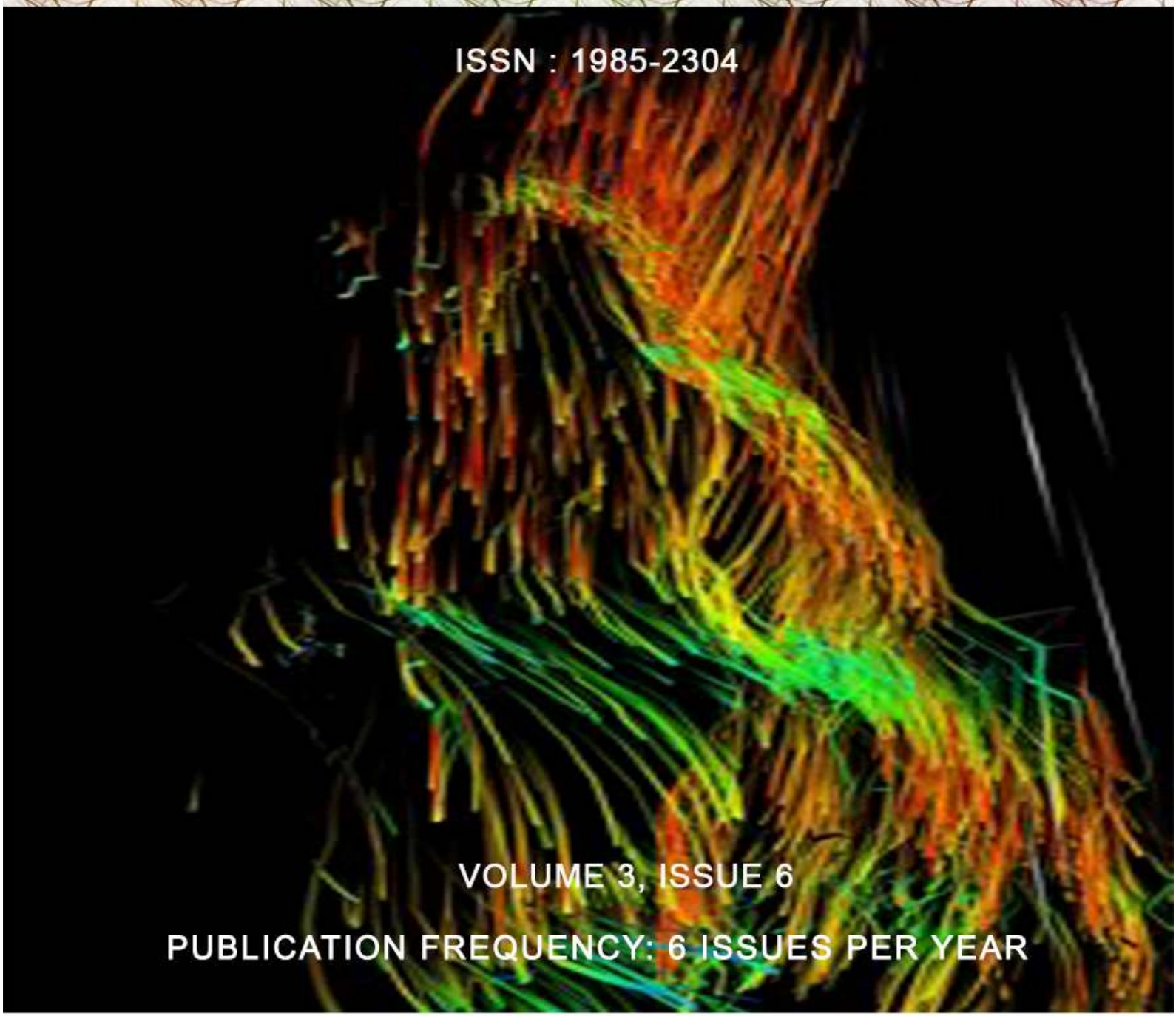
International Journal of Image Processing (IJIP)

ISSN : 1985-2304



VOLUME 3, ISSUE 6

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

International Journal of Image Processing (IJIP)

Volume 3, Issue 6, 2010

Edited By
Computer Science Journals
www.cscjournals.org

International Journal of Image Processing (IJIP)

Book: 2010 Volume 3 Issue 6
Publishing Date: 31-01-2010
Proceedings
ISSN (Online): 1985-2304

IJIP Journal is a part of CSC Publishers
http://www.cscjournals.org

Published in Malaysia

Typesetting: Camera-ready by author, data conversation 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 sixth issue of volume three 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)

Associate Editors (AEiCs)

Editorial Board Members (EBMs)

Table of Content

Volume 3, Issue 6, January 2010

Pages

Face Hallucination using Eigen Transformation in Transform Domain

Abdu Rahiman V vkarahim@gmail.com
Department of Electronics and Communication
Government College of Engineering,
Kannur, Kerala, India

Jiji C. V. jiji@ee.iitb.ac.in
Department of Electronics and Communication
College of Engineering, Trivandrum
Kerala, India

Abstract

Faces often appear very small in surveillance imagery because of the wide fields of view that are typically used and the relatively large distance between the cameras and the scene. In applications like face recognition, face detection etc. resolution enhancement techniques are therefore generally essential. Super resolution is the process of determining and adding missing high frequency information in the image to improve the resolution. It is highly useful in the areas of recognition, identification, compression, etc. Face hallucination is a subset of super resolution. This work is intended to enhance the visual quality and resolution of a facial image. It focuses on the Eigen transform based face super resolution techniques in transform domain. Advantage of Eigen transformation based technique is that, it does not require iterative optimization techniques and hence comparatively faster. Eigen transform is performed in wavelet transform and discrete cosine transform domains and the results are presented. The results establish the fact that the Eigen transform is efficient in transform domain also and thus it can be directly applied with slight modifications on the compressed images.

Keywords: Face hallucination, Super resolution, Eigen transformation, wavelets, discrete cosine transform.

1. INTRODUCTION

In most electronic imaging applications, images with high spatial resolution are desired. High resolution (HR) means that pixel density within an image is high, and therefore an HR image can offer more details than its low resolution counter part. The performance of face recognition or detection in computer vision can be improved if an HR image is provided. The direct solution to increase spatial resolution is to reduce the pixel size by sensor manufacturing techniques. As the pixel size decreases, however, the amount of light available also decreases, which generates shot noise that degrades the image quality severely. Also to reduce the pixel size, there exists a minimum limit, which is already achieved. An alternate approach is to use signal processing techniques to obtain an HR image from one or more low-resolution (LR) images. Recently, such a

resolution enhancement approach has been one of the most active research areas, and it is called super resolution (SR) image reconstruction or simply Super resolution. The major advantage of the signal processing approach to improve resolution is that it is less costly and the existing LR imaging systems can be still utilized. The SR image reconstruction has wide fields of application like medical imaging, Synthetic zooming, forensics, satellite imaging and video applications. Another application is in the conversion of an NTSC video signal to HDTV format.

There are two different types of super resolution approaches. In the first type, more than one low resolution images are used to produce a high resolution image. It is generally called multi-frame super resolution. In multi-frame super resolution, HR image is synthesized from input images alone, so these are also called reconstruction based super resolution. Another type of super resolution uses a single low resolution image as the input to produce a high resolution image. This method is called single frame super resolution. Most of the single frame image super-resolution algorithms use a training set of HR images and the additional details of the HR image to be synthesized is learnt from these HR training set. Such algorithms are called learning based super resolution algorithms.

The simplest signal processing technique to increase resolution is the direct interpolation of input images using techniques such as nearest neighbor, cubic spline, etc. But it does not add any extra information to the image. Also its performance become poor if the input image is too small in size.

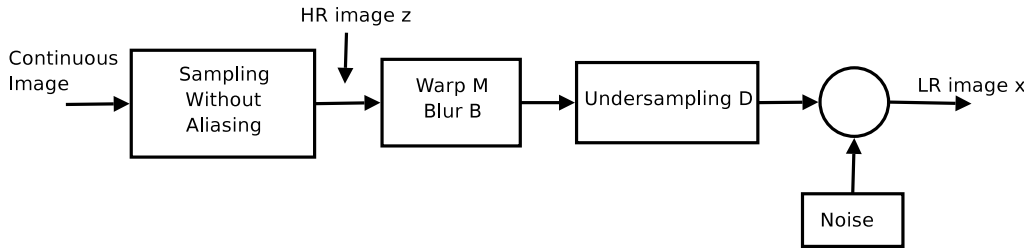


FIGURE 1: A digital low resolution image acquisition model

1.1 LR Image Formation Model

A LR image formation model is shown in Figure 1, the observed LR images result from warping, blurring, and subsampling operations performed on the HR image ***z***. Assuming that LR image is corrupted by additive noise, we can then represent the observation model as

$$x = DBMz + \eta \quad (1)$$

Where ***D*** is the decimation or subsampling matrix, ***B*** is the blur matrix and ***η*** represents noise vector. The motion that occurs during the image acquisition is represented by warp matrix ***M***. It may contain global or local translations, rotations and so on. Blurring may be caused by an optical system (e.g., out of focus, diffraction limit, aberration, etc.) and the point spread function (PSF) of the LR sensor. Its effects on HR images are represented by the matrix ***B***. The subsampling matrix D generates aliased LR images from the warped and blurred HR image.

Face hallucination, the term coined by Baker and Kanade [1] is the super resolution of face image, which is the process of synthesizing a high resolution face image from low resolution observation. Figure 2 shows the schematic of face hallucination algorithm. Face hallucination techniques can be useful in surveillance systems where the resolution of a face image is normally low in video, but the details of facial features which can be found in an HR image may be crucial for identification and further analysis. The standard super resolution techniques may also introduce some unwanted high frequency components. However, hallucinating faces is more challenging because people are so familiar with the face image. This specialized perception of faces requires that a face synthesis system be accurate at representing facial features and the process should not introduce many unwanted details. A small error, e.g. an asymmetry of the

eyes, might be significant to human perception [17], whereas for super resolution of generic images the errors in textured regions, e.g. leaves, grasses, etc. are often ignored.

This work studies the feasibility of Eigen transformation based super resolution in transform domain for synthesizing high resolution face images. Advantage of Eigen transformation based technique is that, it does not require iterative optimization, which considerably reduces the processing time. Eigen transform is performed in wavelet transform and discrete cosine transform domains and the results are presented. This work establishes the fact that the Eigen transform is efficient in transform domain and thus it can be directly applied to the compressed images.

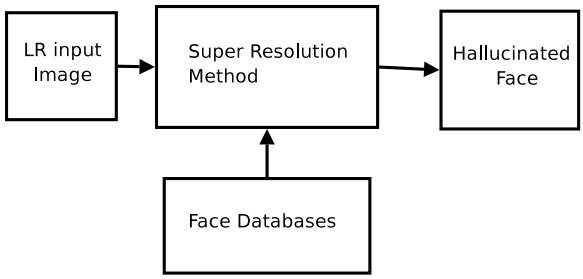


FIGURE 2: block schematic of face hallucination system

2. RELATED WORKS

In the paper "Hallucinating Faces", Kanade and Baker introduced the term face hallucination for super resolution of face image [1][2]. They use a single LR observation to synthesize an HR face image, making use of a training set of HR face images. High frequency details of the HR face image are learned by identifying local features from the training set. In the above hallucination approach, a Gaussian image pyramid is formed for every image in the training set as well as for the LR observation. A set of features are computed for every image in the image pyramid resulting in a feature database. The feature vector for hallucinated image is learned from these feature vector database. Hallucinated face is estimated using maximum a posteriori (MAP) framework, which uses leaned prior in its cost function. The final gray level image is then obtained by gradient descent optimization to fit the constraints learned from the facial features. So here the high frequency part of the face image is purely fabricated by learning the properties from the similar HR images.

The images hallucinated by Baker and Kanade appear to be noisy at places, especially where the test image and training set images have significantly different features. As the magnification increases the noise increases as well. Liu, Shum and Zang [3] argued that face hallucination algorithms should consider the following constraints.

i. The result must be very close to the input image when smoothed and down sampled.
ii. The result must have common characteristics of human face, *e.g* eyes, mouth, nose, symmetry, etc.
iii. The result must have specific characteristics of this face image with realistic local features.

Baker and Kanade considered the first condition but not focused on the next two. Liu, Shum and Zhang [3] proposed a two step approach to take the above constraints into account. It is done by incorporating the local features of face image as well as the global structure. Local features are learned by using a patch based method. Global structure of the face is determined by learning principal component coefficients. Locally learned patches are then combined with the global face structure to give the hallucinated face image.

Capel and Zisserman proposed a principal component analysis (PCA) based learning method for face image super resolution [6]. A collection of registered face images are used as the training set and the image is modeled using PCA basis computed from these training images. In this method the face image is divided in to six regions or subspaces. The intuition here is that these regions are relatively uncorrelated and that, by considering small regions, better models can be

learnt than would be by performing PCA on the whole face. Each of the subimages is separately super resolved using PCA based learning method. The reconstructed subspaces are combined to give hallucinated face.

Jiji et al. [8] proposed a wavelet based single frame super resolution method for the super resolution of general images. It makes use of a training set consist of HR images from different categories. In this method, observed image as well as the images in data base are decomposed using wavelet transform. Wavelet coefficients of the super resolved image are learned from the coefficients of images in the database. The HR image is estimated under a MAP frame work using the learned wavelet prior. An edge preserving smoothness constraint is used to maintain the continuity of edges in the super resolved image. This method is applied on face images, but as it is formulated for general images, it does not consider the structural properties of face image and therefore the results are not good for higher magnification factors.

A promising approach for face hallucination is proposed by Wang and Tang [13][14]. This method is based on the face recognition using Eigen faces [11] and it is computationally much efficient than any previous algorithms. It does not require iterative optimization techniques. A registered HR face image training set is used here and a corresponding LR training set is prepared by down sampling the HR images. If the blur matrix is known, it can be incorporated by filtering the HR images with blur matrix, before down-sampling it to produce LR training set. LR observation image is then represented as the linear combination of LR database images. The linear combination coefficients are determined from the PCA coefficients. The super resolution is achieved by finding the linear combination of the HR images with the same coefficients. To avoid abnormalities in the image, regularization is done with respect to Eigen values. Besides other methods, Eigen transformation based method give better results even with higher magnification factors.

3. EIGEN TRANSFORMATION BASED SUPER RESOLUTION

Eigen transformation (ET) based super resolution makes use of a registered set of HR face images as training set. PCA models are formulated for HR and LR image space using the respective training sets. This section discusses the PCA in brief followed by super resolution using Eigen transformation.

3.1 Principal Component Analysis

PCA is a powerful tool for analyzing data by performing dimensionality reduction in which the original data or image is projected on to a lower dimensional space. An image in a collection of images can be represented as the linear combination of some basis images. Let there be M images with N pixels each, in a collection, all images in the collection are arranged into column vectors by scanning them in raster scan order. Let $\boldsymbol{x_i}$ be the individual image vectors and $\overline{x}$ be the mean image vector, and then the mean removed image is given by

$$L_i = x_i - \overline{x} \tag{2}$$

All the mean removed images are arranged in columns to form the matrix $\boldsymbol{L = [l_1, l_2, \ldots, l_M]}$ Covariance matrix of $\boldsymbol{L}$ can be found as

$$C = L \times L^T \tag{3}$$

Let $\boldsymbol{E}$ be the matrix of Eigen vectors of the matrix $\boldsymbol{C}$ and $\boldsymbol{S}$ be the Eigen values. The Eigen vectors are arranged in such a way that respective Eigen values are in decreasing order. A given image can be projected on to these Eigen vectors and the coefficients $\boldsymbol{w}$ thus obtained are called PCA coefficients.

$$w = E^T \times (x_i - \overline{x}) = E^T \times l_i \tag{4}$$

The mean removed image can be reconstructed as $\hat{l}_i = E \times w$. Adding the mean image vector to $\hat{l}_i$ gives the actual image vector. In the discussions followed, image is considered as the mean removed image unless otherwise mentioned. An important fact about PCA coefficients is that the image can be reconstructed with minimum mean square error, using only the first few coefficients.

3.2 Super resolution with Eigen Transformation

Here we discuss the use of PCA for super resolution. First we determine the significant Eigen vectors of ***C*** as described in [11]. Define the matrix $K = L^T \times L$. Let ***Λ*** be the diagonal matrix consisting of the Eigen values and ***V*** is the matrix containing Eigen vectors of ***K***. Most significant ***M*** Eigen vectors of ***C*** can be determined by

$$E_M = L \times V \times \Lambda^{-1/2} \tag{5}$$

M significant PCA coefficients of l_i can be found by projecting it on to E_M, ie. $w_l = E_M^T \times l_i$. The reconstructed image $\hat{l}$ is then obtained as

$$\hat{l} = E_M \times w_l = L \times c \tag{6}$$

where

$$c = V \times \Lambda^{-1/2} \times w_l \tag{7}$$

In the super resolution process, we use databases of registered HR images and corresponding LR images. Let ***H*** be the matrix of mean removed image vectors of HR images in database, corresponding to the matrix ***L*** discussed above. The given LR image is represented as the linear combination of the image vectors as shown in equation (6). Hallucinated face image can be determined by using the same coefficients but by using the HR image vectors ***H***

$$h_{SR} = H \times c \tag{8}$$

where h_{SR} is the hallucinated face image. It means that if LR image is the linear combination of image vectors in the LR face images, then the corresponding HR image will be linear combination of the respective HR image vectors while keeping the same coefficients. If the test image is a very low resolution image, then the hallucinated image will have lot of artifacts. We minimize these artifacts by applying a constraint based on the Eigen values. Let ***Q*** be the resolution enhancement factor and ***α*** be a positive constant. To apply the constraint, PCA coefficients w_h of the super resolved image is found. Let E_h be the Eigen vectors of HR image space, then constrained PCA coefficients, $\hat{w}_h(i)$ of the i^{th} eigen vector is given by

$$\hat{w}_h(i) = \begin{cases} w_h(i) & \text{if } |w_h(i)| < \lambda_i^{1/2} \alpha / Q^2 \\ sign(w_h(i)) \lambda_i^{1/2} & \text{otherwise} \end{cases} \tag{9}$$

where the λ_i are the eigen values corresponding to E_h. These new coefficients, $\hat{w}_h$ is used to reconstruct the super resolved images from HR eigen vectors. Super-resolved image x_h is given by

$$x_h = E_h \times \hat{w}_h + \overline{x}_h \tag{10}$$

$\overline{x}_h$ is the mean of HR images in the database. As the value of ***α*** increase, super resolved image may have more high frequency details. This may introduce spurious high frequency components also. On the other hand, when ***α*** is reduced, the super resolved image tends towards mean face image.

4. EIGEN TRANSFORMATION IN WAVELET TRANSFORM DOMAIN

In this section we discuss the use of Eigen transformation in the wavelet transform domain for face hallucination.

4.1 Discrete Wavelet Transform

Wavelets are functions defined over a finite interval and are used in representing data or other functions. The basis functions used are obtained from a single mother wavelet, by dilations or contractions and translations. The Discrete Wavelet Transform (DWT) is used with discrete signals. Wavelet coefficients of an image are determined using filters arranged as shown in Figure 3. ***g(n)*** and ***h(n)*** are the half band high pass and low pass filters respectively.

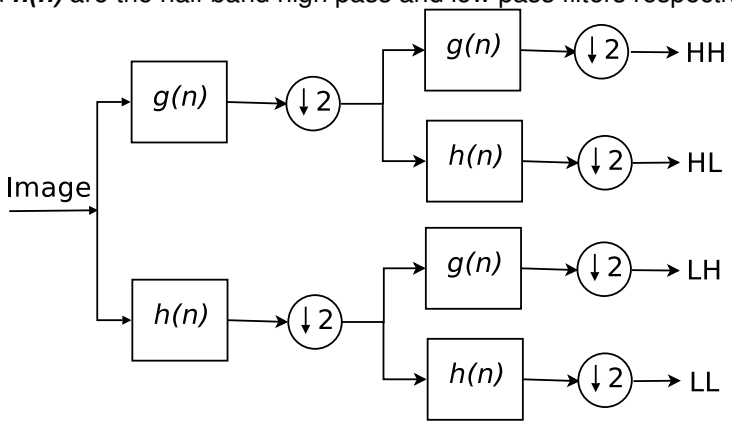


FIGURE 3: Filter structure for the wavelet decomposition of an image.

Resulting wavelet subbands of a face image are depicted in figure 4. Perfect reconstruction of the image is possible from the wavelet coefficients, using inverse DWT (IDWT). Wavelet subbands preserve the locality of spatial and spectral details in the image [18]. This property of spectral and spatial localization is useful in problems like image analysis, especially in super resolution. The type of filters used for ***g(n)*** and ***h(n)*** is determined by the wavelet associated. Face recognition experiments performed by Wayo Puyati, Somasak W. and Aranya W.[15] claims that Symlets give better performance in PCA based face recognition, over other wavelet. In this work, we have tested the proposed algorithm with Coiflets, Symlets and Daubechies wavelets.

FIGURE 4: Single level wavelet decomposition of face image.

4.2 Why Eigen Transformation in Wavelet Domain?

Face image has a specific structure and this prior information is utilized in face hallucination algorithms. In a specific class of properly aligned face images, contours, patterns and such facial features will be closely aligned. Discrete wavelet transform (DWT) decomposition of face image splits the image into four spectral bands without losing spatial details. Details in respective subbands will be more similar for different face images. It can be observed from Figure 5 that in any given subband other than LL subband, the patterns are similar for all images. Therefore, using very less number of Eigen images we will be able to capture the finer details accurately.

Hence a PCA based super resolution scheme in wavelet domain will be more efficient and computationally less expensive. Another importance of such transform domain approach is that, all images are stored in compressed format and most of the popular image compression techniques are in transform domain. Wavelet based compression is used in JPEG2000, MPEG4/H.264 and in many other standard image and video compression techniques. Therefore, the proposed algorithm can be directly applied on compressed images. This will considerably reduce the computational cost.

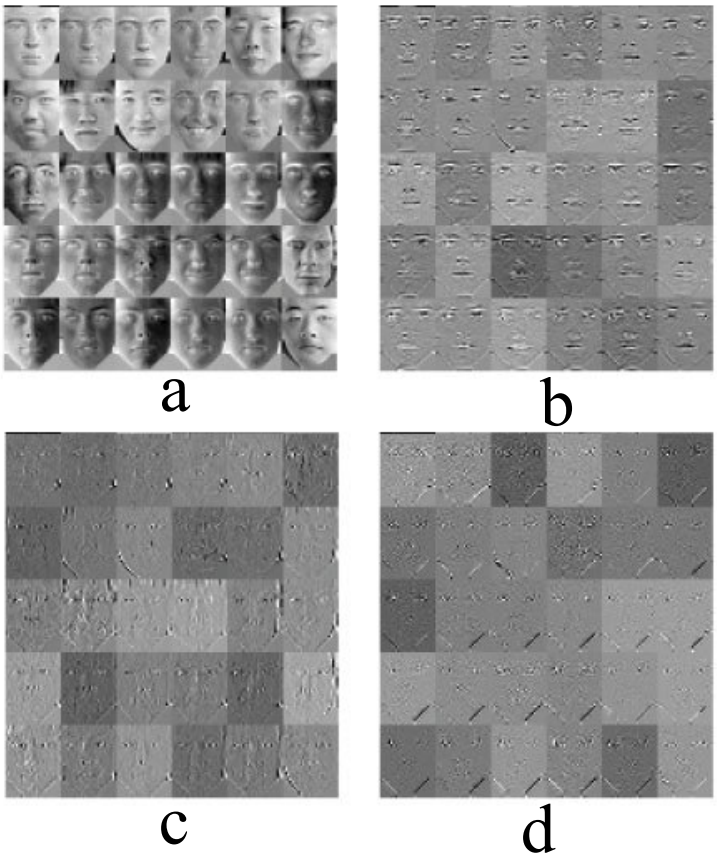


FIGURE 5: Wavelet subband images of face images in the training set. (a) LL subband, (b) LH subband, (c) HL subband and (d) HH subband images.

