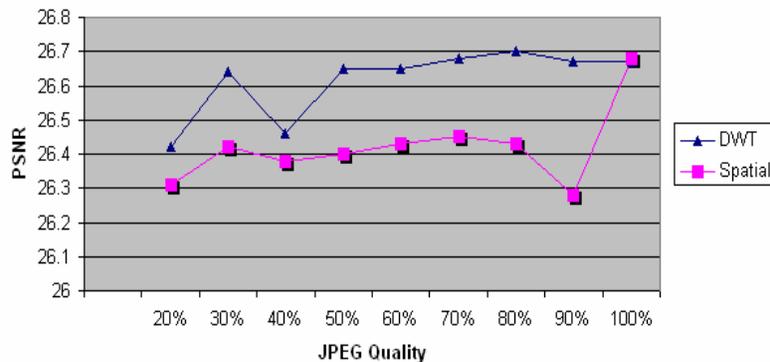


FIGURE 12: PSNR (dB) comparison for JPEG Quality variation

## 5 CONSLUSION & FUTURE WORK

In this paper, we proposed a digital color image watermarking system using wavelet(DWT) domain embedding and adopting Blind Source Separation theory along with RGB decomposition to extract watermark. The novelty of our scheme lies in determining the mixing matrix for BSS model, based on energy content of the image using Quasi Newton (BFGS) method. This makes our method image adaptive to embed any image watermark into original host image. The effort has been also put to check the feasibility of proposed method in device dependent color spaces for the application of image watermarking. The watermark is readily detected by Principal Component Analysis(PCA) whitening process. The watermark is further separated by using BSS/ICA algorithm based on Joint diagonalization of the time delayed covariance matrices. The performance of the proposed method can be evaluated in terms of normalized correlation coefficient and PSNR with respect to spatial domain watermark embedding. Experimental results demonstrate the proposed watermarking scheme is more robust to various attacks as compared to spatial domain watermark embedding.

In future research work, it is proposed to implement the watermark extraction process in time-frequency domain using DWT in order to improve the performance for different types of images as well as to make the proposed watermarking scheme more robust against various attacks.

## 6 REFERNCES

- [1] Peter H W Wong, Oscar C Au and YM Yeung, "A novel blind multiple watermarking technique for Images", in IEEE transactions on circuits and ystems for video technology, vol.13, No. 8,pp 813 -830,yr 2003
- [2] Chiou-Ting Hsu, Ja-Ling Wu, "Hidden Digital Watermarks in Images," in IEEE transaction on image processing, vol.8 No.1, pp 58-68,Jan1999.
- [3] Ibrahim Nasir et al. " a New Robust Watermarking Scheme for Color Image in spatial Domain", [ieeexplore.ieee.org/iel5/4618740/4618741/04618875.pdf](http://ieeexplore.ieee.org/iel5/4618740/4618741/04618875.pdf).
- [4] Moilim Amir, Abdellah Adib et al, "Color Images Watermarking by Means of Independent Component Analysis",in IEEE transaction pp 347-350,yr-2007.
- [5] Tsz Kin Tsui et al, " Quaternion Image Watermarking using the Spatio-Chromatic Fourier Coefficients Analysis", in ACM 1-59593-447-2/06/0010, yr-2006.
- [6] Thang V. Nguyen et al., "A Novel Digital Image Watermarking Scheme Using Blind Source Separation", Proc of IEEE, ICIP,pp2649-2652,Yr 2004.
- [7] Yang Su -min,Wang Jia-zhen,ZHANG Zheng-bao Ding Guo-liang, "A digital watermark based on blind signal separation algorithm," in ICIP 2006 proceedings.
- [8] Thang Viet et al. "A Simple ICA -based Digital Image Watermarking Scheme",DSP archive ,Volume 18,issue 5,yr-2008.

- [9] Peter H W Wong, Oscar C Au and YM Yeung, "A novel blind multiple watermarking technique for Images", in IEEE transactions on circuits and systems for video technology, vol.13, No. 8,pp 813 - 830,yr 2003
- [10] Chiou-Ting Hsu, Ja-Ling Wu, "Hidden Digital Watermarks in Images," in IEEE transaction on image processing, vol.8 No.1, pp 58-68,Jan1999.
- [11] Shang-Lin Hsieh et al , "a Color Image Watermarking Scheme Based on Secret Sharing and Wavelet Transform" ,IEEE,pp-2143-2147,Yr 2007.
- [12] Dan Yu ,Farook Sattar et al , "Watermark Detection and Extraction using Independent component analysis Method",EURASIP Journal on Applied Signal Processing 2002:1,92-104
- [13] A.Cichoki,Shun-ichi Amari , "Adaptive Blind Signal and Image Processing –Learning Algorithms and applications", Oct-2005,pp 0-175.
- [14] L F Escudero, "On Diagonally preconditioning the 2-Steps BFGS Method With Accumulated Steps for Supra Scale Linearly Constrained Non Linear Programming"published in QUESTIO-1982,vol-06 no-04, Yr 1982.
- [15] A Hyvarinen and E. Oja , " Independent Analysis a Tutorial", <http://www.cis.hut.fi/projects/ica/April99>.
- [16] SUN Jiande et al. "A Blind Video Watermarking Scheme based on ICA and Shot Segmentation",Science in China Series F Information Sciences,vol-49, No-3,pp302-312 Yr 2006.

# CALL FOR PAPERS

**Journal:** International Journal of Image Processing (IJIP)

**Volume:** 4 **Issue:** 1

**ISSN:** 1985-2304

**URL:** <http://www.cscjournals.org/csc/description.php?JCode=IJIP>

## About IJIP

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.

## IJIP 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
- 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

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

## **CFP SCHEDULE**

**Volume:** 4

**Issue:** 2

**Paper Submission:** March 31 2010

**Author Notification:** May 1 2010

**Issue Publication:** May 2010

## CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for ***International Journal of Image Processing (IJIP)***. CSC Journals would like to invite interested candidates to join **IJIP** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at <http://www.cscjournals.org/csc/byjournal.php>. Interested candidates may apply for the following positions through <http://www.cscjournals.org/csc/login.php>.

*Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.*

Feel free to contact us at [coordinator@cscjournals.org](mailto:coordinator@cscjournals.org) if you have any queries.

## **Contact Information**

### **Computer Science Journals Sdn Bhd**

M-3-19, Plaza Damas Sri Hartamas  
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607  
          +603 2782 6991  
Fax:     +603 6207 1697

### **BRANCH OFFICE 1**

Suite 5.04 Level 5, 365 Little Collins Street,  
MELBOURNE 3000, Victoria, AUSTRALIA

Fax: +613 8677 1132

### **BRANCH OFFICE 2**

Office no. 8, Saad Arcad, DHA Main Bulevard  
Lahore, PAKISTAN

### **EMAIL SUPPORT**

Head CSC Press: [coordinator@cscjournals.org](mailto:coordinator@cscjournals.org)  
CSC Press: [cscpress@cscjournals.org](mailto:cscpress@cscjournals.org)  
Info: [info@cscjournals.org](mailto:info@cscjournals.org)



COMPUTER SCIENCE JOURNALS SDN BHD  
M-3-19, PLAZA DAMAS  
SRI HARTAMAS  
50480, KUALA LUMPUR  
MALAYSIA