4.3 Super resolution with Eigen Transformation in Wavelet Domain

In this section we describe our face hallucination method using eigen transformation in the wavelet domain. The HR and LR face images in database are decomposed using DWT to form LR and HR wavelet coefficient database.
Define,

$$[L_{xx}] = DWT(L) \tag{11}$$

$$[H_{xx}] = DWT(H) \tag{12}$$

where ***xx*** stands for LL, LH, HL and HH wavelet subbands. The test image is also decomposed with DWT and then ET based super resolution method described in section 3.2 is applied on these wavelet subbands separately. Resulting wavelet coefficients, h_{SR-xx}, are given by

$$h_{SR-xx} = H_{xx} \times c_{xx} \tag{13}$$

where c_{xx} represents the coefficients for linear combination in different subbands, calculated using equation (7). The constraint based on Eigen value, as given in equation (9), is applied individually on all the super resolved wavelet subbands to obtain $\hat{h}_{SR-xx}$. Super resolved face image h_{SR} is computed by determining the IDWT of the coefficients $\hat{h}_{SR-xx}$.

$$h_{SR} = IDWT(\hat{h}_{SR-xx}) \tag{14}$$

The complete algorithm for face hallucination using ET in wavelet domain is summarized below.

Step 1: Prepare the HR and LR image databases and compute the wavelet subbands of all the images in the databases.
Step 2: For all the wavelet coefficients, find the vectors ***L***, the matrix ***K*** and the eigen vectors ***V*** as in section 3.2.
Step 3: Determine the significant Eigen vectors of ***C***
Step 4: Find the PCA coefficients w_I of the test image
Step 5: Compute the coefficients ***c***, using equation (7).
Step 6: The super resolved coefficients are obtained from equation (8).
Step 7: Modify the coefficients by applying the eigen value based constraints using equation (9).
Step 8: Reconstruct the wavelet subbands from the modified coefficients
Step 9: Reconstruct the super resolved images by finding the IDWT of super resolved wavelet coefficients.

5. FACE HALLUCINATION USING EIGEN TRANSFORMATION ON SUBSPACES IN WAVELET DOMAIN

In this section we describe a subspace based method for face hallucination in wavelet domain.

5.1 Super resolution using Eigen Transformation in Subspaces

In the case of a normal face image, some of the portions like eyes, nose, etc. are highly textured and more significant, so it needs more attention during super resolution. Bicubic interpolation will be sufficient for smooth regions like forehead, cheeks, etc. In our subspace based approach, face image is split into four subimages. They are left eye, right eye, mouth with nose and the remaining area as shown in figure 6. These regions are the subspaces of the entire face space.

Eigen transformation based super resolution technique out performs other hallucination methods if the test image is in the column span of Eigen vectors. If sub images are used for super resolution, only small number of images are required in the database compared to the case of whole face image for a given reconstruction error. Eigen transform based hallucination is applied on all the subimages separately and the resulting super resolved regions are combined along with the interpolated version of remaining area. The computational cost associated with this method is much less because it is comparatively easy to compute the Eigen vectors of smaller subspace images.

5.2 Eigen Transformation on Subspaces in Wavelet Domain

The subspace method proposed for face hallucination is an extension of the algorithm propsed in previous section. In this method HR and LR face images in database as well as the LR test image are split in to four regions as shown in figure 6. Then the three textured regions are individually super resolved using the algorithm explained in section 4.3. The fourth region is interpolated using bicubic interpolation and the three super resolved regions are combined with the fourth region to form the hallucinated face image. This subspace technique in wavelet domain for super resolution, reduces computational cost considerably, because the size of the subimages are small and therefore the computation required to determine wavelet coefficients are very less.

PCA in wavelet domain further reduces the memory required for implementation. This method is not suitable where input image resolution is very less, because it is not feasible to split and align test image into subimages when the input image resolution is very less.

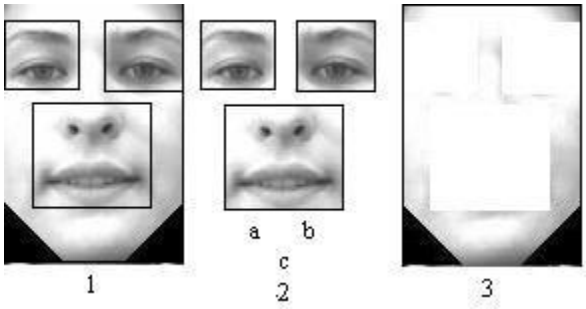


FIGURE 6: Face image divided into subspaces. (1) Entire face image with regions marked, (2 a, b, c) Textured regions, left eye, right eye and mouth with nose respectively. (3) Remaining smooth region.

The steps involved in this method for face are listed below:

Step 1: Split all the face images in the HR and LR databases into mouth with nose, left eye, right eye and remaining area.

Step 2: Determine the wavelet coefficients of eyes and mouth with nose.

Step 3: Repeat steps 2 to 9 of the algorithm described in section 4.3 on all the three textured portions.

Step 4: Combine the super resolved regions with interpolated version of remaining part to form the hallucinated face image.

6. EIGEN TRANSFORMATION IN DISCRETE COSINE TRANSFORM DOMAIN

In this section, we explain the usefulness of Discrete Cosine Transform (DCT) for face hallucination using Eigen transformation based approach. The DCT helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the visual quality of the image).

DCT has excellent energy compaction performance and therefore it is widely used in image compression [9]. Block wise DCT is usually used in image compression applications. DCT of image x_i with a DCT block size $N \times N$ is computed as

$$DCT(x_i) = \sum_x \sum_y x_i(x, y) g(x, y, u, v) \quad \text{for } u, v = 0 \text{ to } N-1 \tag{15}$$

where

$$g(x, y, u, v) = \alpha(u)\alpha(v) \cos\left[\frac{(2x+1)\pi u}{2N}\right] \cos\left[\frac{(2y+1)\pi v}{2N}\right] \tag{16}$$

After normalization of the DCT coefficients of LR and HR images, the low frequency side of HR coefficients and DCT coefficients of the corresponding LR images are very close and they represent the low frequency information [7]. The remaining DCT coefficients of HR image correspond to the high frequency information. Thus the process of super resolution in DCT domain is determining these remaining coefficients from the low frequency coefficients of LR image. The super resolution is applied on the block wise DCT coefficients of the image. Let Q be the magnification factor in x and y directions, $b \times b$ be the size of one DCT block for HR image and the $b/Q \times b/Q$ be the size of DCT block used for LR image and test image. Find the block wise DCT of all images in HR and LR databases as well as test image and then all coefficients are normalized with the maximum values in the DCT of each image. Now the values of DCT

coefficients in the ***b/Q × b/Q*** block corresponding to the low frequency side of HR image are very close to the corresponding DCT coefficients of LR image as shown in figure 7.

$$x_{dct-i} = DCT(x_i) \tag{17}$$

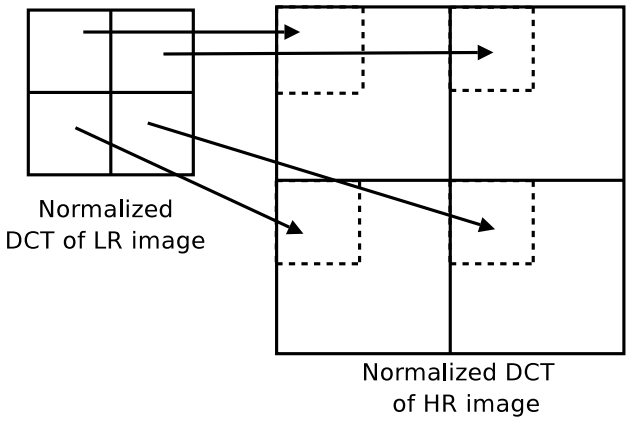


FIGURE 7: Relation between the normalized DCT coefficients of LR and corresponding HR images. Values of the DCT coefficients in the blocks connected by arrows correspond to low frequency details.

To perform ET, compute the matrices $\boldsymbol{L_{dct}}$, $\boldsymbol{H_{dct}}$ and $\boldsymbol{c_{dct}}$ as explained in equations (2) through (7). The DCT coefficients of the super resolved image are determined as explained in section 3.2. The resulting coefficients are the normalized DCT coefficients of hallucinated face image. Mean of the value used for normalizing the DCT coefficients of HR image is used to find the de-normalized DCT $\boldsymbol{x_{dct-h}}$. Now compute the inverse DCT to give the hallucinated face image $\boldsymbol{x_h}$.

$$x_h = IDCT(x_{dct-h}) \tag{18}$$

In this proposed algorithm for face hallucination in DCT domain, images are divided into fixed size blocks and the DCT of these blocks are determined as in the case of JPEG compression. All these blocks are considered together as a single unit for determining the SR image. With minimum modifications, this algorithm can be customized to use directly with different DCT based compression schemes.

7. EXPERIMENTAL RESULTS

All experiments in this paper are performed using manually aligned face images taken from PIE, Yale and BioID face databases. 100 front facial images are selected from the above databases. All the images are manually aligned using affine transformation and warping such that the distance between the centres of eyes is 50 pixels. Also the eyes, lip edges and tip of nose of all the images are aligned. Images are then cropped to *128 × 96* pixels. The high resolution images are having a resolution *128 × 96* and the low resolution images are derived from these HR images by subsampling them. Database is formed using these HR and LR images. Test image is also chosen from the LR image database as per leave one out policy, ie. the testing image is excluded from the training set for that particular experiment. Performances of the proposed techniques are quantified in terms of peak signal to noise ratio (PSNR), Mean structural similarity measure (MSSIM) [16] and correlation coefficient (CC). Structural Similarity Measure (SSIM) is defined by the relation

$$SSIM = \frac{(2\mu_x\mu_y)(2\sigma_{xy})}{(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2)} \tag{19}$$

SSIM is measured with a local window size of *8 × 8* pixels. Mean value of SSIM (MSSIM) is used as the metric. Ideal value of MSSIM is 1. Correlation coefficient (CC) between two images is computed by the following relation. Here also the ideal value is unity.

$$CC = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{20}$$

$$\sigma_{xy} = \frac{1}{N_1 N_2} \sum_{\forall i} \sum_{\forall j} (x(i,j) - \mu_x)(y(i,j) - \mu_y) \tag{21}$$

where N_1 and N_2 are the dimensions of the image.

$$\sigma_x^2 = \frac{1}{N_1 N_2} \sum_{\forall i} \sum_{\forall j} (x(i,j) - \mu_x)^2 \tag{22}$$

$$\mu_x = \frac{1}{N_1 N_2} \sum_{\forall i} \sum_{\forall j} x(i,j) \tag{23}$$

7.1 Eigen Transformation in Wavelet Domain

Experimental results of the method described in section 4.3 are shown in figure 8. Experiments are performed; a) to evaluate the visual quality of the hallucinated image, b) to test the performance of the algorithm with noisy observation and c) to find the variation in performance with different types of wavelet functions.

FIGURE 8: Hallucinated faces with the Eigen transformation in wavelet domain. Input, original, Bicubic interpolated and hallucinated images. For magnification factors of four (top), eight (middle) and eleven (bottom).

Figure 8 shows the hallucination result of the Eigen transformation in wavelet domain with magnification factors 4, 8 and 11. Experiments using this method are done with Daub2 wavelet function. Result of hallucination result is much better when the images in database precisely represent the features of the test face. But the result seems to be noisy when the test face is

significantly different from those in database. As it can be observed from figure 8, the hallucinated result is much better than the bicubic interpolation, for higher values of ***Q***. But if the number of pixels in the input image is very less, the proposed method fails to find the super resolved image. In our experiment, the resolution of HR image is *128 × 96* pixels. Therefore, it is observed that when the value of ***Q*** is above 11, size of input image will be less than *11 × 9* pixels and the algorithm fails to produce correct result. Figure 9 show the result of the proposed algorithm, when the test image not similar to the images in the database.

FIGURE 9: Hallucination result with a test image not similar to the database images. Input, original, Bicubic interpolated and hallucinated face images.

Next we perform the experiment with noisy test image. Gaussian noise is added to the test image. In this case, the test image and its corresponding HR version is included in the LR and HR databases respectively and the corresponding result is shown in Figure 10. This result shows the recognition performance of the algorithm with noisy observations.

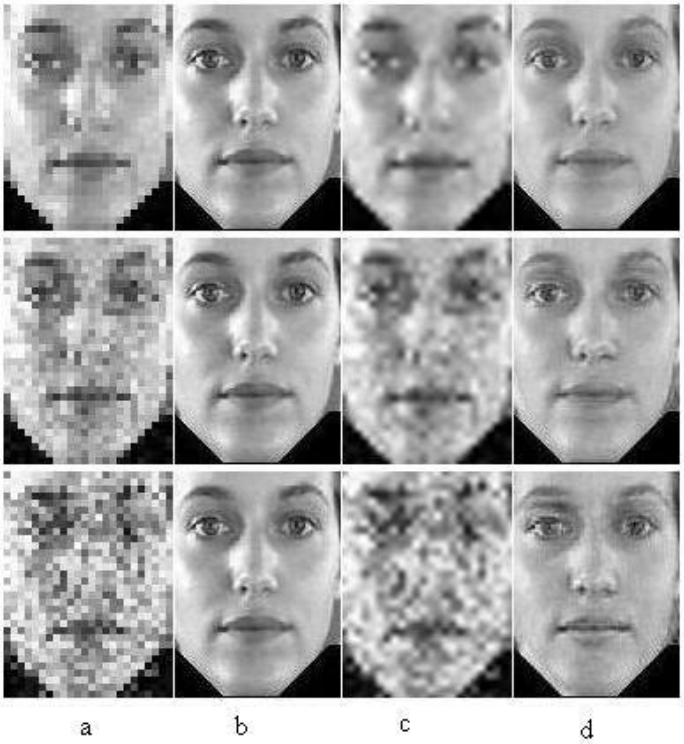


FIGURE 10: {Hallucinated faces using Eigen transformation in wavelet domain, with noisy input image. (a) Input image with Gaussian noise, (b) Original image, (c) Bicubic interpolated image and (d) hallucinated image. Noise variance σ=0.001 (top), σ =0.01 (middle) and σ =0.1 (bottom).

The proposed algorithm is then tested for the variation in performance with the different types of wavelets. In this particular case, algorithm increases the resolution by a factor of two both horizontally and vertically (magnification factor is two, ***Q****=2*). Experiments are performed with different types of wavelet functions. Table 1 show that results are better for daubechies2, coiflet5, symlets2 and symlets5 and the best result is obtained for symlet9.

FIGURE 11: Textured regions reconstructed using the algorithm proposed in section 4.4. Hallucinated, original and bicubic interpolated images.

FIGURE 12: Hallucinated face image with subspace PCA in wavelet domain. Hallucinated face, Original face and bicubic interpolated face image.

7.2 Eigen Transformation on Subspaces in Wavelet Domain

Experimental results of the face hallucination technique using Eigen transformation on subspaces in wavelet domain is given here. In order to implement the subspace based super resolution, face image is split in to four regions as explained in section 5. All the images in the database are aligned and thus the coordinates of all the subimages edges are predetermined. All the images in the database as well as the test image are split in to the subimages and the Eigen transformation based super resolution in wavelet domain is separately performed on all the subimages. Super resolved subimages are separately shown in figure 11 along with their original and bicubic interpolated versions. These results are for a magnification factor of two (***Q****=2*). Smooth regions in the image are interpolated and then the super resolved subimages are combined with the interpolated image to form the final hallucinated face. Figure 12 shows the final hallucinated face. Eyes, nose, lips etc are sharper than the bicubic interpolated version. Boundaries of the subimages are barely visible in this image, but it will become more visible as the magnification factor increases.

The proposed algorithm is then tested for the variation in performance with the different types of wavelets. Table 1 gives the change in performance with wavelet types. The subspace based method has best results when symlet7 and coiflet3.

7.3 Eigen Transformation in DCT Domain

Finally we show the results of face hallucination using Eigen transformation in DCT domain. Block wise DCT of all the images are computed. The values of ***b*** are chosen such that the value of ***b/Q*** is at least 2. If this value is less, PCA based super resolution in Eigen Transformation will be weak and the result will be noisy.

Wavelet Type	ET in Wavelet Domain (PSNR)	Subspace ET in Wavelet Domain (PSNR)
Symlet2	28.777	24.630
Symlet5	28.773	24.046
Symlet7	28.706	25.048
Symlet9	28.794	24.891
Coif3	28.641	25.048
Coif4	28.719	25.029
Coif5	28.765	24.961
Daub2	28.777	24.630
Daub3	28.671	24.802
Daub7	28.684	24.875

TABLE 1: Change in PSNR of hallucinated image for Eigen transformation in wavelet domain with different types of wavelets

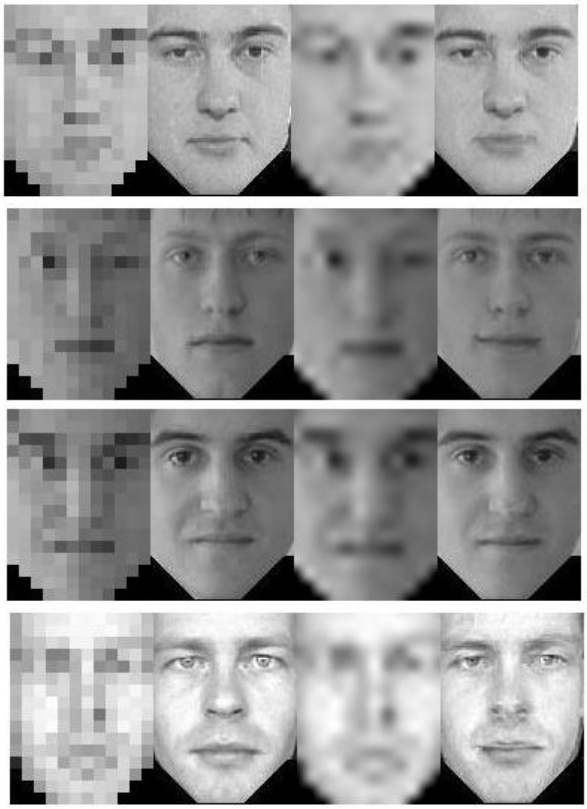


FIGURE 13: Result of hallucination experiments in DCT domain. Input image (first column), original image (second column), bicubic interpolated (third column) and Hallucinated result (Fourth column) for ***Q****=8*.

In our experiment the values of ***b*** is chosen as 16. Computed coefficients are normalized with the maximum values in the DCT of each image. Algorithm represents the DCT coefficients of LR

image as the linear combination of DCT coefficients in the LR image database. Resulting images are shown in figure 13 along with the input image, original image and bicubic interpolated image.

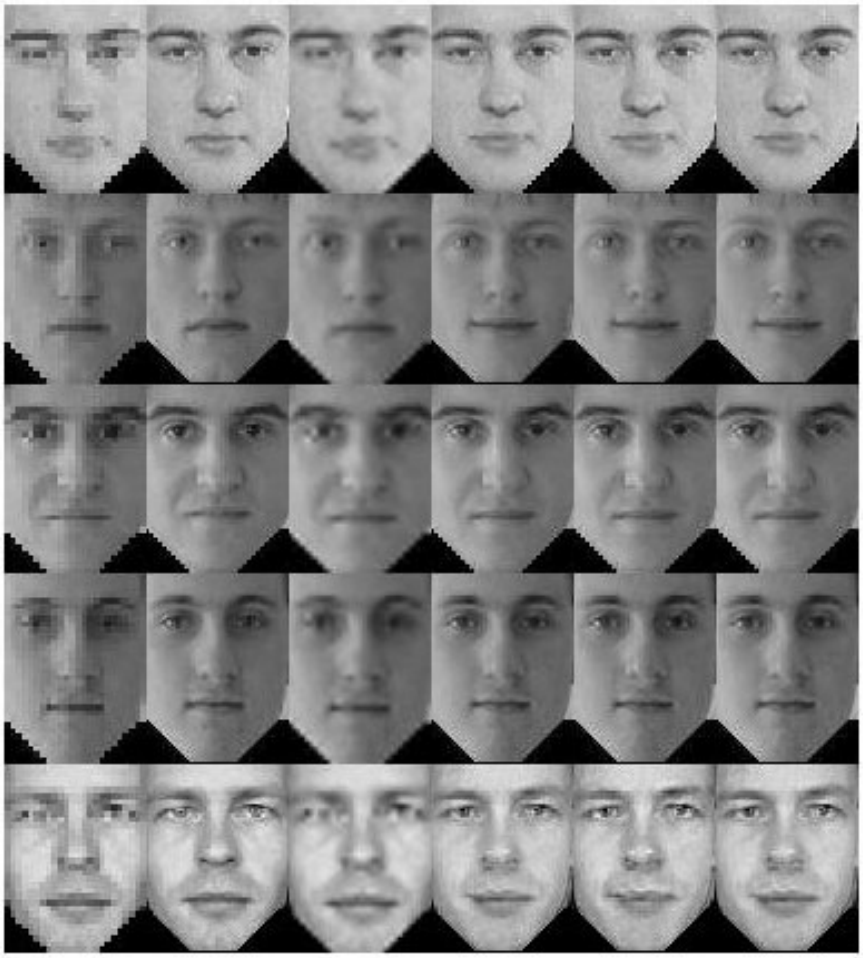


FIGURE 14: Result of hallucination experiments. Input image (first column), original image (second column), bicubic interpolated (third column), Eigen Transformation in spatial domain (fourth column), in wavelet domain (fifth column) and in DCT domain (sixth column). For ***Q=4***.

Test Image	ET in spatial Domain (*Q=4*)	ET in Wavelet Domain (*Q=4*)	ET in DCT Domain (*Q=4*)	Bicubic Inter-polation. (*Q=4*)	ET in spatial Domain (*Q=8*)	ET in Wavelet Domain (*Q=8*)	ET in DCT Domain (*Q=8*)
a	29.803	29.514	29.585	20.157	29.780	30.021	29.328
b	31.486	31.337	31.633	23.735	30.777	29.958	29.105
c	33.424	33.114	33.290	22.501	31.601	30.705	31.050
d	32.027	31.820	32.030	22.684	31.222	28.958	31.402
e	23.623	23.106	23.737	19.179	22.404	20.755	22.456

TABLE 2: Comparison of performance of Eigen transformation based face hallucination algorithms in spatial, wavelet and DCT domains. PSNR for magnification factors Q=4 and Q=8.

7.4 Comparison of hallucination results

Figure 14 and figure 15 shows the hallucination results of Eigen transformation in spatial domain, wavelet domain and DCT domain respectively for magnification factors 4 and 8. The first three

images in each set are input LR, original and bicubic interpolated images. Tables 2 and 3 compare Eigen transformations in three domains with respect to the parameters PSNR, MSSIM and CC respectively. Eigen transformation in DCT domain has the best performance followed by Eigen transformation in spatial domain and then in the wavelet domain, in terms of the above three parameters.

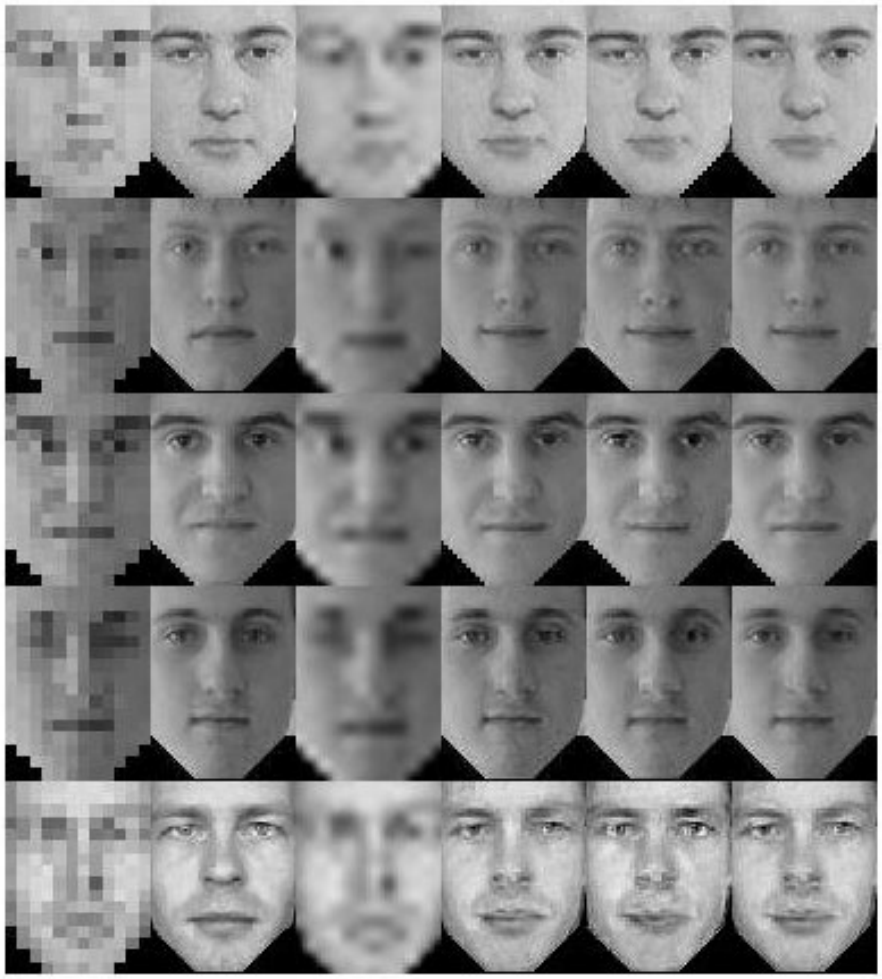


FIGURE 15: Result of hallucination experiments. Input image (first column), original image (second column), bicubic interpolated (third column), Eigen Transformation in spatial domain (fourth column), in wavelet domain (fifth column) and in DCT domain (sixth column). For ***Q****=8*.

8. CONSLUSION & FUTURE WORK

In this work, feasibility of Eigen transformation in transform domain for face super resolution is studied. Eigen transformation is applied in wavelet and DCT domains and the performance are compared. A subspace based super resolution method is also proposed in wavelet domain. The results show that Eigen transformation is applicable in transform domain, which means that the Eigen transform can be directly applied with slight modifications on the compressed images as well as on compressed video streams. Results obtained indicate that Eigen transformation in DCT and spatial domain has the best performance followed by Eigen transformation in wavelet domain. Results of Eigen transform based method are much better and it can be used for higher magnification factor. A disadvantage with Eigen transform based method is that, it depends on the alignment of images as well as the structural similarity of images. The effect of image alignment can be reduced by using pose and illumination invariant features instead of transform

coefficients. Our future work is intended on the study of the performance of Eigen transformation based hallucination on these features. Another possible extension is that study of the performance of Eigen transformation on individual DCT blocks, instead of the DCT of entire image.

Parameter	Test Image	ET in spatial Domain (*Q=4*)	ET in Wavelet Domain (*Q=4*)	ET in DCT Domain (*Q=4*)	ET in spatial Domain (*Q=8*)	ET in Wavelet Domain (*Q=8*)	ET in DCT Domain (*Q=8*)
MSSIM	**a**	0.84049	0.81407	0.83341	0.83905	0.84037	0.83562
	b	0.88562	0.87446	0.88476	0.88422	0.83214	0.86278
	c	0.90586	0.89278	0.90643	0.87119	0.85279	0.86685
	d	0.88597	0.87900	0.88627	0.87721	0.83214	0.87867
	e	0.67083	0.61993	0.66972	0.62173	0.54340	0.62760
Correlation Coefficient (CC)	**a**	0.98905	0.98810	0.98821	0.98869	0.98325	0.98746
	b	0.98795	0.98750	0.98785	0.98540	0.98267	0.98714
	c	0.99058	0.99168	0.99263	0.98917	0.98676	0.98913
	d	0.99261	0.99000	0.99056	0.98886	0.98151	0.98761
	e	0.96853	0.96347	0.96855	0.95821	0.94136	0.95913

TABLE 3: Comparison of performance of Eigen transformation based face hallucination algorithms in spatial, wavelet and DCT domains. Values of MSSIM and Correlation coefficient for a magnification factor Q=4 and Q=8.

9. REFERENCES

1. Simon Baker and Takeo Kanade, *"Hallucinating Faces"*, In Proceedings of Fourth International Conference on Automatic Face and Gesture Recognition, 2000.

2. Simon Baker and Takeo Kanade, *"Limits on Super resolution and how to break them"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.

3. Ce Liu, Heung-Yeung Shum and Chang Shui Zhang, *"A Two Step Approach to Hallucinating Faces: Global Parametric Model and Local Parametric Model",* In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2001.

4. Ce Liu, Heung-Yeung Shum and William T. Freeman, *"Face Hallucination: Theory and Practice",* International Journal of Computer vision Springer, 2007.

5. I.Daubechies, "*Ten Lectures on Wavelets*", SIAM, Philadelphia, 1992.

6. David Capel and Andrew Zisserman, *"Super-resolution from multiple views using learnt image models"*, In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), 2001.

7. A. Hossen and U. Heute, *"2D Subband Transforms: Theory and Applications"*, In Proceedings of IEE Vis. Image Signal Processing, Vol. 151, No. 5, October, 2004.

8. Jiji C.V., M.V. Joshi and Subhasis Chaudhuri, *"Single frame Image Super-resolution Using Learned Wavelet Coefficients"*, International Journal of Imaging Systems and Technology, 2004.

9. J. Makhoul, *"A Fast Cosine Transform in One and Two Dimensions"*, IEEE Tran. Acoustic and Speech Signal Processing, 28(1), 1980.

10. Todd K. Moon and Wynn C. Stirling, *"Mathematical Methods and Algorithms for Signal Processing"*, Pearson Education, 2005.

11. M.Turk and A. Pentland, *"Eigenface for Recognition"*, Journal of Cognitive Newroscience,1991.

12. Gonzalez and Woods, *"Digital Image Processing"*, Prentice Hall India.

13. X. Wang and X. Tang, *"Face Hallucination and Recognition"*, In Proceedings of 4th Int. Conf. Audio and video based Personal Authentication, IAPR, University of Surrey, Guildford, UK, 2003.

14. X. Wang and X. Tang, "Hallucinating Faces by Eigen transformation", IEEE Transactions on systems, man and cybernetics- Part C: Applications and Reviews, 2005.

15. Wayo Puyati, Somsak Walairacht and Aranya Walairacht, *"PCA in wavelet domain for face recognition"*, Department of computer Engineering, King Mongkut's Institute of technology, Bankok, ICACT 06, 2006.

16. M. Choi, R. Y. Kim, M. R. Nam, and H. O. Kim., *"Fusion of Multi-spectral and Panchromatic Satellite Images Using the Curvelet Transform"*, IEEE Transactions on Geosciences and Remote Sensing, 2(2):136--140, 2005.

17. J.K. Kailash and N. T. Sanjay, *"Independent Component Analysis of Edge Information for Face Recognition"*, International Journal of Image Processing (IJIP) Volume (3) : Issue (3), 120-130, 2009.

18. Abdu Rahiman V. and Jiji C.V., *"Face Hallucination using PCA in Wavelet Domain"*, In Proceedings of Third International Conference on Computer Vision Theory and Applications, VISAPP2008, 2008.

COLOR IMAGE SEGMENTATION BASED ON JND COLOR HISTOGRAM

Kishor Bhoyar kkbhoyar@ycce.edu
Assistant Professor, Department of Information Technology
Yeshwantrao Chavan College of Engineering
Nagpur 441110 India

Omprakash Kakde ogkakde@yahoo.com
Professor, Department of Computer Science Engineering
Vishweswarayya National Institute of Technology
Nagpur 440022 India

Abstract

This paper proposes a new color image segmentation algorithm based on the JND (Just Noticeable Difference) histogram. Histogram of the given color image is computed using JND color model. This samples each of the three axes of color space so that just enough number of visually different color bins (each bin containing visually similar colors) are obtained without compromising the visual image content. The number of histogram bins are further reduced using agglomeration successively. This merges similar histogram bins together based on a specific threshold in terms of JND. This agglomerated histogram yields the final segmentation based on similar colors. The performance of the proposed algorithm is evaluated on Berkeley Segmentation Database. Two significant criteria namely PSNR and PRI (Probabilistic Rand Index) are used to evaluate the performance. Results show that the proposed algorithm gives better results than conventional color histogram (CCH) based method and with drastically reduced time complexity.

Keywords: Color Image Segmentation, Just noticeable difference, JND Histogram.

1. INTRODUCTION

Color features of images are represented by color histograms. These are easy to compute, and are invariant to rotation and translation of image content. The potential of using color image histograms for color image indexing is discussed by [1]. However color histograms have several inherent limitations for the task of image indexing and retrieval. Firstly, in conventional color histogram (CCH) two colors will be considered totally different if they fall into two different bins even though they might be very similar to each other for human perception. That is, CCH considers neither the color similarity across different bins nor the color dissimilarity in the same bin. Therefore it is sensitive to noisy interferences such as illumination changes and quantization errors. Secondly, CCH's high dimensionality (i.e. the number of histogram bins) requires large computations on histogram comparison. Finally, color histograms do not include any spatial information and are therefore not suitable to support image indexing and retrieval, based on local image contents. To address such issues various novel approaches were suggested, like spatial color histogram [2], merged color histogram [3], and fuzzy color histogram [4].

Segmentation involves partitioning an image into a set of homogeneous and meaningful regions, such that the pixels in each partitioned region possess an identical set of properties. Image segmentation is one of the most challenging tasks in image processing and is a very important pre-processing step in the problems in the area of image analysis, computer vision, and pattern recognition [5,6]. In many applications, the quality of final object classification and scene interpretation depends largely on the quality of the segmented output [7]. In segmentation, an image is partitioned into different non-overlapping homogeneous regions, where the homogeneity of a region may be composed based on different criteria such as gray level, color or texture.

The research in the area of image segmentation has led to many different techniques, which can be broadly classified into histogram based, edge based, region based, clustering, and combination of these techniques [8,9] . Large number of segmentation algorithms are present in the literature, but there is no single algorithm that can be considered good for all images [7]. Algorithms developed for a class of images may not always produce good results for other classes of images.

In this paper we present a segmentation scheme based on JND (Just Noticeable Difference) histogram. Color corresponding to each bin in such histogram is visually dissimilar from that of any other bin; whereas each bin contains visually similar colors. The color similarity mechanism is based on the threshold of similarity which is based on Euclidean distance between two colors being compared for similarity. The range of this threshold for fine to broad color vision is also suggested in the paper, based on sampling of RGB color space suggested by McCamy [10].

The rest of the paper is organized as follows. Section 2 gives the brief overview of JND model and computation of color similarity threshold in RGB space, and computation of JND histogram. Section 3 presents the algorithm for agglomeration of JND histogram and the subsequent segmentation based on JND histogram. Section 4 presents the results of the proposed algorithm on BSD, and its comparison based on two measures of segmentation quality namely PSNR and PRI. Section 5 gives the concluding remarks and future work.

2. JND COLOR MODEL AND JND HISTOGRAM

2.1 Overview of JND Color model

The JND color model in RGB space based on limitations of human vision perception as proposed in [11] is briefed here for ready reference. The human retina contains two types of light sensors namely; rods and cones, responsible for monochrome i.e. gray vision and color vision respectively. The three types of cones viz, Red, Green and Blue respond to specific ranges of wavelengths corresponding to the three basic colors Red, Green and Blue. The concentration of these color receptors is maximum at the center of the retina and it goes on reducing along radius. According to the three color theory of Thomas Young, all other colors are perceived as linear combinations of these basic colors. According to [12] a normal human eye can perceive at the most 17,000 colors at maximum intensity without saturating the human eye. In other words, if the huge color space is sampled in only 17,000 colors, a performance matching close to human vision at normal illumination may be obtained. A human eye can discriminate between two colors if they are at least one ‘just noticeable difference (JND)’ away from each other. The term ‘JND’ has been qualitatively used as a color difference unit [10].

If we decide equal quantization levels for each of the R, G and B axes, then we require approximately 26 quantization levels each to accommodate 17000 colors. But from the physiological knowledge, the red cones in the human retina are least sensitive, blue cones are moderately sensitive and the green cones are most sensitive. Keeping this physiological fact in mind, the red axis has been quantized in 24 levels and the blue and green axes are quantized in 26 and 28 levels [11]. The 24x26x28 quantization in the RGB space results in slight over-

sampling (17,472 different colors) but it ensures that each of the 17,000 colors is accommodated in the sampled space. Heuristically it may be verified that any other combination of quantization on the R,G and B axes results in either large under sampling or over-sampling as required to accommodate 17000 colors in the space. Although the actual value of the just noticeable difference in terms of color co-ordinates may not be constant over the complete RGB space due to non-linearity of human vision and the non-uniformity of the RGB space, the 24x26x28 quantization provides strong basis for deciding color similarity and subsequent color segmentation as demonstrated in this work.

Using this sampling notion and the concept of 'just noticeable difference' the complete RGB space is mapped on to a new color space J_r J_g J_b where J_r, J_g and J_b are three orthogonal axes which represent the Just Noticeable Differences on the respective R,G and B axes. The values of J on each of the color axes vary in the range (0,24) ,(0,26) or (0,28) respectively for red, blue and green colors. This new space is a perceptually uniform space and offers the advantages of the uniform spaces in image analysis.

2.2 Approximating the value of 1 JND_h

For a perfectly uniform color space the Euclidean distances between two colors is correlated with the perceptual color difference. In such spaces (e.g. CIELAB to a considerable extent) the locus of colors which are not perceptually different from a given color, forms a sphere with a radius equal to JND. As RGB space is not a perceptually uniform space, the colors that are indiscernible form the target color, form a perceptually indistinguishable region with irregular shape. We have tried to derive approximate value of JND by 24x26x28 quantization of each of the R,G, and B axes respectively. Thus, such perceptually indistinguishable irregular regions are modeled by 3-D ellipsoids for practical purposes.

The research in physiology of human eye indicates two types of JND factors involved in the human vision system. The first is the JND of human eye referred to as JND_{eye} and the second is the JND of human perception referred to as JND_h. It is found that the neural network in human eye is more powerful and can distinguish more colors than those ultimately perceived by the human brain. The approximate relationship between these two [11] is given by equation (1) .

$$\mathrm{JND}_{\mathrm{h}} = 3.\mathrm{JND}_{\mathrm{eye}} \quad ---(1)$$

Let C_1 and C_2 be two RGB colors in the new quantized space. Let C_1= (Jr_1,Jg_1,Jb_1) =(0,0,0) and its immediate JND neighbour, that is 1 noticeable difference away is C_2= (Jr_2,Jg_2,Jb_2) =(255/24,255/28,255/26). Hence JND_{eye}= sqrt ((255/24)^2 + (255/28)^2 + (255/26)^2)) = sqrt (285.27). Using equation (1) the squared JND threshold of human perception is given by equation (2).

$$\Theta = \mathrm{JND}_{\mathrm{h}}^{\ 2} = 2567 \quad ---(2)$$

In equation (1), the squared distance is used, to avoid square root computation and hence to reduce time complexity. The use of Θ as a squared threshold is very convenient mechanism to exploit perceptual redundancy inherent in digital images, as it gives opportunity to work in sampled color space without compromising on the visual quality of results. For practical applications the range of Θ for fine to broad vision is $\mathrm{JND}_{\mathrm{eye}}^{2} \le \Theta \le \mathrm{JND}_{\mathrm{h}}^{2}$.

2.3 Computing JND histogram

Histogram of an image manifests an important global statistics of digital images, which can be used for a number of analysis and processing algorithms. In color image histograms, a large number of colors may be present as required for representing real life images. All of these colors may not even be noticed as different colors by normal human eye [10], hence as the first step the

histogram on each of its axis has been sampled suitably to accommodate all the human distinguishable colors.

In this section, we propose an algorithm for computing histogram of a color image in RGB space. As the structure of this histogram is four dimensional it is difficult to represent it in a 2-D plane and hence it is hardly possible to plot. Thus the histogram of an RGB image I=$f(1),f(2),f(3),\ldots\ldots,f$(m x n) is given by H(r,g,b) as in equation(3), where m and n are rows and columns of the image respectively and I represents the color intensity values[13]. N is a counter variable and r, g, b represents the color coefficients.

$$H(r,g,b) = \sum_{i=1}^{m.n} N\Big|_{f=f(r,g,b)} \quad ---(3)$$

In the proposed histogram, the first data structure is a table of size nx4, where n indicates number of different colors in the image. Out of the four columns three are used for the RGB color intensities and the fourth is for population of that color. The second data structure is a table of (r x c) rows (where r and c indicate rows and columns in an image) and out of the three columns the first one is used for the color index i.e. row number in the first data structure and the remaining two are for storing the respective color position information in terms of the x and y coordinates of the color pixel. In this form the color image histogram becomes a solid cube while the density of the cube at a point in it represents the frequency and the three orthogonal edges of the cube represent the basic R, G and B colors. A traditional histogram does not contain any positional information. With the positional information stored in the proposed histogram, it just becomes transform of an image. In other words, the image can be obtained back from the new histogram with the positional information. The spatial color distribution information also plays an important role in the image analysis. Both of these histogram tables have been shown in Table 1 and Table 2. These new histogram data structures will be collectively called as JND histogram.

For practical purposes already discussed, the color image histograms have to be sampled on R, G and B axis suitably to reduce the number of colors. Most of the literature till now either uses uniform sampling of the R, G, B axis or uses images represented in uniform color spaces. Such a uniformly sampled histogram can be represented by equation (4) with the same symbols. δ represents sampling interval on each axis and p is an integer variable.

$$H(p\delta r, p\delta g, p\delta b) = \sum_{i=1}^{m.n} N\Big|_{f=f(p\delta r, p\delta g, p\delta b)} \quad ---(4)$$

The four dimensional color image histogram is represented by two linked structures as given in Table 1 and Table 2. Actual implementation of the Table 2 may contain only m.n integer entries for JND color index, as the pixel entries are sorted spatially from left to right and from top to bottom.

JND Color Index	R	**G**	**B**	**H**
1	5	15	20	335
2	10	100	20	450
3	20	50	10	470
.	.	.	.	.
K	25	72	90	200

TABLE 1: Color population

X	**Y**	JND Color Index for (x_i,y_i) from Table 1
x_1	y_1	1
x_1	y_2	1
.	.	.
x_i	y_i	K
.	.	.

TABLE 2: Color index-pixel location relation

Thus Table 1 contains the R, G, B coordinates and the respective frequency information or population (H) of the tri-color stimulus, while Table 2 contains the respective color index (row index) in Table 1 and the x and y positional co-ordinates in the image. The number of rows in Table 1 is equal to the number of different color shades available in the image. In Table 1 there will be one entry for each color shade while in Table 2 there will be one entry for each pixel. The color shades which are not present in the image are not allotted any row in Table 1 and hence in Table 2. The color vectors are entered in the Table 1 in the order of their appearance in the image or in other words, as they are encountered during the scan of the image which starts from the top left corner of the image. The population (H) in Table 1 must satisfy equation (5).

$$\sum H = m.n \quad ---(5)$$

The proposed histogram computation procedure given below finds out the color shades available in the image and arranges them in the said format in a single scan of the complete image. Thus it does not require three scans of the complete image as in [14] neither it requires as many passes as minimum of the R,G and B frequencies as in [15]. It also simultaneously notes the positional information in a separate data structure which may further be used by different algorithms like shell clustering algorithms [16,17], which require positional information. The histogram computing algorithm with our approach has been presented below.

Algorithm for Computing the Basic JND Histogram

i) Initialize two data structures Table 1 and Table 2. Initialize the first entry in Table 1 by the first color vector in the image i.e. top left pixel color vector [R,G,B] and the frequency(population) by one. Initialize the first entry in Table 2 by the current row index value of Table 1 i.e. 1 , and the top left pixel position row and column i.e. y(column)=1 and x(row)=1. Also initialize a (row, column) pointer to top left corner of the image. Select a proper similarity threshold Θ_1 ($JND_{eye}^2 \leq \Theta_1 \leq JND_h^2$) depending on the precision of vision from fine to broad as required by the application.
ii) Read the next pixel color vector in scan line order.
iii) Compare the new pixel color vector with all the previous entries in Table 1 one by one and if found similar to any of them, then accommodate it in the respective bin. Update Table 2 by entering the current index and the current row and column values and go to step v.
iv) If the new color vector is not equal to any of the previously recorded color vectors in Table 1, increment the row index of Table 1, enter the new color vector in it, set the population to 1 , make the index , row and column entry in Table 2 and go to step ii.
v) Repeat step ii) to iv) for all the pixels in the image.
vi) Sort Table 2 in the increasing order of the color index.
vii) Save the Table 1 and Table 2 for latter analysis of the histogram.

The histogram computed using JND_h^2 derived in section 2.2 as threshold, using above algorithm has miraculously reduces the colors in the natural images. The drastic reduction in number of colors in a natural image brings it to the range suitable for the machine analysis in real time. The k visually different colors in Table 1 found by the basic algorithm are further reduced using agglomeration procedure discussed in the next section.

3. HISTOGRAM AGGLOMERATION AND SEGMENTATION

Agglomeration in chemical processes attributes to formation of bigger lumps from smaller particles. In the digital image segmentation, the similar pixels (in some sense) are clustered together under some similarity criteria. And thus it was inspired that the agglomeration may contribute considerably in the process of color image segmentation. In this section, a basic agglomeration histogram processing algorithm is presented. The multidimensional histogram peak detection and thresholding are complex and time consuming tasks. The agglomeration techniques can be thought of as the powerful alternatives to the other image thresholding techniques. After the compressed histogram of a real life image is obtained using the basic JND

histogram algorithm given in section 2, the agglomeration technique can further be used to reduce the number of colors by combining the smaller segments (less than .1% [18] of the image size) with similar colored larger segments. To implement this scheme a merging threshold Θ_2 which is slightly greater than Θ_1 is used, typically $\Theta_2 = \Theta_1 + 100$ works well here. This stimulates the process of merging of small left over segments (after building basic JND histogram presented in section 2.3) with larger similar color segments. This helps in minimizing over segmentation. The basic agglomeration algorithm has been presented below.

Algorithm for Computing Agglomerated Histogram using JND colors

i) Arrange the Table 1 in decreasing order of population.
ii) Starting from the first color in Table 1, compare the color with the next color in table 1.
iii) If the population of the smaller segment is smaller than .1% of the image size and the two segments are similar using Θ_2, merge the i_{th} color with the previous one (the first in Table 1), their populations will be added and the color of larger population will represent the merger.
iv) The merged entry will be removed from Table 1. This reduces number of rows in Table 1. In Table 2, the color index to be merged is changed by the index to which it is merged.
v) Thus the first color in the Table 1 will be compared with every remaining color in Table 1 and step ii is repeated if required.
vi) Step ii, iii and iv are repeated for every color in the Table 1.
vii) Steps ii to v are repeated till the Table 1 does not reduce further i.e. equilibrium has reached.
viii) Table 2 is sorted in ascending order of the color index.

The human retina performs a low pass filtering operation following Poisson's distribution around every point on the retinal image and the neural activity initially notes and interprets the predominant or above average outputs of the retinal sensors passed to the brain via visual cortex [12]. Though we have not implemented the classical Poisson's distribution based spatial integration, the agglomeration in this work has carried out the task of low pass filtering in the color space. This reduces the number of colors in an image from several thousands to a few tens. Based on this human physiological background, the prominent segments of the image can be estimated from the agglomerated histogram.

Segmentation procedure is straightforward with the data structures given in Table 1 and Table 2. In Table 2, the pixel entries are sorted spatially from left to right and from top to bottom. Segmented image can simply be formed by assigning to each pixel position a JND color from Table 1 as pointed to by the respective index in Table 2.

4. EXPERIMENTAL RESULTS

In this section, we demonstrate the segmentation results of the proposed algorithm on natural images from Berkeley Segmentation Database (BSD)[19]. It Contains 300 real life RGB images of different categories and same size 481x321 pixels. It also contains benchmark segmentation results (ground truth database) of 1633 segmented images manually obtained from 30 human subjects. i.e. multiple ground truth hand segmentations of each image. For each image, the quality of segmentation obtained by any algorithm can be evaluated by comparing it with ground truth hand segmentations.

The results of proposed segmentation algorithm are presented here and its effectiveness is compared with the conventional histogram based segmentation, using two quantitative measures, namely, the Probabilistic Rand Index (PRI)[20] and Peak signal to Noise Ratio (PSNR). PRI Counts the fraction of pairs of pixels whose labeling are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for variation in human perception. This measure takes the values in the interval [0,1]; more is better. We will consider the segmentation 'good' if for any pair of pixels x_i, x_j we would

like the labels of those pixels l_i^{Stest}, l_j^{Stest} to be the same in the test segmentation if the labels l_i^{Sk}, l_j^{Sk} were the same in the ground truth segmentations, and vice versa.

PSNR represents region homogeneity of the final partitioning. The higher the value of PSNR the better is segmentation. The PSNR measure [21] between the image I and the first order approximation based on the segmentation result S is calculated by equation (6).

$$PSNR(I, S) = 10 \cdot \log_{10}\left(\frac{255^2 \text{ rows} \cdot \text{columns} \cdot \text{channels}}{\sum_i^{rows} \sum_j^{columns} \sum_k^{channels} [I(i,j,k)-S(i,j,k)]^2}\right) \quad ----(6)$$

The algorithm is applied to BSD database of 300 images. The average PRI and PSNR values of all the images for two algorithms is given in Table 3. The quantitative comparison as given in Table and the qualitative (visual) comparison presented in figure 2 clearly demonstrate the superiority of proposed algorithm.

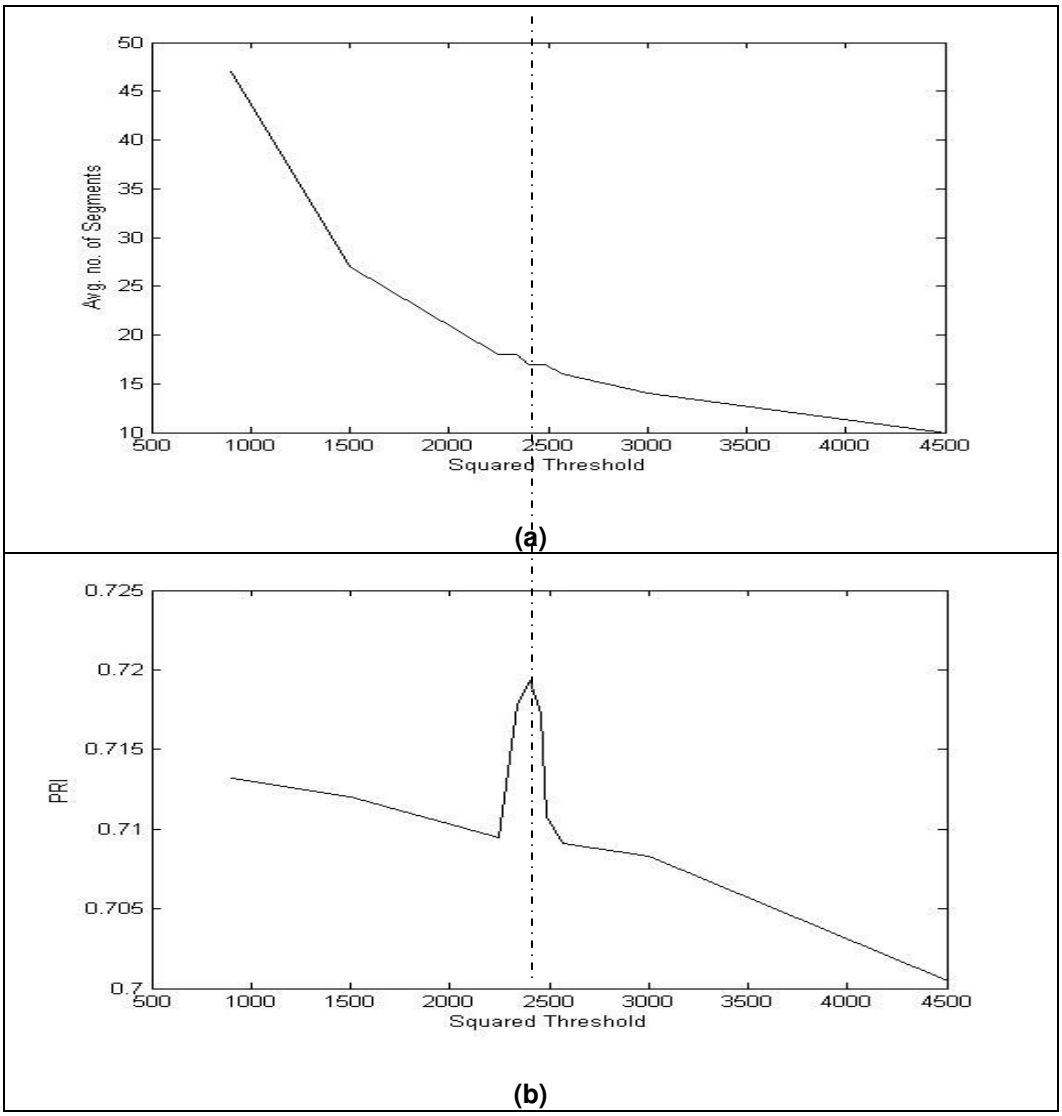


FIGURE 1: Graphs showing the effect of different values of squared threshold Θ_1 on average number of segments in BSD (a) and on PRI of BSD (b). The best results are obtained at Θ_1=2400. Θ_2= Θ_1+100 for all experiments

The Segmentation experiments for various values of Θ_1 are performed to find the best segmentation results. The Figure 1 shows the graphs plotted to analyse the performance of the proposed algorithm. The graphs shows the best performance (PRI=.7193) at value 2400 (shown with vertical dotted line at Θ=2400) which is very close to the derived value of JND_h^2=2567. Also note that the value of average number of segments (approximately 17) at Θ_1=2400 is reasonable. More the value of Θ_1, less are the average number of segments and vice versa. This is obvious as increased value of Θ_1 accepts more pixels as similar to given pixel and hence increases the size of the segments; thus producing lesser number of segments for any given image. From this discussion it can be concluded that we can implement fine to broad vision by varying value of Θ from JND_{eye} to JND_h.

Seg. Method	PRI	PSNR	Time* for Segmentation of 300 BSD images
CCH	0.7181	21.37	1.12 Hrs
JND based Color Histogram Θ=2400	0.7193	25.60	0.3575 Hrs

TABLE 3: Average Performance on BSD *On AMD Athlon 1.61 GHz processor

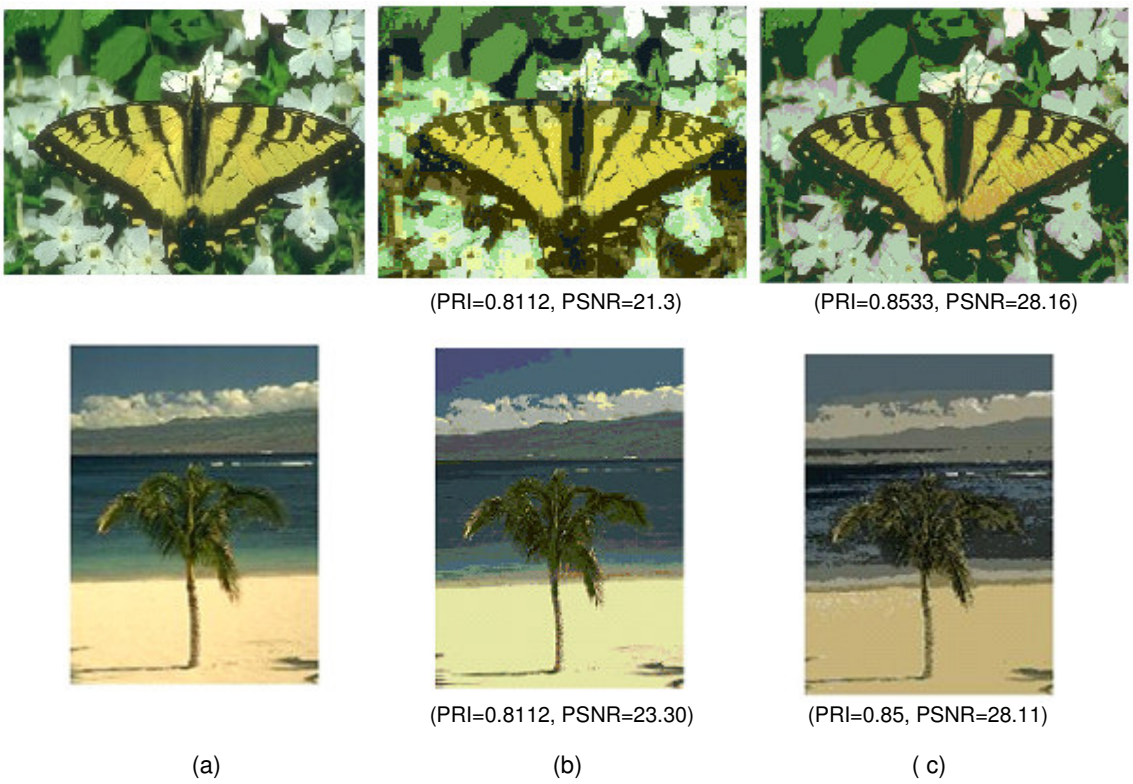


FIGURE 2: Few Segmentation results on BSD database: Column (a) Original Images, Column (b) CCH Segmented images, and Column (c) Segmented images with proposed approach.

5. CONCLUSION AND FUTURE WORK

Observing the graphs given in Figure 1, the effect of squared threshold Θ_1 on JND Histogram based Segmentation can be summarized as follows. It is observed that as Θ_1 increases; average number of segments (Avgk) over BSD exponentially decreases. As Θ_1 increases, PRI decreases. Optimal value of Θ_1 (considering both Avgk and PRI) should be around 2567. It is observed that the results on Segmentation on natural images in BSD are optimal near JND_h^2=2400. This approves our claim on derived value of JND for RGB color space.

The result comparison of JND Histogram Segmentation approach with CCH based segmentation approach on BSD images is summarized in Table-3. It can be observed that the proposed segmentation approach outperforms CCH in terms of PRI as well as PSNR. Also note the reduction in time required to perform the segmentation of all the 300 images in the database, with the proposed approach.

The information obtained about number of segments, and their cluster centers can be used to initialize Fuzzy C-means Segmentation algorithm. In future work we are proposing a modified FCM segmentation algorithm that works with the histogram bins as data for clustering instead of individual pixel values.

6. REFERENCES

1. M.Swain and D. Ballard,"*Color indexing*", International Journal of Computer Vision, Vol.7, no. 1,1991.
2. W. Hsu, T.S. Chua, and H. K. Pung, "*An Integrated color-spatial approach to Content-Based Image Retrieval*", ACM Multimedia Conference, pages 305-*313, 1995.*
3. *Ka-Man Wong, Chun-Ho Chey, tak-Shing Liu, Lai-Man Po, "Dominant color image retrieval using merged histogram"*, Circuits and Systems,ISCAS'03 Proceedings of 2003 International Symposium, Vol. 2, pp II-908 – II-911, 2003
4. Ju Han and Kai-Kuang Ma, "*Fuzzy color Histogram and its use in color image retrieval*", IEEE Transactions on Image Processing, Vol. 11, No. 8, 2002.
5. Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J., "*Color image segmentation: Advances and prospects"*, Pattern Recognition 34,2259–2281, 2001.
6. Liew, A.W., Yan, H., Law, N.F., "*Image segmentation based on adaptive cluster prototype estimation*", IEEE Trans. Fuzzy Syst. 13 (4), 444–453, 2005.
7. Pal, N.R., Pal, S.K., "*A review on image segmentation techniques*", Pattern Recognition 26 (9), 1277–1294, 1993.
8. Aghbari, Z. A., Al-Haj, R., "*Hill-manipulation: An effective algorithm for color image segmentation*", Image Vision Comput. 24 (8), 894–903, 2006..
9. Cheng, H.D., Li, J., "*Fuzzy homogeneity and scale-space approach to color image segmentation*", Pattern Recognition 36, 1545–1562, 2003.
10. Gaurav Sharma, "*Digital color imaging*", IEEE Transactions on Image Processing, Vol. 6, No.7, , pp.901-932, July1997.
11. K. M. Bhurchandi, P. M. Nawghare, A. K. Ray, "*An analytical approach for sampling the RGB color space considering limitations of human vision and its application to color image analysis*",, Proceedings of ICVGIP 2000, Banglore, pp.44-49.
12. A. C. Guyton, "*A text book of medical Physiology*", W.B.Saunders company, Philadelphia, pp.784-824, (1976).
13. A. Moghaddamzadeh and N. Bourbakis, "*A fuzzy region growing approach for segmentation of color images*", Pergamon,Pattern Recognition, Vol.30,No.6, pp.867-881, 1997.

14. Sang Ho Park, Il Dong Yun and Sang Uk Lee, "*Color image segmentation based on 3-D clustering: morphological approach*", Pergamon, Pattern Recognition, Vol.44, No.8, pp. 1061-1076, 1998.
15. Liang-Kai Huang and Mao-Jiun J.Wang, "*Image thresholding by minimizing the measures of fuzziness*", Pergamon,Pattern Recognition, Vol.28,No.1, pp.41-51, 1995.
16. Raghu Krishnapuram, Hichem Frigui and olfa Nasraoui, "*Fuzzy possiblistic shell clustering Algorithms and their application to boundary detection and surface approximation- part I*", IEEE Transactions on Fuzzy Systems, Vol.3,No.1, pp.29 -43, February1995.
17. Raghu Krishnapuram, Hichem Frigui and olfa Nasraoui, "*Fuzzy possiblistic shell clustering Algorithms and their application to boundary detection and surface approximation- part II*", IEEE Transactions on Fuzzy Systems, Vol.3,No.1, pp.44-60, February1995.
18. Milind M. Mushrif, Ajoy K. Ray,"*Color image segmentation:Rough-set theoretic approach*" ,Elsevier Pattern Recognition Letters, pp 483-493,2008.
19. D. Martin, C. Fowlkes, D. Tal, J. Malik, "*A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*", Proceedings of IEEE International Conference on Computer Vision, 2001, pp.416 –423
20. R. Unnikrishnan, M. Hebert, "*Measures of Similarity*", IEEE Workshop on Computer Vision Applications, pp. 394–400, , 2005.
21. D. Suganthi, S. Purushothaman, "IMRI Segmentation using echo state neural network", International Journal of Image Processing, Volume (2):Issue (1), pp 1-9.

A DUAL TREE COMPLEX WAVELET TRANSFORM CONSTRUCTION AND ITS APPLICATION TO IMAGE DENOISING

Sathesh sathesh_ece@yahoo.com

Assistant professor / ECE / School of Electrical Science
Karunya University, Coimbatore, 641114, India

Samuel Manoharan samuel1530@gmail.com

Phd Scholar / ECE / School of Electrical Science
Karunya University, Coimbatore, 641114, India

Abstract

This paper discusses the application of complex discrete wavelet transform (CDWT) which has significant advantages over real wavelet transform for certain signal processing problems. CDWT is a form of discrete wavelet transform, which generates complex coefficients by using a dual tree of wavelet filters to obtain their real and imaginary parts. The paper is divided into three sections. The first section deals with the disadvantage of Discrete Wavelet Transform (DWT) and method to overcome it. The second section of the paper is devoted to the theoretical analysis of complex wavelet transform and the last section deals with its verification using the simulated images.

Keywords: Complex Discrete Wavelet Transform (CDWT), Dual-Tree, Filter Bank, Shift Invariance, Optimal Thresholding.

1. INTRODUCTION

The application of wavelets to signal and image compression and to denoising is well researched. Orthogonal wavelet decompositions, based on separable, multirate filtering systems have been widely used in image and signal processing, largely for data compression. Kingsbury introduced a very elegant computational structure, the dual - tree complex wavelet transform [5], which displays near-shift invariant properties. Other constructions can be found such as in [11] and [9]. As pointed out by Kingsbury [5], one of the problems of mallat-type algorithms is the lack of shift invariance in such decompositions. A manifestation of this is that coefficient power may dramatically re –distribute itself throughout subbands when the input signal is shifted in time or in space.

Complex wavelets have not been used widely in image processing due to the difficulty in designing complex filters which satisfy a perfect reconstruction property. To overcome this, Kingsbury proposed a dual-tree implementation of the CWT (DT CWT) [7], which uses two trees of real filters to generate the real and imaginary parts of the wavelet coefficients separately. The two trees are shown in Fig. 3 for 1D signal. Even though the outputs of each tree are downsampled by summing the outputs of the two trees during reconstruction, the aliased components of the signal can be suppressed and approximate shift invariance can be achieved. In this paper CDWT, which is an alternative to the basic DWT the outputs of each tree are downsampled by summing the outputs of the two trees during reconstruction and the aliased

components of the signal are suppressed and approximate shift invariance is achieved. The DWT suffers from the following two problems.

- Lack of shift invariance - this results from the down sampling operation at each level. When the input signal is shifted slightly, the amplitude of the wavelet coefficients varies so much.
- Lack of directional selectivity - as the DWT filters are real and separable the DWT cannot distinguish between the opposing diagonal directions.

These problems hinder the use of wavelets in other areas of image processing. The first problem can be avoided if the filter outputs from each level are not down sampled but this increases the computational costs significantly and the resulting undecimated wavelet transform still cannot distinguish between opposing diagonals since the transform is still separable. To distinguish opposing diagonals with separable filters the filter frequency responses are required to be asymmetric for positive and negative frequencies. A good way to achieve this is to use complex wavelet filters which can be made to suppress negative frequency components. The CDWT has improved shift-invariance and directional selectivity than the separable DWT.

The work described here contains several points of departure in both the construction and application of dual tree complex wavelet transform to feature detection and denoising.

2. DESIGN OVERVIEW

The dual-tree CWT comprises of two parallel wavelet filter bank trees that contain carefully designed filters of different delays that minimize the aliasing effects due to downsampling[5]. The dual-tree CDWT of a signal x(n) is implemented using two critically-sampled DWTs in parallel on the same data, as shown in Fig. 3. The transform is two times expansive because for an N-point signal it gives 2N DWT coefficients. If the filters in the upper and lower DWTs are the same, then no advantage is gained. So the filters are designed in a specific way such that the subband signals of the upper DWT can be interpreted as the real part of a complex wavelet transform and subband signals of the lower DWT can be interpreted as the imaginary part. When designed in this way the DT CDWT is nearly shift invariant, in contrast to the classic DWT.

3. TRANSLATION INVARIANCE BY PARALLEL FILTER BANKS

The orthogonal [8] two-channel filter banks with analysis low-pass filter given by the *z*-transform $H_0(z)$, analysis highpass filter $H_1(z)$ and with synthesis filters $G_0(z)$ and $G_1(z)$ is shown in figure.1

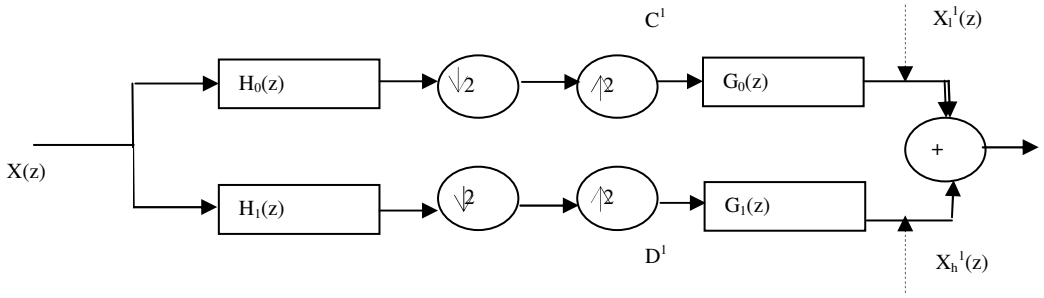


Figure 1: DWT Filter Bank

For an input signal X(z), the analysis part of the filter bank followed by upsampling produces the low-pass and the high-pass coefficients respectively, and decomposes the input signal into a low frequency part $X_l^1(z)$ and a high frequency part $X_h^1(z)$, the output signal is the sum of these two components.

$$C^1(z^2) = \left(\frac{1}{2}\right)\left(X(z)H_0(z) + X(-z)H_0(-z)\right) \quad (1)$$

$$D^1(z^2) = \left(\frac{1}{2}\right)\left(X(z)H_1(z) + X(-z)H_1(-z)\right) \quad (2)$$

$$Y(z) = X_l^1(z) + X_h^1(z) \quad (3)$$

Where

$$X_l^1(z) = C^1(z^2)G_0(z)$$

$$= \left(\frac{1}{2}\right)\left(X(z)H_0(z)G_0(z) + X(-z)H_0(-z)G_0(z)\right) \quad (4)$$

$$X_h^1(z) = D^1(z^2)G_1(z)$$

$$= \left(\frac{1}{2}\right)\left(X(z)H_1(z)G_1(z) + X(-z)H_1(-z)G_1(z)\right) \quad (5)$$

This decomposition is not shift invariant due to the terms in X(–z) of eqn 4 and eqn 5, respectively, which are introduced by the downsampling operators. If the input signal is shifted, for example $z^{-1}X(z)$, the application of the filter bank results in the decomposition

$$z^{-1}X(z) = \tilde{X}_l^1(z) + \tilde{X}_h^1(z) \quad (6)$$

For an input signal $z^{-1}X(z)$ we have

$$C^1(z^2) = \left(\frac{1}{2}\right)\left(z^{-1}X(z)H_0(z) + (-z^{-1})X(-z)H_0(-z)\right) \quad (7)$$

and

$$\tilde{X}_l^1(z) = \left(\frac{1}{2}\right)\left(X(z)H_0(z)G_0(z) - X(-z)H_0(-z)G_0(z)\right) \quad (8)$$

and similarly for the high-pass part, which of course is not the same as $z^{-1}X_l^1(z)$ if we substitute for z^{-1} in eqn 4. From this calculation it can be seen that the shift dependence is caused by the terms containing X(–z), the aliasing terms.

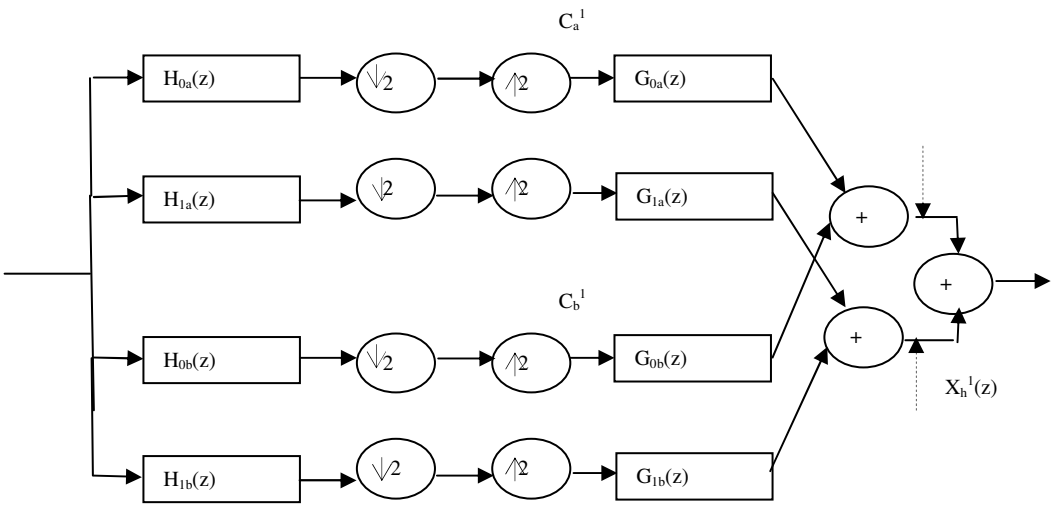


Figure 2: One level complex dual tree.

One possibility to obtain a shift invariant decomposition can be achieved by the addition of a filter bank to figure 1 with shifted analysis filters $z^{-1}H_0(z)$, $z^{-1}H_1(z)$ and synthesis filters $zG_0(z)$, $zG_1(z)$ and subsequently taking the average of the lowpass and the highpass branches of both filter banks as shown in figure 2.

If we denote the first filter bank by index a and the second one by index b then this procedure implies the following decomposition

$$X(z) = X_1^1(z) + X_{II}^1(z) \quad (9)$$

where for the lowpass channels of tree a and tree b we have

$$X_1^1(z) = \left(\frac{1}{2}\right)\left(C_a^1(z^2)G_{0a}(z) + C_b^1(z^2)G_{0b}(z)\right)$$
$$= \left(\frac{1}{2}\right)X(z)\,H_0(z)G_0(z) \quad (10)$$

and similarly for the high-pass part. The aliasing term containing X(-z) in X_l^1 has vanished and the decomposition becomes indeed shift invariant.

Using the same principle for the design of shift invariant filter decomposition, Kingsbury suggested in [4] to apply a 'dual-tree' of two parallel filter banks are constructed and their bandpass outputs are combined. The structure of a resulting analysis filter bank is shown in Fig. 3, where index *a* stands for the original filter bank and the index *b* is for the additional one. The dual-tree complex DWT of a signal x(n) is implemented using two critically-sampled DWTs in parallel on the same data.

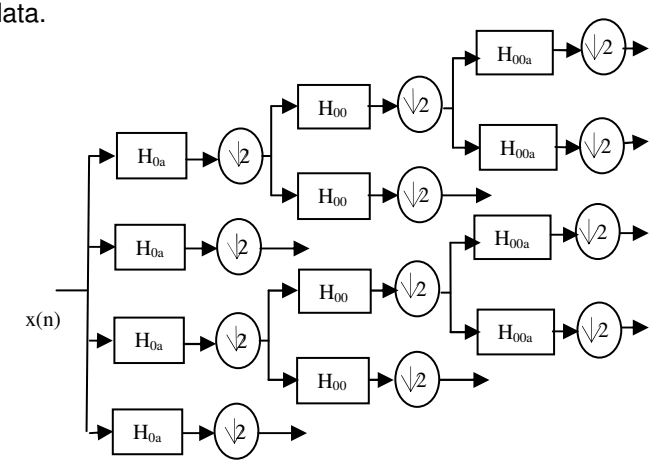


Figure 3: Three level Complex dual tree

In one dimension, the so-called dual-tree complex wavelet transform provides a representation of a signal x(n) in terms of complex wavelets, composed of real and imaginary parts which are in turn wavelets themselves. In fact, these real and imaginary parts essentially form a quadrature pair.

H_{0a}	H_{1a}	H_{0b}	H_{1b}
0	0	0.01122679	0
-0.08838834	-0.01122679	0.01122679	0
0.08838834	0.01122679	-0.08838834	-0.08838834
0.69587998	0.08838834	0.08838834	-0.08838834
0.69587998	0.08838834	0.69587998	0.69587998
0.08838834	-0.69587998	0.69587998	-0.69587998
-0.08838834	0.69587998	0.08838834	0.08838834
0.01122679	-0.08838834	-0.08838834	0.08838834
0.01122679	-0.08838834	0	0.01122679
0	0	0	-0.01122679

TABLE 1: First Level DWT Coefficients

The dual-tree CDWT uses length-10 filters [6], the table of coefficients of the analyzing filters in the first stage is shown in table 1 and the remaining levels are shown in table 2. The reconstruction filters are obtained by simply reversing the alternate coefficients of the analysis filters.

To extend the transform to higher-dimensional signals, a filter bank is usually applied separably in all dimensions. To compute the 2D CWT of images these two trees are applied to the rows and then the columns of the image as in the basic DWT.

Tree a		Tree b	
H_{00a}	H_{01a}	H_{00b}	H_{01b}
0.03516384	0	0	-0.03516384
0	0	0	0
-0.08832942	-0.11430184	-0.11430184	0.08832942
0.23389032	0	0	0.23389032
0.76027237	0.58751830	0.58751830	-0.76027237
0.58751830	-0.76027237	0.76027237	0.58751830
0	0.23389032	0.23389032	0
-0.11430184	0.08832942	-0.08832942	-0.11430184
0	0	0	0
0	-0.03516384	0.03516384	0

TABLE 2: Remaining Levels DWT Coefficients

This operation results in six complex high-pass subbands at each level and two complex low-pass subbands on which subsequent stages iterate in contrast to three real high-pass and one real low-pass subband for the real 2D transform. This shows that the complex transform has a coefficient redundancy of 4:1 or 2m : 1 in m dimensions. In case of real 2D filter banks the three highpass filters have orientations of 0^{θ}, 45^{θ} and 90^{θ}, for the complex filters the six subband filters are oriented at angles $\pm 15^{\theta}, \pm 45^{\theta}, \pm 75^{\theta}$. This is shown in figure 4.

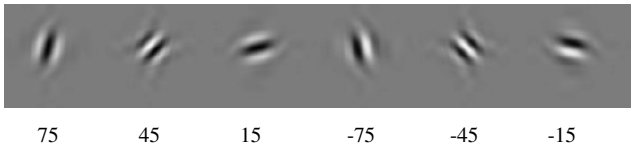


Figure 4: Complex filter response showing the orientations of the complex wavelets

The CDWT decomposes an image into a pyramid of complex subimages, with each level containing six oriented subimages resulting from evenly spaced directional filtering and subsampling, such directional filters are not obtainable by a separable DWT using a real filter pair but complex coefficients makes this selectivity possible.

4. RESULTS AND DISCUSSION

The shift invariance and directionality of the CWT may be applied in many areas of image processing like denoising, feature extraction, object segmentation and image classification. Here we shall consider the denoising example. For denoising a soft thresholding method is used. The choice of threshold limits σ for each decomposition level and modification of the coefficients is defined in the following equation.

$$\bar{c}_s(k) = \{\operatorname{sinc}(k)(|c(k)| - \sigma) \text{ if } |c(k)| > \sigma$$
$$\bar{c}_s(k) = 0 \text{ if } |c(k)| \leq \sigma \quad (11)$$

To compare the efficiency of the DWT with the basic DWT the quantitative mean square error (MSE) is used. In all cases the optimal thresholds points σ were selected to give the minimum square error from the original image, showing a great effectiveness in removing the noise compared to the classical DWT as shown in table 3.

Figure 5: (a) Input Image, (b) Denoised with real CWT,(c) Denoised with dual tree CWT

From figure 5(b) it may be seen that DWT introduces prominent worse artifacts, while the DT CWT provides a qualitatively restoration with a better optimal minimum MSE error.

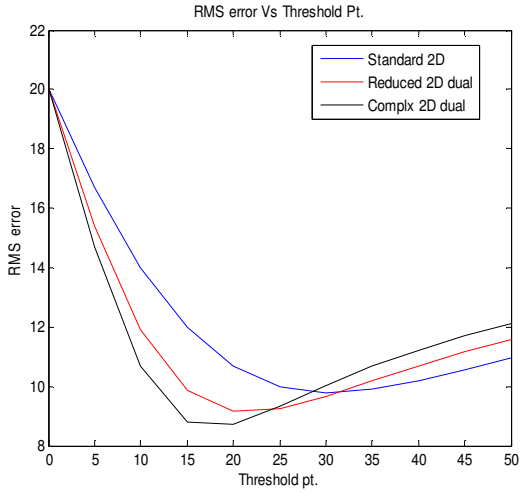


Figure 6: Optimal threshold points for the three different methods

The table 3 gives the comparison between the various methods in terms of their Mean Square Error (MSE) and Signal-to-Noise Ratio (SNR) Values.

Type of method	MSE	SNR [dB]
noisy image	0.0418	20.8347
DWT	0.0262	25.4986

real CWT	0.0255	25.7601
CWT	0.0240	26.3751

TABLE 3: Mean Square Error (MSE) and Signal – to – Noise Ratio (SNR) Values

The DT CWT is shift invariant and forms directionally selective diagonal filters. These properties are important for many applications in image processing including denoising, deblurring, segmentation and classification. In this paper we have illustrated the example of the application of complex wavelets for the denoising of Lena images. To obtain further improvements, it is also necessary to develop principled statistical models for the behavior of features under addition of noise, and their relationship to the uncorrupted wavelet coefficients. This remains to be done.

5. REFERENCES

[1] R.Anderson, N.Kinsbury, and J. Fauqueur. 'Determining Multiscale Image feature angles from complex wavelet phases' In international conference on Image processing (ICIP), September 2005

[2] C. Kervrann and J.Boulanger, " Optimal spatial adaptation for patch based image denoising," *IEEE* Trans. Image Process., Vol. 15, no.10, pp. 2866 – 2878, Oct. 2006.

[3] S.G. Chang, Y.Bin, and M.vetterli, "Adaptive wavelet thresholding for image denoising and compression", *IEEE* Transaction on image processing., Vol.9, No.9, pp.1532 – 1546, Sep.2000.

[4] Ming Zhang and Bahadir K. Gunturk, " Multiresolution Bilateral Filtering for Image Denoising" *IEEE* Transaction on Image processing ,Vol.17, No.12 Dec. 2008.

[5] N. G. Kingsbury," The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters", In the Proceedings of the *IEEE* Digital Signal Processing Workshop, 1998.

[6] N. G. Kingsbury,"Image processing with complex wavelets", Phil. Trans. Royal Society London – Ser. A., vol.357, No.1760, pp. 2543 – 2560,Sep 1999.

[7] N. G. Kingsbury,"A dual-tree complex wavelet transform with improved orthogonality and symmetry properties", In Proceedings of the *IEEE* Int. Conf. on Image Proc. (ICIP), 2000.

[8] J.Scharcanskim C.R.Jung and R.T.Clarke, "Adaptive image denoising using scale and space consistency", *IEEE* Transaction on image processing ., Vol.11, No.9,pp.1092 – 1101,Sep.2002.

[9] J. Neumann and G. Steidl, "Dual–tree complex wavelet transform in the frequency domain and an application to signal classification", International Journal of Wavelets, Multiresolution and Information
Processing IJWMIP, 2004.

[10] K.Hirakawa and T.W. Parks, "Image denoising using total least squares," *IEEE* Trans. Process., vol. 15, No. 9, pp. 2730 – 2742, Sep. 2006.

[11] J. K. Romberg, H. Choi, R. G. Baraniuk, and N. G. Kingsbury,"Hidden Markov tree models for complex wavelet transforms", Tech. Rep., Rice University, 2002.

[12] L. Sendur and I.W. Selesnick, " Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," IEEE Transactions on signal processing, vol.50, no,11,pp.2744-2756, November 2002.

Enhanced Morphological Contour Representation and Reconstruction using Line Segments

Santhosh P Mathew mathewsantosh@yahoo.com
Professor/Computer Science & Engineering
Saintgits College of Engineering
Kottayam, Kerala, India PIN 686 532

Philip Samuel philipsamu@yahoo.com
Reader & Head/Information Technology
Cochin University of Science & Technology
Cochin, Kerala, India PIN 682 022

Justin Varghese justinv@saintgits.org
Professor & Head/Computer Science & Engineering
Infant Jesus College of Engineering
Tirunelveli, Tamilnadu, India PIN 628 851

Saudia Subhash saudias@yahoo.com
Lecturer/Center for Information Technology & Engineering
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India PIN 627 012

Abstract

The paper proposes an enhanced morphological contour/edge representation algorithm for the representation of 2D binary shapes of digital images. The concise representation algorithm uses representative lines of different sizes and types to cover all the significant features of the binary contour/edge image. These well characterized representative line segments, which may overlap among different types, take minimum representative points than that of most other prominent shape representation algorithms including MST and MSD. The new algorithm is computationally efficient than most other algorithms in the literature and is also capable of approximating edge images. The approximated outputs produced by the proposed algorithm by using minimal number of representative points are more natural to the original shapes than that of MST and MSD.

Keywords: closing, dilation, erosion, opening, representation

1. INTRODUCTION

Humans recognize objects mainly on the basis of their shapes and so its representation is an important issue of concern in image processing and computer vision to provide the foundation for image coding [4], shape matching and object recognition [1], content-based video processing [12], [13], image data retrieval [14], character recognition [2], automatic visual inspection and medical diagnostics [3]. A good shape representation algorithm should be precise, well defined, accurate, complete, easily reconstruct able and computationally efficient.

Targeting these requirements, through these years a number of representation algorithms have evolved focusing the shape characteristics [4]-[5], [7], [9], [15]-[16]. Charif and Schonfeld [17] made a thinning based shape representation algorithm. Multiple structuring elements and minimal enclosing structure elements are proposed in the scheme of Pitas et. al [7]. Algorithms including parity check [8] and chain code [18] are also proposed for representing shape images. But the conventional parity check is not reversible unless the shapes are very simple. Maragos attempted to

represent an image as a minimal union of translated and scaled patterns from a finite basic pattern class. Y. M. Y. Hasan and L. J. Karam [10] proposed a reversible contour representation algorithm in which shape images is decomposed into residual and ambiguous contours but the algorithm is too complex in its representation and reconstruction phases.

Many basic morphological shape representations have also been employed [5], [11], [16] on binary shape images. The morphological skeleton transform (MST) [4] represents shape as a union of all maximal disks but the disks tend to be larger in size and highly overlapping. As a variation to the morphological skeleton transform a decomposition scheme is proposed in [16]. The morphological shape decomposition (MSD) [5] decomposes the binary shape into non-overlapping components of overlapping disks each, which tend to be smaller affecting the reconstruction efficiency. Though these algorithms could meet some of the basic requirements of a typical representation algorithm, they lose their focus in parallely meeting other vital aspects like computational efficiency, lesser bit-rate, minimal representative points and so on.

In this paper, we propose a new representation algorithm where a given shape is represented as a union of a number of overlapping lines contained in the given shape. The paper is organized in 6 sections. Section 2 gives an overview of the fundamental morphological operations involved in the shape representation scheme and explains the features of MSD and MST. Section 3 explains the features and improved characteristics of the proposed shape representation and reconstruction algorithms. The simulation results are provided in Section 4. Further Scope is explored in section 5 and Conclusions are finally made in Section 6.

2. MORPHOLOGICAL REPRESENTATION OPERATORS

The binary shapes extracted from images are represented before being stored in the knowledge base. The shape so represented is reconstructed and filled to be matched or recognized later on with a suitable input as shall be the requirement. The shape representation scheme passes the image to be represented through a morphological pre-processing set-up. It is then subjected to the reversible shape representation algorithms. The general morphological operations involved in these steps, the features of the top-ranking internal shape representation algorithms, MST and MSD are highlighted in this section

2.1 Basic Morphological Operations

In the morphological analysis of binary images, a *2D* image is defined as a subset of *2D* Euclidean space $R \times R$ or its digital equivalent $Z \times Z$ .For a digital image $A \in Z \times Z$ and a point, $b \in Z \times Z$,the translation of *A* by *b* is defined as

$$(A)_b = \{a + b \mid a \in A\} \quad (1)$$

The morphological dilation of the image *A* by the structuring element (SE), *B* expands the image while morphological erosion of *A* by *B* shrinks the image. They are defined respectively in (2) and (3).

$$A \oplus B = \bigcup_{b \in B} (A)_b \quad (2)$$

$$A \ominus B = \bigcap_{b \in B} (A)_{-b} \quad (3)$$

Opening of the binary image *A* by structuring element *B* denoted as $A \circ B$, is defined as

$$A \circ B = (A \ominus B) \oplus B \quad (4)$$

Closing of the binary image *A* by structure element *B* denoted as $A \bullet B$, is defined as

$$A \bullet B = (A \oplus B) \ominus B \quad (5)$$

2.2 The MST and MSD

The morphological skeleton transform (MST) [4] is a simple and efficient shape representation scheme where a binary shape *X* is represented as union of all the maximal disks contained in it. These maximal disks of different sizes may overlap with each other as can be determined directly from the shape.

$$X = \bigcup_{i=0}^{N} S_i \oplus iB \quad (6)$$

Where

$$S_i = (X \ominus iB) \setminus ((X \ominus iB) \circ B) \quad (7)$$

'\ ' is the logical difference operator, *N* is the largest integer such that $X \ominus NB \neq \phi$ and $iB = B \oplus B \oplus \ldots \oplus B$ *i times* is a disk of size *i.* The skeleton subset *Si* contains the centers of all maximal

inscribable disks of size *i.* A maximal disk cannot be contained in a representative disk of larger size and the maximal disks of different sizes may overlap. So, in general

$$(S_i \oplus iB) \cap (S_i \oplus jB) \neq \phi, \quad for\ i \neq j \quad (8)$$

Another interpretation on these skeleton subsets is that *Si* is the set of centers of all disks of size *i* in *X* that are not contained in any representative (maximal) disks of larger sizes, the shape can be reconstructed as

$$X = ((((S_N \oplus B) \cup S_{N-1}) \oplus B) \cup S_{N-2}) \oplus \ldots \cup S_0 \quad (9)$$

Therefore, *N* dilations with *B* will be needed to reconstruct *X* from all the skeleton subsets. The MST usually uses comparatively fewer numbers of larger, overlapping disks to represent a given shape. There is no simple and obvious way of combining representative disks into more meaningful shape components due to heavy overlapping.

The morphological shape decomposition (MSD) [5] decomposed a binary shape into a union of certain non overlapping disks contained in the shape with minimum morphological operations. A binary shape *X* is represented by the MSD as a union of certain disks contained in *X* .

$$X = \bigcup_{i=0}^{N} L_i \oplus iB \quad (10)$$

Where $L_N = X \ominus NB$ and

$$L_i = \left(X \setminus \left(\bigcup_{j=i+1}^{N} L_j \oplus jB \right) \right) \ominus iB, \quad 0 \leq i < N \quad (11)$$

Again, *N* is the largest integer such that $X \ominus NB \neq \phi$. The sets of centers of representative disks of different sizes $L_N, L_{N-1} \ldots L_0$ are determined in the order given and then

$$(L_i \oplus iB) \cap (L_i \oplus jB) = \phi, \quad for\ i \neq j \quad (12)$$

The centers, *Li* of all the disks of size *i* contained in *X* that do not intersect with any representative disks of larger sizes are determined by removing all the representative disks of larger sizes from the given shape and then finding all the centers of representative disks of size *i* in the remaining areas. Overlapping between disks of the same size still exists. Similar to the MST, the original image *X* can be reconstructed using *N* dilations with *B* such that

$$X = ((((L_N \oplus B) \cup L_{N-1}) \oplus B) \cup L_{N-2}) \oplus \ldots . \cup L_0 \quad (13)$$

Where some of the *Li's* can be empty.

Thus shape can be easily represented by the components generated by MSD using the larger number of smaller, non-overlapping disks. Though the level of redundancy and the reconstruction cost when compared to MST are less, the numbers of disks used by the MSD are higher.

3. PROPOSED REPRESENTATION ALGORITHM

The proposed representation algorithm represents shape as a union of a number of representative lines contained in its boundary extracted image. This algorithm is efficient than many other shape representation algorithms, represents all type of images, even the edge extracted images with less number of representative points. It is featured towards cost effective representation of shapes, reduced representation error and least burden in pattern recognition applications. The operations involved in representation and reconstruction of the shape of the object is schematically shown in Figure 1.

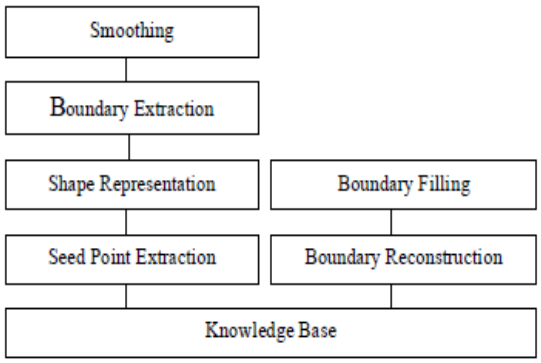


Figure 1. Block diagram of Representation Scheme

3.1. Boundary Extraction and Smoothing

The noise manifestations in the shape of the object is smoothened by performing closing after opening the input binary shape images and is represented as

$$(A \circ B) \bullet B \quad (14)$$

In the proposed representation scheme, the internal boundary, *K* of the shape, *A* or the set of object pixels that have at least one non-object neighbor is extracted from the morphological erosion gradient (EG). $K = EG(A, B) = A \setminus A \ominus B$ (15)

where *B* is a *3x3 4* or *8* connected flat structural element. The boundary extracted image is then represented by proposed representation algorithm.

3.2. Proposed Representation Algorithm

The new approach uses four types of structural elements which are shown in Figure 2. These structural elements represent the boundary extracted binary shape image, *K* as a union of maximal representative lines such that

$$K = \bigcup_{j=1}^{4} \bigcup_{i=0}^{N} C_{ij} \oplus iB_j \quad (16)$$

where $0 \leq i \leq N$ and *j=1....4* .

$$C_{ij} = \left(K \setminus \left(\bigcup_{k=i+1}^{N} C_{kj} \oplus kB_j \right) \right) \ominus iB_j \quad (17)$$

and

$$C_{Nj} = K \ominus NB_j$$

If there arise any two points x_{ij}, y_{ij} such that $x_{ij}, y_{ij} \in C_{ij}$

and if $(x_{ij} \oplus iB_j) \cap (y_{ij} \oplus iB_j) \neq \phi$

one of them is arbitrarily selected and other is rejected. The sets of centers of representative disks of different sizes $C_{Nj}, C_{N-1j}, .., C_0$ must be determined in the order given, then

$$(C_{ij} \oplus iB_j) \cap (C_{ki} \oplus kB_i) = \phi, \quad for\ j = l \quad (18)$$ and

$$(C_{ij} \oplus iB_j) \cap (C_{ki} \oplus kB_i) \neq \phi, \quad for\ j \neq l \quad (19)$$

This means that overlapping of representative lines is allowed only between lines of different types.

(a)B_1 (b)B_2 (c)B_3 (d)B_4

Figure 2. Four basic structural elements

Here, in this new approach the representative points used to represent a shape are very less when compared with MSD and MST. The four basic structural elements when dilated twice with themselves can be written as $3B_1$, $3B_2$, $3B_3$, $3B_4$ and are shown in Figure 3.

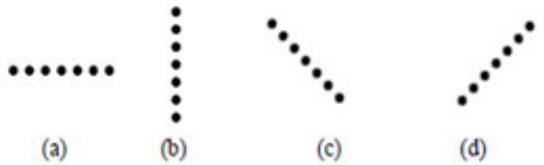


Figure 3. Structural elements (a)$3B_1$ (b)$3B_2$ (c)$3B_3$ (d) $3B_4$

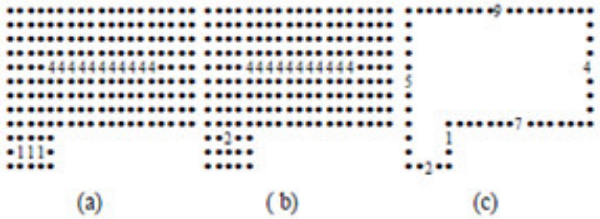


Figure 4. Representation example: (a) MSD, (b) MST, and (c) Proposed algorithm

In this implementation, the unit line B_1 is defined as B_1 the size two line $2B_1$ is defined as $B_1 \oplus B_1$ and the size three line $3B_1$ is defined as $(B_1 \oplus B_1) \oplus B_1$. In general, a line of size *i* and type *j* is defined

$$iB_j = (i-1)B_j \oplus B_j, where\ 1 \le j \le 4 \quad (20)$$

In this proposed representation algorithm, overlapping is allowed only between representative line segments of different types. A line is selected if it matches with some parts of the given boundary image. Compared to MST and MSD, the overlapping level is much lower since the overlapping between two lines segments is always reduced to a single point. Therefore the redundancy level is much lower when compared to MST and MSD; the moderate overlapping reduces the number of representative points needed. Consider the image in Figure 4. The MST uses twelve representative points, the MSD uses a maximum number of fourteen points and the proposed method uses the fewest number of six points to represent it. For an image *X* , the proposed representation algorithm produces a sequence of center point sets: C_{Ni}, C_{N-1j},..,C_0 .

3.3. Representation Seed Point Extraction

If N_{ijk} is the four/eight neighbors of C_{ij} then

$$Seed_{ij} = \{k / for\ any\ N_{ijk} \subset X \Theta SB\} \quad (21)$$

where *SB* is a 4 or 8 connected structural element which depends upon the seed filling algorithm. The cardinality $\left(CARD\left(Seed_{ij}\right)\right)$ on being greater than one, any one of them is arbitrarily selected as the seed point, $Seed_{i,j}$. The final representative set, *RP* for the proposed algorithm is a combination of C_{ij} and $Seed_{ij}$. i.e.,

$$RP_{ij} = \{ (RC_{ij}, R_k) / RC_{ij} = C_{ij}\ and\ R_k = Seed_{ij} \} \quad (22)$$

The seed point $Seed_{ij}$ is stored to refer it as one of the constraints in the representative table which ranges between zero to seven and further the region filling in the reconstruction phase of the algorithm uses these seed point indices to reconstruct the original shape.

3.4. The Reconstruction Algorithm

The representative point set, *RP* of the boundary extracted binary shape generated by the representation algorithm is used by the proposed lossless reconstruction algorithm. Similar to the MST and MSD, the boundary image *K* can be reconstructed using *N* dilations of *RCij* with *Bj*

$$K = \bigcup_{j=1}^{4} (((C_{N,j} \oplus B_j) \cup C_{N-1,j}) \oplus B_j) \oplus ... \cup C_0 \quad . \quad (23)$$

for all *j=1…4*

N ×4 dilations are needed to reconstruct the contour/ boundary of the given shape. This reconstruction algorithm is much faster than most other algorithms in the literature since the overlapping level is reduced to some points of the lines. Looking into the R_k field of the representative point set, *RP* the initial seed point for region filling is identified and is fed to the appropriate seed-filling

algorithm for reconstructing the original shape. In this scheme the traditional area filling algorithm is used.

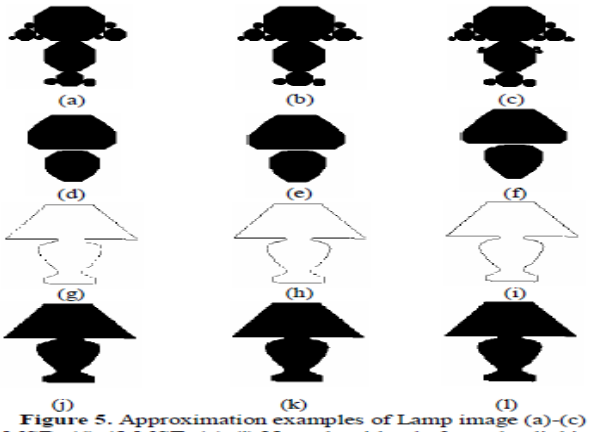


Figure 5. Approximation examples of Lamp image (a)-(c) MSD, (d)-(f) MST, (g)-(i) New algorithm before edge linking and filling, (j)-(l) New algorithm after edge linking and filling. 1st, 2nd and 3rd columns respectively use 80, 100 and 120 RPs.

4. Experimental Results and Simulation Analysis

The proposed representation algorithm is tested on a variety of binary shape images of varied sizes and complexity of which Teapot, Lamp, Telephone, Temple, Puzzle, Letters, Digits, Lena, House, Tree, Building are used here for subjective and objective comparisons. The approximation examples of Lamp image shown in Figure 5. confirm the improved subjectiveness and efficiency of the proposed representation algorithm over the MSD and MST. The reconstructed images generated by the proposed algorithm are shown through Figure.6 (g) to Figure.6 (i). These outputs are edge liked by using the effective edge linking algorithm proposed in [6] and are then filled by the traditional flood filling algorithm as are shown in (j) to (l) of Figure 5. The reconstructed shapes from the new algorithms are more equivalent to the original binary shapes though different number of RPs are used by the three representation schemes; 80, 100 and 120 respectively as are shown in the first three columns of Figure 5.

Table 1. Representative Points (RP) and Computation time (CT) in Seconds of MST, MSD, J. Xu Approach and Proposed Algorithms.

Images	MST		MSD		J. Xu Approach		Proposed Algorithm	
	RP	CT	RP	CT	RP	CT	RP	CT
Teapot	188	49.31	349	41.62	217	45.83	141	26.7
Lamp	238	51.74	439	46.37	248	47.62	140	28.3
Telephone	300	62.98	587	54.89	305	59.37	189	32.4
Temple	1715	85.23	1837	70.42	1879	78.91	697	37.1
Puzzle	286	57.70	450	43.25	281	48.48	281	27.1
Letters	567	65.49	585	55.27	460	60.62	460	29.6
Digits	731	75.25	717	65.85	613	71.47	613	32.2
Lenna	1501	84.32	1600	71.36	1347	78.27	712	35.8
House	1005	70.88	1101	60.85	963	67.68	963	30.1
Tree	1919	97.73	1956	75.38	1658	84.94	952	36.1
Building	2346	98.16	2876	85.32	2149	91.82	1476	46.1

(a) (b) (c)

Figure 6. Approximation example (a) Canny edge detected Image of cameraman. (b) Image reconstructed by new algorithm using 800 RP (c) Image reconstructed by new algorithm using 1000 RP

The proposed representation algorithm is also capable of approximating edge images. This can be observed from Figure.6 where the edge detected cameraman image is approximated by the proposed algorithm with fewer representative points. Figure.6 (b) and (c) respectively are the images reconstructed by the proposed algorithm using800RP and 1000RP. Table.1 shows the number of Representative Points (RP) and the Computation time(CT) in Seconds of MST, MSD and the Proposed Algorithms. The representative points used by the proposed algorithm for representing the binary shape is far less when compared to the conventional MST, MSD, J. Xu Approach [11] as recorded in the Table 1 which also shows the improved computational efficiency of the proposed representation algorithm over the MST and the MSD.

5. CONSLUSION & FUTURE WORK

In this paper, a new morphological shape and edge representation and reconstruction algorithm is proposed which represents a boundary extracted binary shape as a union of a number of maximal lines contained in the shape. The experimental results have shown that this algorithm performs better than the more prevalent morphological boundary representation algorithms, the MST and the MSD in terms of representative points and computational efficiency. In case of multi-contour images, multi-component images, there will be many seed points to be stored in the representation. In case of round or circular shapes, the number of line segments will be many to give a compact representation of the contour. Attempts to refine the algorithm to address the above issues can be made.

6. REFERENCES

1. P. E. Trahanias, "*Binary shape recognition using the morphological skeleton transform*," Pattern Recognition, 25(11):1277–1288, 1992.

2. P. Yang and P. Maragos, "*Morphological Systems for Character Recognition*", In Proceedings of the IEEE Int'l Conf. Acoustics and Speech and Signal Processing, 1993.

3. G.K. Matsopoulos and S. Marshall, "*Use of Morphology Image Processing Techniques for the Measurement of Fetal Head from Ultrasound Images*", Pattern Recognition, 27(10):1,317-1,324, 1994.

4. P. A. Maragos and R. W. Schafer, "*Morphological skeleton representation and coding of binary images*", IEEE Transactions on Acoustics, Speech & Signal Processing,34(5):1228–1244, 1986.

5. I. Pitas and A. N. Venetsanopoulos, "*Morphological shape decomposition*," IEEE Transactions on Pattern Analysis & Machine Intelligence,12(1):38–45, 1990.

6. F.L.Miller, J.Maeda, H.Kubo, "*Template Based Method of Edge Linking Using a Weighted Decision*,"In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Japan, 1993.

7.I. Pitas and A. N. Venetsanopoulos, "*Morphological shape representation*," Pattern Recognition, 25(6):555–565, 1992.

8. B.D. Ackland and N. Weste, "*The Edge Flag Algorithm Fill Method for Raster Scan Display*", IEEE Transactions on Computers & Graphics,30(1):41-47, 1981.

9. J.M.Reinhardt and W. E. Higgins, "*Efficient morphological shape representation*," IEEE Transactions on Image Processing, 5(1): 89–101, 1996.

10. Y. M. Y. Hasan and L. J. Karam, "*Morphological reversible contour representation*," IEEE Transactions on Pattern Analysis & Machine Intelligence, 22(3):227–240, 2000.

11. Justin Varghese et al, "*An efficient Morphological Reversible Contour/Edge Representation using overlapping Line Components*", In Proceedings of the International Conference on Emerging Trends in Engineering and Technology (ICETET), 2008

12. J. Xu, "*Efficient Morphological Shape Representation with Overlapping Disk Components*", IEEE Trans.Image Processing,10(9):1346- 1356, 2005.

13. R. S. Jasinschi and J. M. F. Moura, "*Content-based video sequence representation*," In Proceedings of the IEEE International Conference on Image Processing, 1995.

14. P. Salembier, P. Brigger, J. R. Casas, and M. Pardas, "*Morphological operators for image and video compression*", IEEE Transactions on Image Processing, 5(6):881–897, 1996.

15. G. Lu, "*An approach to image retrieval based on shape*," Journal of Information Science, 23(2):119–127, 1997.

16. J. Xu, "*Morphological decomposition of 2-D binary shapes into simpler shape parts*," Pattern Recognition Letters*,* 17(7):759–769, 1996.

17. A.C.P. Loui, A.N. Venetsanopoulos, and K.C. Smith, "*Morphological Autocorrelation Transform: A New Representation and Classification Scheme for Two- Dimensional Images*", IEEE Transactions on .Image Processing,1(7):337-353, 1992.

18. Raman Maini, Himanshu Aggarwal "*Study and Comparison of Various Image Edge Detection Techniques*" International Journal of Image Processing (IJIP), 3(1):1-11, 2009

19. M. Charif and D. Schonfeld, "*On the Invertability of Morphological Representation of Binary Images*", IEEE Transactions on Image Processing, 3(11): 847-849, 1994.

20. Z. Cai,"*Restoration of Binary Images Using Contour Direction Chain Codes Description*", Computer Vision, Graphics, and Image Processing, 41(1):101-106, 1988

21. Chandra Sekhar Panda, Srikanta Patnaik, "Filtering Corrupted Image and Edge Detection in Restored Grayscale Image Using Derivative Filters", International Journal of Image Processing (IJIP),3(3):105-119

DATA HIDING METHOD with HIGH EMBEDDING CAPACITY CHARACTER

Wen-Chung Kuo simonkuo@nfu.edu.tw
Department of Computer Science and Information Engineering, National Formusa University, Yunlin 632, Taiwan, R.O.C

Jiin-Chiou Cheng chiou@mail.stut.edu.tw
Department of Computer Science and Information Engineering, Southern Taiwan University, Tainan 710, Taiwan, R.O.C

Chun-Cheng Wang 96g0216@webmail.stut.edu.tw
Department of Computer Science and Information Engineering, Southern Taiwan University, Tainan 710, Taiwan, R.O.C

Abstract

Recently, the data hiding method based on the high embedding capacity by using improved EMD method was proposed by Kuo et al.[6]. They claimed that their scheme can not only hide a great deal of secret data but also keep high safety and good image quality. However, in their scheme, the sender and the receiver must share the synchronous random secret seed before they transmit the stego-image each other. Otherwise, they can not recover the correct secret information from the stego-image. In this paper we propose an improved scheme based on EMD and LSB matching method to overcome the above problem, in other words, the sender does not share the synchronous random secret seed the receiver before the stego-image is transmitted. Observing the experimental results, they show that our proposed scheme acquires high embedding capacity and acceptable stego-image quality.

Keywords: Data-hiding, Cover-image, Stego-image, EMD, LSB match method.

1. Introduction

With the rapid development of network technology, vast multimedia data would be communicated over the network. Although network transmission is convenient and fast, the multimedia data passing through the network is often attacked and tampered by malicious attackers. From the literatures many people are devoted to study the security for multimedia data. In general there are two methodologies to deal with such work: one is the cryptography and the other is steganography. Using the cryptography methodologies, the only specific user with the private key can decrypt the ciphertext when the plaintext is encrypted. An attacker cannot find out the content of message even though he gets the encryption message from the Internet. Nevertheless, the ciphertext will still be insecure if the private key is stolen or broken. Another way to promote the security of multimedia data is to hide secret data behind a meaningful image. The major goal of data hiding scheme is

not only to raise the hiding amount in the stego-image but also keep the quality of the stego-image. In the past literatures, many well data-hiding schemes had be suggested [4,6,9].

In 2006, an efficient embedding scheme based on the Exploiting Modification Direction (for short, EMD-scheme) was proposed by Zhang and Wang [9]. The scheme uses the relationship of adjacent pixels to embed the secret data. The secret data will be embedded within two adjacent pixels, that is, only one of two pixels in the EMD scheme – add one, subtract one, or stay the same. From a spatial point of view, two pixels just have five situations - moving upward, downward, left, right, or not moving at all. From their experimental simulations and discussions, the EMD-scheme can enhance the capacity of secret message and the quality of the stego-image. Recently, Lee et al. [4] proposed an improved data-hiding scheme, say LWC-scheme, which catches both of two adjacent pixels at a time and improves the possible situations from five to eight. As a result of LWC-scheme, it can promote the capacity 1.5 times approximately the former. Since the data embedding process uses the fixed evaluating parameters in both of EMD-scheme [9] and LWC-scheme [4], they will be cracked easily and leak the secret message within the stego-image while their technology are disclosed. Therefore, some concerns about the security issues will be considered. Later, Kuo et al. (for short KWSK-scheme) [6] proposed two high capacity EMD data hiding techniques with changing-evaluating-value to improve the shortcoming of above schemes, in other words, the stego-images will still be safe even when it publishes the embedding formulas. According to KWSK-scheme, they used the synchronous generator of random numbers to minimize the possibility of message disclosure and improve the lack of open method but there is an open problem of synchronization of random seeds before the stego-image is transmitted between the sender and the receiver. In this paper, we will propose an improvement scheme based on EMD and LSB matching method to overcome the synchronization problem, in other words, the sender does not send the synchronous random secret seed to the receiver before the stego-image is transmitted. According to the experimental simulations and discussions, we show that the proposed scheme still keeps high safety and good image quality.

The rest of this paper is organized as follows. In Section 2, we will introduce the EMD-method, LSB matching method and LWC-scheme briefly. Then, we will propose the improvement scheme to overcome the synchronization problem and give the experimental result in Section 3 and Section 4, respectively. Finally, conclusions will be drawn in the Section 5.

2. REVIEW THE DATA HIDING SCHEME WITH HIGH EMBEDDING CAPACITY TECHNIQUES

2.1. The Exploiting Modification Direction Method

In 2006, Zhang and Wang [9] used the relationship of adjacent pixels to promote the data embedding scheme. In their method, they transfer the secret message into (2n+1)-ary system and then embed the modified secret message into a group of n pixels in cover image by using the following equation:

$$f(g_1, g_2, \cdots g_n) = \left[\sum_{i=1}^{n} (g_i \cdot i) \right] \bmod (2n+1) \quad (1)$$

g_i is the i-th value of pixel and n is the number of pixels. Due to the limit of paper page, we cannot explain their embedding and extracting procedures in detail here. For more details about those methods, the reader can refer to the Ref. [9].

2.2. The High Embedding Capacity by Improving Exploiting Modification Direction (EMD)

According to Lee et al.'s analysis, they find only five situations - moving upward, downward, left, right, or not moving at all to embed the secret data into two adjacent pixels by using the EMD scheme. To elevate the capacity of EMD-scheme, Lee et al. improve the number of variable situations from five to eight and then propose a steganographic scheme [4] with high embedding capacity in 2007. Here, we just only describe the embedded procedure in LWC-scheme as following steps:

Step 1. Transfer the secret message to message s, which is 8-ary system.

Step 2. Take two adjacent pixels (X, Y) as a group and perform the following extraction process,

$$f_e(X, Y) = (X \times 1 + Y \times 3) \bmod 8 \quad (2)$$

Step 3. Adjust (X, Y) according to the following rule:

(3-1) If $s = f_e(X, Y)$, $X = X$, $Y = Y$.

(3-2) If $s = f_e(X+1, Y)$, $X = X+1$.

(3-3) If $s = f_e(X\text{-}1, Y)$, $X = X\text{-}1$.

(3-4) If $s = f_e(X, Y\text{+}1)$, $Y = Y\text{+}1$.

(3-5) If $s = f_e(X, Y\text{-}1)$, $Y = Y\text{-}1$.

(3-6) If $s = f_e(X\text{+}1, Y\text{+}1)$, $X = X\text{+}1$, $Y = Y\text{+}1$.

(3-7) If $s = f_e(X\text{+}1, Y\text{-}1)$, $X = X\text{+}1$, $Y = Y\text{-}1$.

(3-8) If $s = fe(X\text{-}1, Y\text{+}1)$, $X = X\text{-}1$, $Y = Y\text{+}1$.

Therefore, the stego-image may be generated as soon as the above modified pixels are embedded into the original image. The secret data can be extracted by using the extracting procedure when the particular user receives the stego-image.

2.3. The Data Hiding Scheme with High Embedding Capacity Based on General Improving EMD Method

Observing Eq. (1) in EMD-scheme and Eq. (2) in LWC-scheme, both uses the change of weight value along with modulus to fulfill the proper position for any point from surrounding area. Although there are outstanding contributions on the hiding capacities in the two techniques, the parameters of embedding function are fixed and their algorithms have to be kept. Otherwise, they will be cracked and the secret message in stego-image will leak out. In order to improve such shortcoming, Kuo et al. [6] proposed two high capacity EMD data hiding techniques with changing-evaluating-value, in other words, the stego-image will still be safe even though it publishes the embedding procedure. The KWSK-scheme is summarized as following:

Step 1. Transfer the secret message s, which is 8-ary system.

Step 2. Take two adjacent pixels (X, Y) as a group.

Step 3. Compute the value of the extract function f_{seed} with a random seed. The extract function is defined as Eq.3:

$$f_{seed}(X,Y) = (X \times a + Y \times b) \bmod 8 \tag{3}$$

Where the coefficients a and b are decided by the modular table shown in Fig.1. Compute the difference $d = (s - f_{seed}) \bmod 8$. Adjust ($X$, Y) by the modular table and the seed.

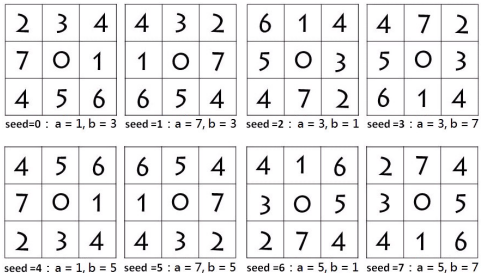


FIGURE 1: The modular tables for different weights.

Similar to the LWC-scheme, the stego-image is generated when the above modified pixels are embedded into the original image. Besides, the secret data will be extracted by using the extracting procedure when the particular user receives this stego-image. Form the experiment simulations, the KWSK-scheme [6] still maintains the high capacity and the image quality is almost the same as the LWC-scheme.

2.4. Least-Significant-Bit (LSB) Matching Method

In order to keep the embedding of the same amount of information as LSB matching and detect the secret data harder than the conventional LSB matching method, Mielikainen proposed a robust LSB matching method [5] in 2006. There are two major properties in his scheme as following:

$$f(l-1,n) \neq f(l+1,n), \forall l,n \in Z.$$

$$f(l,n) \neq f(l,n+1), \forall l,n \in Z.$$

Therefore, embedding message is performed for two pixels *X* and *Y* of a cover image at a time and then adjusting one pixel of the (*X*, *Y*) to embed two secret bits message s_1s_2. The embedding flowchart is shown in Fig.2 and the embedding procedure is described as following:

Step 1. If the LSB of *X* is the same as s_1, go to step 2.
Otherwise, go to step 3.

Step 2. If the value of $f(X,Y)$ is the same as s_2, do not change any pixel. Otherwise, the value of pixel *Y* is increased or decreased by 1.

Step 3. If the value of $f(X-1,Y)$ is the same as s_2, the value of pixel *X* is decreased by 1. Otherwise, the value of pixel *X* is increased by 1.

Where the function $f(X,Y)$ is defined as Eq.4:

$$f(X',Y') = LSB\left(\left\lfloor \frac{X'}{2} \right\rfloor + Y'\right) \tag{4}$$

Since this new LSB matching method just only increase or decrease 1 in two adjacent pixels, the difference of the two neighborhood pixel between cover image and stego-image is very small. Hence, it can keep high quality while hiding data.

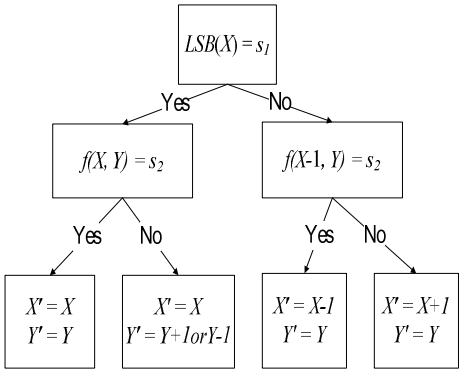


FIGURE 2: The LSB matching embedding procedure.

3. THE PROPOSED DATA HIDING SCHEME

By using more changes of weight, a robust embedded method can be proposed, which will enhance the security of the secret data within the stego-image[6]. Unfortunately, it needs to produce many random seeds before the stego-image will be processed and send them to the receiver for extracting secret message from the stego-image. How to transmit the additional information from sender to receiver is an important issue. However, such issue does not be discussed in [6]. In order to improve the lack, we will propose an efficient data hiding method based on the improved EMD and LSB matching methods, in which the seeds are embedded into stego-image at the same time and the receiver can extract these seeds and secret data from the stego-image.

3.1. The Embedding Secret Message Procedure

In our scheme, the embedding procedure is performed over three cover image pixels at a time. First, we embed the secret message by using the improvement EMD method, and then use the following functions f_1 and f_2 to embed the random seeds into the stego-image.

$$f_1(X,Y) = LSB(X+Y) \tag{5}$$

$$f_2(X,Z) = LSB(\left\lfloor \frac{X}{2} \right\rfloor + Z) \tag{6}$$

, where *X*, *Y*, *Z* are the first, second and third pixel in a group respectively. The flowchart of embedding message is shown in Fig.3. The steps are described as follows:

Step 1. Divide the modular tables into two groups G_0 and G_1 shown in Fig.4.
Step 2. Take three adjacent pixels (X, Y, Z) as a group.
Step 3. Let the result of a hash function $H(\cdot) = 0$ or 1. Compute the hash value $H(x_1||x_2||x_3||x_4||x_5||x_6)=i$ and decide to use group G_0 or G_1, where x_i is the ith bit of pixel X. Then, we also use the random generate to produce a seed $s_a \in \{0,1,2,3\}$.

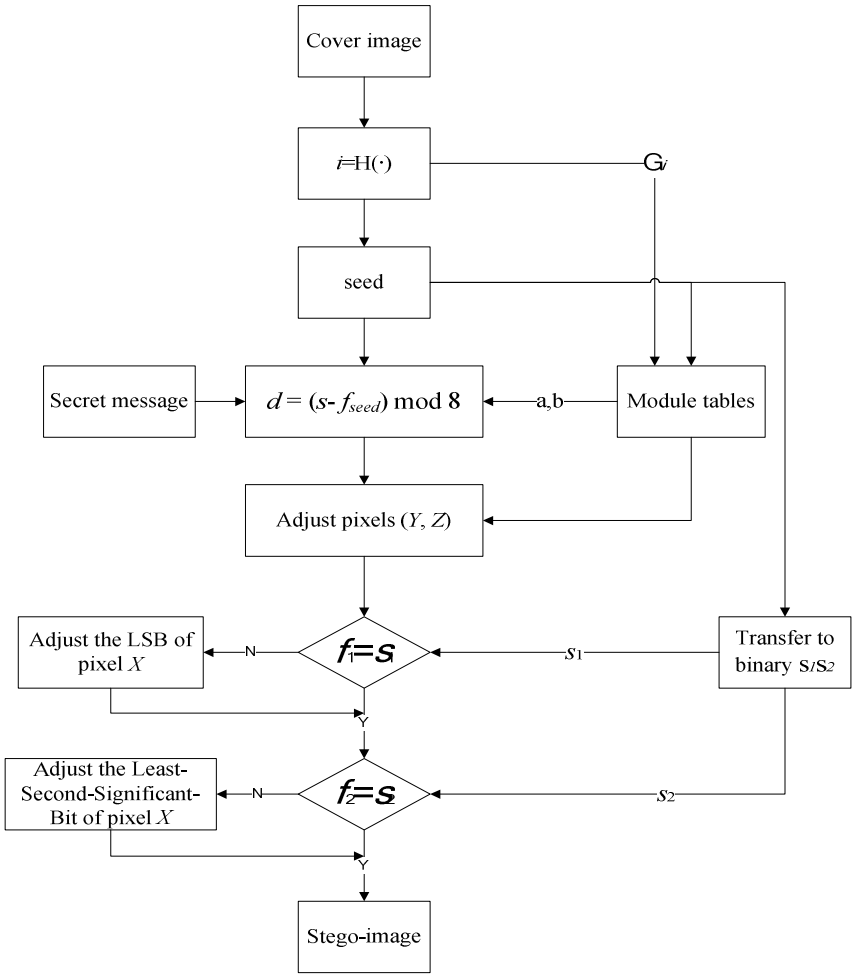


FIGURE 3: The embedding secret message procedure.

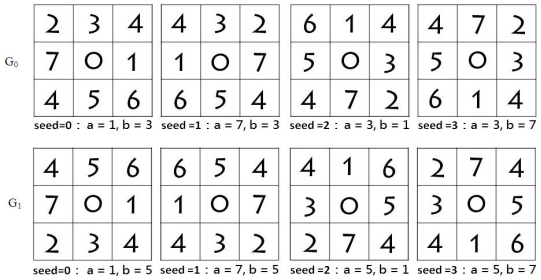


FIGURE 4: The group modular tables.

Step 4. Embed the secret message into pixels (Y, Z) by using the improved EMD method.
Step 5. Transfer the seed s_a to the binary stream s_1s_2.
Step 6. Compute v_1, which is the value of f_1, and check whether v_1 is equal to s_1 or not. If v_1 is equal to s_1, then keep the original LSB of pixel X. Otherwise, we adjust the LSB of pixel X.
Step 7. Compute v_2, which is the value of f_2, and check whether v_2 is equal to s_2 or not. If v_2 is equal to s_2, then keep the original Least-Second-Significant-Bit of pixel X. Otherwise, we adjust the Least-Second-Significant-Bit of pixel X.

3.2. The Extracting Secret Message Procedure

The flowchart of extracting secret message is shown in Fig.5. There are five steps in this procedure. Now, they are described as follows:

Step 1. Compute the value i, which is first six bits of pixel X of $H(\cdot)$, to decide group G_i.
Step 2. Extract the first bit of random seed s_1 by computing f_1.
Step 3. Extract the second bit of random seed s_2 by computing f_2.
Step 4. Transfer the binary s_1s_2 to decimal value to extract seed.
Step 5. Take pixels (Y, Z) and the weight of seed in G_i to extract the secret message by computing extract function f_{seed}.

Therefore, the receiver can recover the secret data by using the extracting procedure.

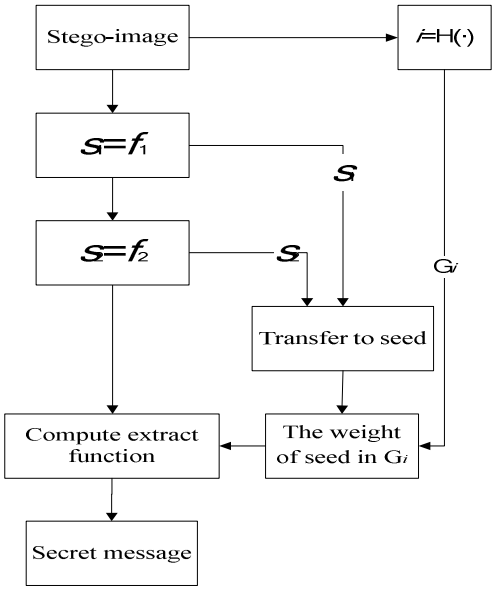


FIGURE 5: The extracting secret message procedure.

4. EXPERIMENTAL RESULT

We perform our scheme over Lena, Pepper, Baboon and Boat, which are common pictures and shown in Fig.6. These cover images are 512×512, 8bits and grayscale. The resultant stego-images are shown in Fig.7. We can't distinguish between cover-images and stego-images with human's eyes.

FIGURE 6: Cover images.

FIGURE 7: Stego-images.

Analysis of the stego-image's PSNR: From Tab.1, we can find out the stego-image's quality by using our method is lower than KWSK-scheme. In KWSK-scheme, Kuo et al. take two adjacent pixels as a group and each pixel is at most increased or decreased by 1. In our scheme, we take three adjacent pixels at a time and it is just only the second or third pixel to increased or decreased by 1 at most but the value of first pixel maybe be changed by difference 3 or 1 in each pixel group. Although the stego-image's quality in our scheme is not good as KWSW-scheme, there is an important merit is that it does not transmit the random number seeds before the sender and receiver communicates each other.

Analysis of embedding capacity: We take three pixels in a group to embed three bits at a time but Kuo et al. [6] take two pixels in a group to embed three bits. Therefore, the embedding capacity of our scheme is about 2/3 of KWSK-scheme and the experiment result shown as Table 1. Similarly, there is an important advantage in our proposed scheme which does not need the synchronous random number seed to carry although the embedding capacity in our scheme is less than KWSK-scheme.

Method	KWSK-scheme[6]		Our scheme	
	Payload (bits)	PSNR (dB)	Payload (bits)	PSNR (dB)
Lena	393,216	50.175	262,143	47.164

Pepper	393,216	50.179	262,143	47.170
Baboon	393,216	50.178	262,143	47.171
Boat	393,216	50.175	262,143	47.074

TABLE 1: The comparison between KWSK-scheme and our scheme.

5. CONCLUSION

In this paper, we propose an improved scheme by using the LSB matching method to embed seeds into the stego-image again to replace to transmit the synchronous random number seeds before the sender and the receiver commune each other, i.e., this can improve the defect of the synchronous random number seeds in KWSK-scheme. The experimental result shows that it can not only keep the acceptable image quality and security but also enhance convenience for transmission in our proposed scheme.

6. ACKNOWLEDGEMENT

This work is supported by National Science Council under NSC 98-2219-E-150-001.

7. REFERENCES

[1] FOR JOURNALS: F. Cayre, C. Fontaine, and T. Furon, “Watermarking Security: Theory and Practice,” IEEE Trans. on Signal Processing Vol.53, No.10, pp.3976-3987, Oct. 2005.

[2] FOR JOURNALS: C. C. Chang and W. C. Wu, “A Novel Data Hiding Scheme for Keeping High Stego-Image Quality,” Proceedings of the 12th International Conference on MultiMedia Modelling, Bijing, China, pp.225-232, January 2006.

[3] FOR JOURNALS: A. Ker, “Steganalysis of LSB Matching in Grayscale Images,” IEEE Signal Processing Letters, Vol.12, No.6, pp.441- 444, June 2005.

[4] FOR JOURNALS: C. F. Lee, Y. R. Wang, and C. C. Chang, “A Steganographic Method with High Embedding Capacity by Improving Exploiting Modification Direction,” IIHMSP 2007, Volume 1, Issue, pp.497 – 500, 26-28 Nov. 2007.

[5] FOR JOURNALS: J. Mielikainen, “LSB Matching Revisited,” IEEE Signal Processing Letters, Vol.13, No.5, pp.285-287, May 2006.

[6] FOR CONFERENCES: W. C. Kuo, L. C. Wuu, C. N. Shyi, and S. H. Kuo, “A Data Hiding Scheme with High Embedding Capacity Based on General Improving Exploiting Modification Direction method” HIS2009, Aug. 2009.

[7] FOR JOURNALS: R. Z. Wang, C. F. Lin, and J. C. Lin, “Image Hiding by Optimal LSB Substitution and Genetic Algorithm,” Pattern Recognition, Vol.34, No.3, pp.671-683, 2001.

[8] FOR JOURNALS: H. C. Wu, N. I. Wu, C. S. Tsai, and M. S. Hwang, “Image Steganographic Scheme Based on Pixel-Value Differencing and LSB Replacement Methods,” IEE Proceedings-Vision, Image and Signal Processing, Vol.152, No.5, pp.611-615, October 2005.

[9] FOR JOURNALS: X. Zhang and S. Wang, “Efficient Steganographic Embedding by Exploiting Modification Direction,” IEEE Comm. Letters, Vol.10, No.11, pp.1-3, Nov. 2006.

Data Steganography for Optical Color Image Cryptosystems

Cheng-Hung Chuang chchuang@asia.edu.tw
Department of Computer Science and Information Engineering
Asia University
Taichung County, 41354, Taiwan

Guo-Shiang Lin khlin@mail.dyu.edu.tw
Department of Computer Science and Information Engineering
Da-Yeh University
Changhua County, 51591, Taiwan

Abstract

In this paper, an optical color image cryptosystem with a data hiding scheme is proposed. In the proposed optical cryptosystem, a confidential color image is embedded into the host image of the same size. Then the stego-image is encrypted by using the double random phase encoding algorithm. The seeds to generate random phase data are hidden in the encrypted stego-image by a content-dependent and low distortion data embedding technique. The confidential image and secret data delivery is accomplished by hiding the image into the host image and embedding the data into the encrypted stego-image. Experimental results show that the proposed data steganographic cryptosystem provides large data hiding capacity and high reconstructed image quality.

Keywords: Data embedding, Data hiding, Image encryption, Optical security, Double random phase.

1. INTRODUCTION

With the fast development of communication and network technology, it is convenient to acquire various multimedia data through Internet. Unfortunately, the problem of illegal data access occurred frequently and popularly. Hence, it is important to protect the content and the authorized use of multimedia data against the pirates. Data encryption is a strategy to make the data unreadable, invisible or incomprehensible during transmission by scrambling the content of data [1]. In an image cryptosystem, it uses some reliable encryption algorithms or secret keys to transform or encrypt secret images into ciphered images. Only the authorized users can decrypt secret images from the ciphered images. The ciphered images are meaningless and non-recognizable for any unauthorized users who grab them without knowing the decryption algorithms or the secret keys.

Dissimilarly, data hiding or steganographic techniques refer to methods of embedding secret data into some host data in such a way that people can not discern the existence of the hidden data. For example, the well-known watermarking which usually hides copyright marks in multimedia data is a kind of data hiding technique [2]. Common methods for data hiding can be categorized into spatial and transform domain methods. The earliest method, which is simple and has high embedding capacity, embedded data into least significant bits (LSBs) of image pixels (i.e. spatial domain). Contrarily, in the transform domain, e.g., discrete cosine transform (DCT), Fourier transform, or wavelets, transformed coefficients of host signals can be manipulated to hide

messages. The image steganographic methods (or called virtual image cryptosystems) [3-6] are proposed to hide the secret images into readable but non-critical host images. They are designed to reduce the notice of illegal users.

For high speed application, image encryption methods based on optical systems have been developed. Many optical image encryption algorithms have been proposed for transmission security [7-11]. The double random phase encoding [7] is a famous and widely used algorithm which employs two random phase masks in the input plane and the Fourier plane to encrypt images into stationary white noise. In [8], an optical image cryptosystem based on the double random phase encryption and a public-key type of data embedded technique is proposed. In [9], a new image cryptosystem with an adaptive steganographic method is proposed to improve the security and visual quality. However, the input image is limited to grayscale in those cryptosystems. In [10], the encryption method using wavelength multiplexing and lensless Fresnel transform hologram is proposed for color image application. In [11], the optical color image encryption scheme is performed in the fractional Fourier transform domain.

In this paper, we propose a data steganographic scheme within an optical color image cryptosystem. A confidential color image is embedded into the phase term of the host image to become the stego-image. Then it is encrypted by using the double random phase algorithm, that is, it is multiplied by two random phase masks. The seeds to generate random phase data are embedded into the LSBs of the encrypted stego-image, in which a zero-LSB sorting technique is applied to find the hiding sequence. Simulations and experiments regarding the hiding method (in comparison with the traditional scheme [8]) are performed. Experimental results show that the proposed color image cryptosystem has a good performance in secure data embedding, large hiding capacity, and high visual quality.

In Section 2, the conventional data embedding technique and the optical cryptosystem are reviewed. Section 3 introduces the proposed steganographic optical color image cryptosystem. Section 4 shows some experimental results to demonstrate the performance of the proposed scheme and a comparison with the previous method [8]. Finally, Section 5 gives conclusion and future work.

2. REVIEW OF OPTICAL IMAGE CRYPTOSYSTEM

In optical image cryptosystems, the double random phase algorithm [7] is a very common encryption and decryption method. In the double-random-phase encoding, an image I is multiplied by a random phase mask P_1 in the input spatial plane and Fourier transformed to frequency domain. It is multiplied by another random phase mask P_2 in the Fourier plane. Then it is inverse Fourier transformed to obtain its ciphered image I_E in the output spatial plane. In the decoding process, the ciphered image I_E is Fourier transformed and then multiplied by the conjugate function of mask P_2 and inverse Fourier transformed to spatial domain. It is multiplied by the conjugate function of mask P_1 to obtain its deciphered image I_D in the output spatial plane. The equations are expressed as follows.

$$I_E = F^{-1}\left[F\left(I \times P_1\right) \times P_2\right] \tag{1}$$

$$I_D = F^{-1}\left[F\left(I_E\right) \times P_2^*\right] \times P_1^* \tag{2}$$

where $P_1 = \exp(i2\pi p_1)$ and $P_2 = \exp(i2\pi p_2)$, p_1 and p_2 are random numbers of the image size between [0, 1], F and F^{-1} define the Fourier and inverse Fourier transforms, and * denotes the conjugate operation. The optical 4f architecture, where f is the focal length of the lens, is shown in Figure 1.

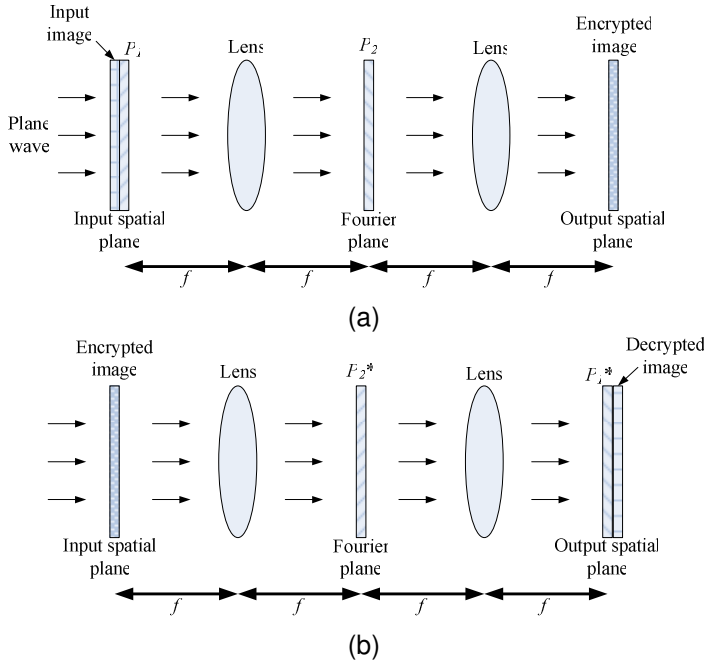


FIGURE 1: Optical 4*f* Architecture of the image cryptosystem. (a) Encryption (b) Decryption.

The data hiding scheme for the optical image cryptosystem proposed in [8] always embeds data in a fixed area of the encrypted image. Although it is a simple and fast way to complete the data embedding and extracting framework, the visual quality of the decrypted images is lower when the hidden data size is large. Therefore, in [9], a new image cryptosystem with an adaptive steganographic method is proposed to improve the visual quality of the reconstructed images. In this paper, the adaptive data hiding method is applied to the proposed optical color image cryptosystem for embedding the seeds which are used to generate double random phase. Besides, confidential or secret images can be embedded into the host images in the proposed cryptosystem.

3. THE PROPOSED METHOD

The proposed optical color image cryptosystem is based on the double random phase encryption theorem [7]. Before encoding, the confidential image I_c is embedded into the phase term of the host image I_h. Then the stego-image I_s is multiplied by a random phase mask P_1 in the input domain and transformed to Fourier plane. It is multiplied by another random phase mask P_2 and converted to the spatial domain for obtaining the encrypted stego-image I_e. The equations are defined as follows.

$$I_s = I_h \exp(i\frac{\pi}{2} I_c) \tag{3}$$

$$I_e = F^{-1}\left[F\left(I_s \times P_1\right) \times P_2\right] \tag{4}$$

where $P_1 = \exp(i2\pi p_1)$ and $P_2 = \exp(i2\pi p_2)$, p_1 and p_2 are random numbers of the image size between [0, 1], and F and F^{-1} define the Fourier and inverse Fourier transforms.

In the decoding step, the encrypted stego-image I_e is transformed to the Fourier plane, multiplied by the conjugate of mask P_2, converted to spatial domain, and multiplied by the conjugate of mask P_1 to obtain its decrypted image I_d. Ideally, the decrypted image I_d is equal to the stego-

image I_s in a lossless manner. The host image can be obtained by computing the complex modulus of the decrypted image I_d. Also the secret image can be retrieved by calculating the complex argument of the decrypted image I_d. The equations are described as follows.

$$I_d = F^{-1}\left[F(I_e)\times P_2^*\right]\times P_1^* \tag{5}$$

$$I_h = \left|I_d\right| \tag{6}$$

$$I_c = \frac{\arg(I_d)}{\pi/2} \tag{7}$$

where P_1^* and P_2^* indicate the conjugate masks of P_1 and P_2, and $\arg(\cdot)$ takes the complex argument.

For color images, they are first separated into three channels: red, green, and blue. Each channel is processed from Equations (3) to (7). However, the three channels can be coded by different random phase masks, i.e. multiplied by random phase masks P_{1R}, P_{1G}, P_{1B} and P_{2R}, P_{2G}, P_{2B}. The seeds to generate random phase data are embedded into the encrypted stego-image I_e. In the receiver side, the first thing is to decode the embedded seeds from the encrypted stego-image I_e. The decoded seeds are used to re-generate the same random numbers which are applied to produce the conjugate random phase masks P_{1R}^*, P_{1G}^*, P_{1B}^* and P_{2R}^*, P_{2G}^*, P_{2B}^*. Thus one can decode the encrypted stego-image I_e to get the decrypted image I_d using the hidden seed data extracted from the encrypted stego-image itself. Figure 2 shows the schema of the proposed optical color image cryptosystem.

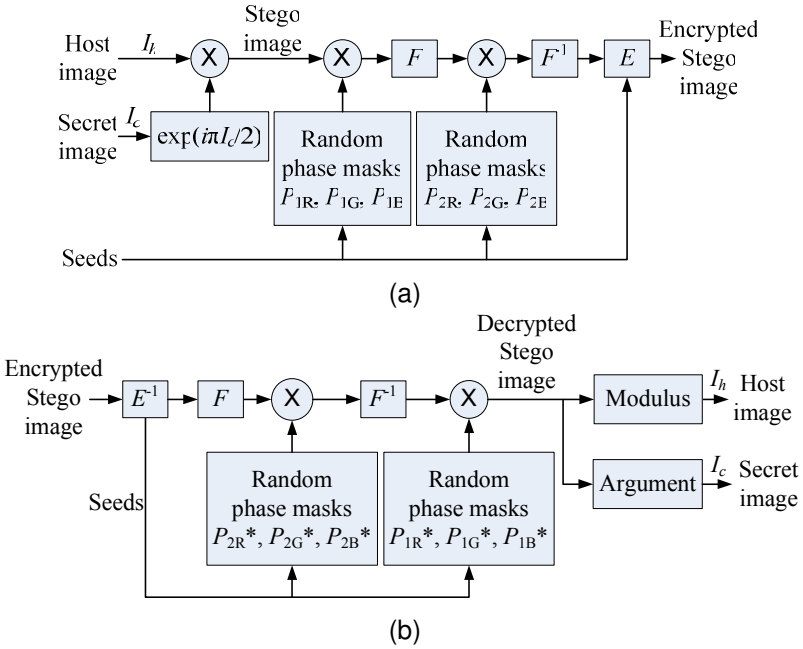


FIGURE 2: Schema of the proposed color image cryptosystem. (a) Encryption, (b) Decryption (X: multiplication, E and E^{-1}: the data embedding and extracting functions, F and F^{-1}: the Fourier and inverse Fourier transforms).

Since the signal values in the optical system are complex number format, there are real and imaginary parts that can be used to embed data. In this paper, we choose the real parts of complex numbers to be the hidden site. That is, the seeds for generating random phase data are embedded into LSBs of the quantized real parts of the encrypted stego-image bit by bit. However,

the quantization and embedding procedure will cause the loss of visual quality in the decrypted host and confidential images. The important issue is how to select the hidden positions that result in low distortion of the decrypted images. It is a simple way to hide the data within a fixed region in the encrypted stego-image [8]. Nevertheless, due to the different image content, the fixed hidden positions are not always suitable for hiding data. To improve the visual quality of the decrypted image and more safely convey the secret seed data, a low distortion, adaptive, and content-dependent data hiding technique [9] is applied to hide the secret data. In our strategy, the positions with smaller absolute values are preferable since they have smaller energy and quantization step size. To keep the embedding and decoding sequences invariant, the LSBs are set to zero and a sorting technique is employed. The detailed data hiding and extraction procedures are described as follows.

3.1 Data Hiding Procedure

Step 1: Assume that there are N bits in the secret data $B=\{b_1,b_2,...,b_N\}$. The values of real parts in the encrypted stego-image I_e are sorted in ascending order with their absolute values. The sorted set of the first N+2 numbers except the maximum and the minimum is chosen and defined as $\Lambda=\{\alpha_1,\alpha_2,\mathrm{K},\alpha_N\}$, where $|\alpha_i|\le|\alpha_{i+1}|$, α_i and $\alpha_{i+1}\in\Lambda$. Note that the maximum and minimum in the first N+2 numbers are not used to be quantized and hidden data because the quantization step size is computed from them.

Step 2: The sorted set Λ is quantized to become $\Lambda_Q=Q_L(\Lambda)=\{\alpha_{q1},\alpha_{q2},...,\alpha_{qN}\}$, where $Q_L(.)$ denotes a quantizer with L levels.

Step 3: The zero-LSB set $\Lambda_{QZ}=\{\alpha_{qz1},\alpha_{qz2},...,\alpha_{qzN}\}$ is obtained by setting all LSBs of Λ_Q to be zero. The elements in Λ_{QZ} are sorted in ascending order with their absolute values to get $\Lambda_{QZS}=\{\alpha_{qzs_1},\alpha_{qzs_2},...,\alpha_{qzs_N}\}$, where $|\alpha_{qzs_i}|\le|\alpha_{qzs_{i+1}}|$, α_{qzs_i} and $\alpha_{qzs_{i+1}}\in\Lambda_{QZS}$.

Step 4: The sequence $S=\{s_1,s_2,...,s_N\}$, where $s_i\in\{1,2,...,N\}$ and $i=1,2,...,N$, generated by the set Λ_{QZS}, is used to be the data hiding index. That is, the secret data is successively embedded into the LSBs of the set Λ_Q according to the sequence S, i.e. $\Lambda_{QS}=\{\alpha_{qs_1},\alpha_{qs_2},...,\alpha_{qs_N}\}$, where $\alpha_{qs_i}\in\Lambda_Q$.

Step 5: The hiding rule is defined as

$$\Lambda_{QS}^E=\Lambda_{QS}+\mathrm{sgn}(B-\mathrm{mod}(\Lambda_{QS},2)) \tag{8}$$

where $\mathrm{sgn}(\cdot)\in\{-1,0,1\}$ is the signum function and $B=\{b_1,b_2,...,b_N\}$ is the secret data. The set with hidden data is $\Lambda_{QS}^E=\{\alpha_{qs_1}^e,\alpha_{qs_2}^e,...,\alpha_{qs_N}^e\}$.

Step 6: Finally, the set Λ_{QS}^E is de-quantized to obtain $\Lambda_S^E=Q_L^{-1}(\Lambda_{QS}^E)=\{\alpha_{s_1}^e,\alpha_{s_2}^e,...,\alpha_{s_N}^e\}$, where $Q_L^{-1}(\cdot)$ is the de-quantizer with L levels.

3.2 Data Extraction Procedure

Step 1: This step is the same as the first step in data hiding procedure to find the sorted set. The set is defined as $\Lambda^E=\{\alpha_1^e,\alpha_2^e,...,\alpha_N^e\}$, where $|\alpha_i^e|\le|\alpha_{i+1}^e|$, α_i^e and $\alpha_{i+1}^e\in\Lambda^E$. The sequence in the sorted set Λ^E is different from that in the sorted set Λ.

Step 2: The sorted set Λ^E is quantized with L levels to be $\Lambda_Q^E=Q_L(\Lambda^E)=\{\alpha_{q1}^e,\alpha_{q2}^e,...,\alpha_{qN}^e\}$.

Step 3: All LSBs of Λ_Q^E are set to zero to obtain the zero-LSB set $\Lambda_{QZ}^E=\{\alpha_{qz1}^e,\alpha_{qz2}^e,...,\alpha_{qzN}^e\}$. The elements in Λ_{QZ}^E are sorted in ascending order with their absolute values to get $\Lambda_{QZS}^E=\{\alpha_{qzs_1}^e,\alpha_{qzs_2}^e,...,\alpha_{qzs_N}^e\}$, where $|\alpha_{qzs_i}^e|\le|\alpha_{qzs_{i+1}}^e|$, $\alpha_{qzs_i}^e,\alpha_{qzs_{i+1}}^e\in\Lambda_{QZS}^E$.

Step 4: Now, the set Λ_{QZS}^{E} is equal to the set Λ_{QZS} with the same sequence $S=\{s_1, s_2, ..., s_N\}$. The hidden data is extracted from the LSBs of the set $\Lambda_{QS}^{E}=\{\alpha_{qs_1}^{e}, \alpha_{qs_2}^{e}, ..., \alpha_{qs_N}^{e}\}$, i.e.

$$\begin{cases} b_i = 0, \text{if } \bmod(\alpha_{qs_i}^{e}, 2) = 0 \\ b_i = 1, \text{if } \bmod(\alpha_{qs_i}^{e}, 2) = 1 \end{cases}, i = 1,2,...,N \tag{9}$$

4. EXPERIMENTAL RESULTS

In the experiment, one hundred 24-bit 512×512-pixel various color images (collected from [12-14]) are examined as host images and the peak signal-to-noise ratio (PSNR) is applied to evaluate the visual quality of the decrypted images. The equation is defined as follows.

$$MSE = \frac{MSE_R + MSE_G + MSE_B}{3} \tag{10}$$

$$PSNR = 10 \times \log_{10} \frac{255^2}{MSE} \tag{11}$$

where MSE_R, MSE_G, and MSE_B are mean square errors in three channels, respectively.

The traditional data hiding scheme [8], where the secret data are embedded in the central square area of the encrypted stego-image, is performed for comparison. For a fair evaluation, the size of hidden data is fixed and set to 480,000 bits. Table 1 shows the average PSNR values of the 100 decrypted host images and the retrieved secret images with different quantization levels, i.e. L = 8, 16, 32, 64, 128, and 256. It is clear that the PSNR values increase about 14 dB both in the decrypted host and secret images of the proposed method. Figure 3 plots detailed PSNR values of 100 decrypted host and secret images with L = 8, where the blue and red curves are the results of the proposed and the traditional methods, respectively. Figure 4 shows some original, decrypted host, and decrypted secret images sampled from the 100 test cases, where the quantization level is 8. The first row of Figure 4 shows the original images, where Figure 4(a)-(c) are host images and Figure 4(d) is the secret image. The second row of Figure 4 is the results of the proposed method, where the PSNR values are 19.14, 20.27, 20.22, and 32.10 dB, respectively. The last row of Figure 4 is the results of the traditional data hiding method [8], where the PSNR values are 4.81, 5.52, 5.79, and 16.96 dB, respectively.

L	Average PSNR (dB)			
	Proposed method		Ref. [8]	
	Host images	Secret images	Host images	Secret images
8	20.88	32.14	6.57	16.89
16	26.88	38.19	12.56	23.70
32	32.89	44.22	18.57	29.85
64	38.91	50.24	24.57	35.89
128	44.93	56.26	30.59	41.91
256	50.95	62.28	36.64	47.97

TABLE 1: Comparisons between the proposed method and the traditional scheme [8] of the average PSNR values of the 100 decrypted host images and the retrieved secret images. (hidden data 480,000 bits)

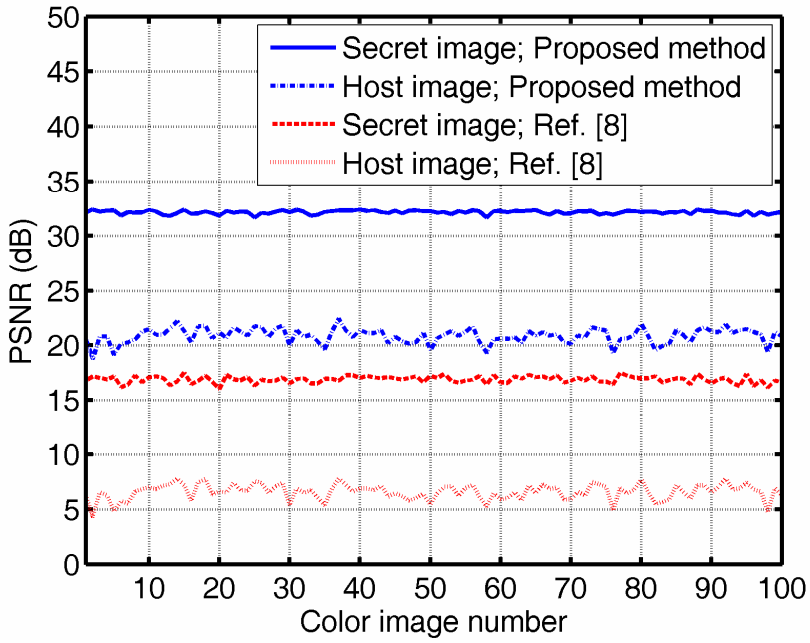


FIGURE 3: PSNR curves of the 100 decrypted host and secret images. The blue and red curves are the results of the proposed and the traditional methods [8], respectively. (L = 8, hidden data 480,000 bits)

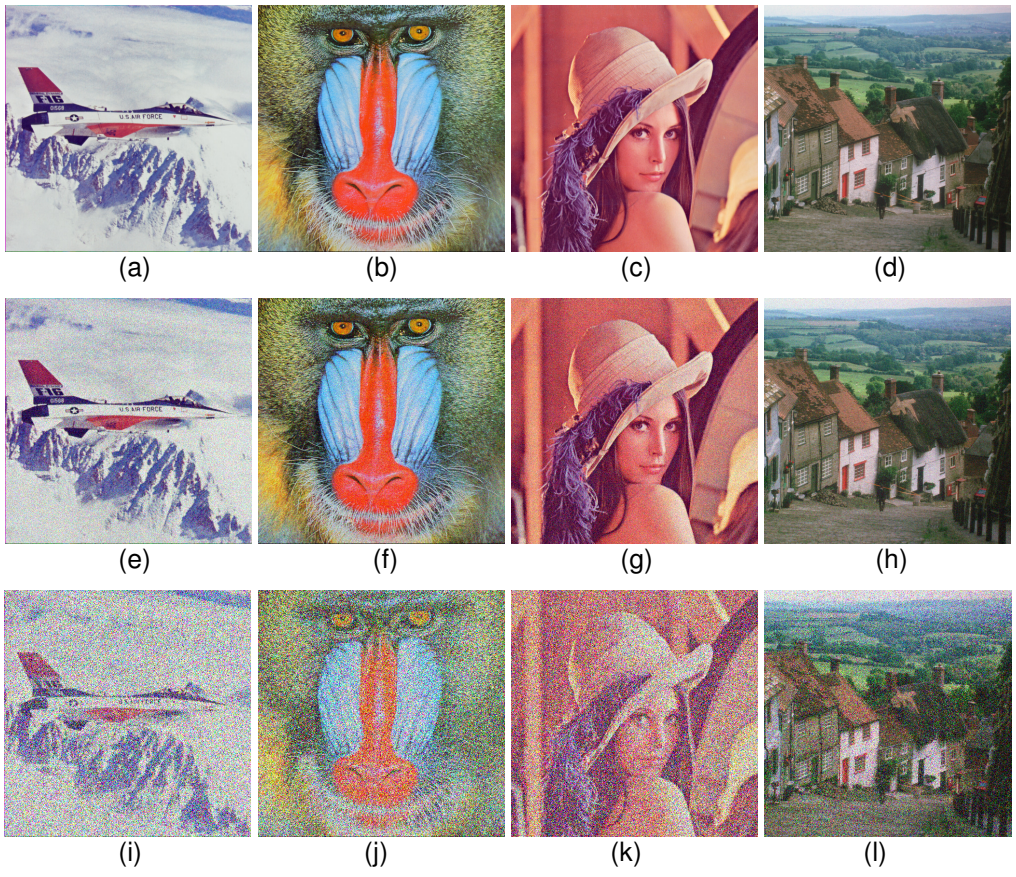


FIGURE 4: (a)(b)(c) The original host images and (d) the secret image, (e)-(h) the decrypted host and secret images by the proposed method, (i)-(l) the decrypted host and secret images by the traditional data hiding scheme. (L = 8, hidden data 480,000 bits)

To evaluate the data hiding capacity versus the visual quality of the decrypted host and secret images, the encrypted stego-images are embedded with different data size ranged from 108 to about 740,000 bits. The quantization level is set to 8. The average PSNR values of the 100 decrypted host and secret images are computed for evaluating the visual quality. Figure 5 shows the curves of the data hiding capacity versus the average PSNR values, where the blue and red curves are the results using the proposed and traditional methods, respectively. The PSNR values in the results of the proposed method are larger than those in the results of the traditional scheme when the sizes of hidden data are the same. It is obvious that the proposed method has a better performance than the traditional one.

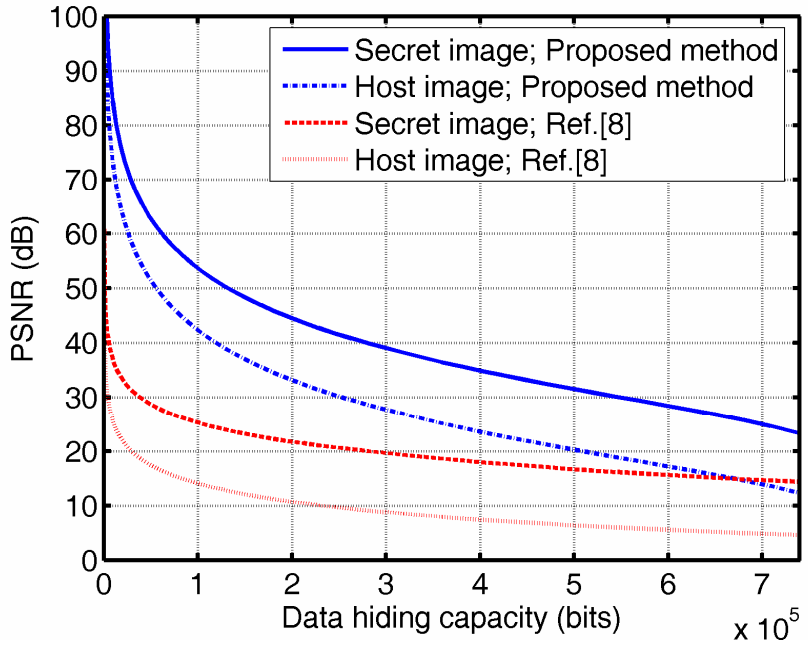


FIGURE 5: Curves of the data hiding capacity versus the visual quality of the decrypted host and secret images. The blue and red curves are the results of the proposed and the traditional methods [8], respectively. (L = 8)

In real applications, pirates may attempt to maintain the reconstructed images recognizable by modifying the encrypted stego-images. However, when the encrypted stego-images are attacked, it is expected that the hidden data will be altered. If the hidden data can not be correctly extracted, the stego-images can not be properly decrypted. In this part of experiment, it is assumed that the hidden data are completely cracked and the amplitude parts of the encrypted stego-images are suffered from three common attacks, i.e. noising, smoothing, and JPEG compression. The Gaussian noise (with zero mean and 0.01 variance) and the 3×3 averaging filter are exploited to disturb the encrypted stego-images for the noising and smoothing attacks. In the JPEG compression, 56.25% (36/64) of the DCT coefficients in the high frequency part in each 8×8 block were discarded (set to zero). The size of hidden data is set to 480,000 bits. The quantization level is also set to 8. The average PSNR values of the 100 decrypted host and secret images are calculated and listed in Table 2, where the encrypted stego-images are suffered from attacks. Without attacks, the average PSNR values in the decrypted host and secret images (shown in Table 1) are 20.88 and 32.14 dB by the proposed method and 6.57 and 16.89 dB by the

traditional scheme, respectively. With the three attacks, the PSNR values of all the decrypted host and secret images are reduced. The visual quality of the decrypted host and secret images in the smoothing attack is almost the worst. However, the results of the proposed method are still better than those of the traditional scheme.

Three common attacks	Average PSNR (dB)			
	Proposed method		Ref. [8]	
	Host images	Secret images	Host images	Secret images
Noising	9.08	19.85	4.42	14.02
Smoothing	7.25	16.31	5.34	12.31
JPEG compression	9.06	19.12	5.69	13.71

TABLE 2: Comparisons between the proposed method and the traditional scheme [8] of the average PSNR values of the 100 decrypted host images and the retrieved secret images when the encrypted stego-images are attacked. (L = 8, hidden data 480,000 bits)

5. CONCLUSION & FUTURE WORK

In this paper, the optical color image cryptosystem with data steganography is proposed. The double random phase encoding algorithm and the adaptive data hiding technique are applied in the proposed color image cryptosystem. The confidential image is hidden in the phase term of the host image. Then the stego-image is encrypted by the double random phase encoding algorithm. The seeds to generate random phase data are embedded into the encrypted stego-image by the proposed data hiding method. In comparison with the traditional hiding scheme, a larger data embedding capacity and higher visual quality of the decrypted host and confidential images are achieved.

For the advanced security, the confidential image and the secret data can be disordered by the scrambling technique before they are hidden. The secret or session keys for scrambling can also be embedded in the encrypted stego-image. Moreover, they can be encrypted by the asymmetric cryptographic algorithm, e.g. the RSA (Rivest-Shamir-Adleman) method. It is verified that the proposed cryptosystem provides a confidential image steganographic method and secret data hiding scheme to improve transmission security of the secret information.

6. ACKNOWLEDGEMENT

This research was supported by the National Science Council, Taiwan, under the grant of NSC97-2221-E-468-006.

7. REFERENCES

1. M. Yang, N. Bourbakis, and Li Shujun, "*Data-image-video encryption*," IEEE Potentials, vol. 23, no. 3, pp. 28-34, 2004.
2. Y. Govindarajan and S. Dakshinamurthi, "*Quality - security uncompromised and plausible watermarking for patent infringement*," International Journal of Image Processing, vol. 1, no. 2, 2007.
3. T.-S. Chen, C.-C. Chang, and M.-S. Hwang, "*A virtual image cryptosystem based on vector quantization*," IEEE Trans. Image Processing, vol. 7, no. 10, pp. 1485-1488, 1998.
4. Y.-C. Hu, "*High-capacity image hiding scheme based on vector quantization*," Pattern Recognition, vol. 39, no. 9, pp. 1715-1724, 2006.

5. C.-C. Chang, C.-Y. Lin, and Y.-Z. Wang, "*New image steganographic methods using run-length approach*," Information Sciences, vol. 176, no. 22, pp. 3393-3408, 2006.
6. W.-Y. Chen, "*Color image steganography scheme using set partitioning in hierarchical trees coding, digital Fourier transform and adaptive phase modulation*," Applied Mathematics and Computation, vol. 185, no. 1, pp. 432-448, 2007.
7. P. Refregier and B. Javidi, "*Optical image encryption based on input plane and Fourier plane random encoding*," Optics Letters, vol. 20, pp. 767-769, 1995.
8. G.-S. Lin, H. T. Chang, W.-N. Lie, and C.-H. Chuang, "*A public-key-based optical image cryptosystem based on data embedding techniques*," Optical Engineering, vol. 42, no. 8, pp. 2331-2339, 2003.
9. C.-H. Chuang and G.-S. Lin, "*An optical image cryptosystem based on adaptive steganography*," Optical Engineering, vol. 47, 047002 (9 pages), April 2008.
10. L. Chen and D. Zhao, "*Optical color image encryption by wavelength multiplexing and lensless Fresnel transform holograms*," Optics Express, vol. 14, pp. 8552-8560, 2006.
11. M. Joshi, Chandrashakher, and K. Singh, "*Color image encryption and decryption using fractional Fourier transform*," Optics Communications, vol. 279, pp. 35-42, 2007.
12. Computer Vision Group (CVG), Department of Computer Science and Artificial Intelligence, University of Granada. Retrieved from http://decsai.ugr.es/cvg/, August 2008.
13. Kodak Lossless True Color Image Suite. Retrieved from http://r0k.us/graphics/kodak/, August 2008.
14. Programming, Image Processing, and Video Codecs Resourses. Retrieved from http://www.hlevkin.com/, August 2008.

Preserving Global and Local Features for Robust Face Recognition under Various Noisy Environments

Ruba Soundar Kathavarayan rubasoundar@yahoo.com
Department of Computer Science and Engineering
PSR Engineering College
Sivakasi, 626140, India

Murugesan Karuppasamy k_murugesan2000@yahoo.com
Principal
Maha Barathi Engineering College
Chinna Salem, 606201, India

Abstract

Much research on face recognition considering the variations in visual stimulus due to illumination conditions, viewing directions or poses, and facial expressions has been done earlier. However, in reality the noises that may embed into an image document will affect the performance of face recognition algorithms. Though different filtering algorithms are available for noise reduction, applying a filtering algorithm that is sensitive to one type of noise to an image which has been degraded by another type of noise lead to unfavorable results. These conditions stress the importance of designing a robust face recognition algorithm that retains recognition rates even under noisy conditions. In this work, numerous experiments have been conducted to analyze the robustness of our proposed Combined Global and Local Preserving Features (CGLPF) algorithm along with other existing conventional algorithms under different types of noises such as Gaussian noise, speckle noise, salt and pepper noise and quantization noise.

Keywords: Biometric Technology, Face Recognition, Noise Reduction, Global Feature and Local Feature

1. INTRODUCTION

Biometric technologies are becoming the foundation of an extensive array of highly secure identification and personal verification solutions. As the level of security breaches and transaction fraud increases, the need for highly secure identification and personal verification technologies is becoming apparent. Biometric authentication has been widely regarded as the most foolproof - or at least the hardest to forge or spoof. The increasing use of biometric technologies in high-security applications and beyond has stressed the requirement for highly dependable face recognition systems. The biometric technology of a face recognition system is used to verify an identity of a person by matching a given face against a database of known faces. It has become a viable and an important alternative to traditional identification and authentication methods such as the use of keys, ID cards and passwords.

Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [1-6]. Even though human can detect and identify faces in a scene with little or no effort, building an automated system that accomplishes such objectives is very challenging. The challenges are even more profound when one considers the

large variations in the visual stimulus due to illumination conditions, viewing directions or poses, facial expressions, aging, and disguises such as facial hair, glasses, or cosmetics [7, 8]. Face recognition technology provides the cutting edge technologies that can be applied to a wide variety of application areas including access control for PCs, airport surveillance, private surveillance, criminal identification and as an added security for ATM transaction. In addition, face recognition system is also currently being used in growing numbers of applications as an initial step towards the next-generation smart environment where computers are designed to interact more like humans.

In recent years, considerable progress has been made in the area of face recognition with the development of many techniques. Whilst these techniques perform extremely well under constrained conditions, the problem of face recognition in uncontrolled noisy environment remains unsolved. During the transmission of images over the network, some random usually unwanted variation in brightness or colour information may be added as noise. Image noise can originate in film grain, or in electronic noise in the input device such as scanner [9], digital camera, sensor and circuitry, or in the unavoidable shot noise of an ideal photon detector. Slow shutter speed and in low light having high exposure of the camera lens are also some of the reasons that noise gets added to the image. Noise causes a wrong conclusion in the identification of images in authentication and also in pattern recognition process. The noise should be removed prior to performing image analysis processes. The identification of the nature of the noise [10] is an important part in determining the type of filtering that is needed for rectifying the noisy image. Noise in imaging systems is usually either additive or multiplicative [11]. In practice these basic types can be further classified into various forms [12] such as amplifier noise or Gaussian noise, Impulsive noise or salt and pepper noise, quantization noise, shot noise, film grain noise and non-isotropic noise. However, in our experiments, we have considered the common noises such as, Gaussian additive noise, speckle multiplicative noise, quantization and salt and pepper impulsive noise.

The previous study [13] proposed several noise removal filtering algorithms. Most of them assume certain statistical parameters and know the noise type a priori, which is not true in practical cases. Applying a filtering algorithm that is sensitive to additive noise to an image that has been degraded by a multiplicative noise doesn't give an optimal solution. Also the difficulty in removing salt/pepper noise from binary image is due to the fact that image data as well as the noise share the same small set of values (either 0 or 1) which complicates the process of detecting and removing the noise. This is different from grey images where salt/pepper noise could be distinguished as pixels having big difference in grey level values compared with their neighbourhood. Many algorithms have been developed to remove salt/pepper noise in document images with different performance in removing noise and retaining fine details of the image. Most methods can easily remove isolated pixels while leaving some noise attached to graphical elements. Other methods may remove attached noise with less ability in retaining thin graphical elements. These conditions in turn stress the importance of the design of robust face recognition algorithms that retain recognition rates even under noisy environments.

In general all the face recognition algorithms uses any one or the combinations of the features namely shape, texture, colour, or intensity to represent the facial image structure. It has been seen from previous works that the appearance based representations that uses the intensity or pixel values produces the better result compared with other techniques. But the intensity features are very vulnerable to image noises that may add with the original image during transmission or during the capturing processes itself. In reality, most of the face recognition algorithms that uses appearance based representations are considered only for the noiseless environments and are not dealing with different type of noises occurred in the image.

From an appearance representation standpoint, Principal Component Analysis (PCA) [14], Multidimensional Scaling (MDS), Linear Discriminant Analysis (LDA) [3], and Locality Preserving Projections (LPP) [4] based techniques are more relevant. In those appearance based face recognition, the global features preserving techniques namely PCA, MDS, and LDA effectively preserves the Euclidean structure of face space or the global features. On the other hand, the local feature preservation technique namely Locality Preserving Projections (LPP) preserves local information and obtains a face subspace that best detects the essential face manifold structure. Global features preserving techniques suffer when the noises affect the global features like the

structure of the facial images, while local features preserving techniques suffer when the image noises affect the local intensity pixels. Hence in our proposed work, for the first time up to our knowledge, we employ the combination of global feature extraction technique LDA and local feature extraction technique LPP, to achieve a high quality feature set called Combined Global and Local Preserving Features (CGLPF) that captures the discriminate features among the samples considering the different classes in the subjects [15]. This increases the robustness of face recognition against noises affecting global features and / or local features. In this work, experiments have been conducted to reveal the robustness of our proposed Combined Global and Local Preserving Features algorithm under different types of noises and the results are compared with that of other traditionally employed algorithms.
The rest of the paper is organized as follows: Section 2 describes various types of common noises that affect the biometric identification of facial images. The basic concepts of proposed CGLPF algorithm is given in section 3. In section 4, the experimental results have been discussed with respect to percentage of correct recognition considering ORL facial image database under various noisy environments for CGLPF in comparison with other traditional PCA, LDA and LPP algorithms. The paper is concluded with some closing remarks in section 5.

2. DIFFERENT CATEGORIES OF NOISES AFFECTING IMAGES

Image Noise [12] is usually an unwanted random variation observed in the brightness or the color information of an image. Image noise can be originated due to an electronic noise in the sensors of the digital cameras or scanners circuitry. Slow shutter speed and in low light having high exposure of the camera lens are some of the reasons that noise gets added to the image. There are different types of noises such as additive noise, multiplicative noise, quantization noise and impulse noise. The identification of the nature of the noise [10] is an important part in determining the type of filtering that is needed for rectifying the noisy image. Most of the filtering algorithms for noise rectification assume certain statistical parameters and the type of noise, which is not true in the practical cases. Applying a filtering algorithm that is sensitive to additive noise to an image degraded by a multiplicative noise doesn't yield an optimal solution. The different types of noises and their properties are discussed here.

Additive Noise

This kind of noise gives a linear impairment to the image. It involves a linear addition of white noise with constant spectral density to the original image. The noise added is constant i.e., additive noises are independent at each pixel and independent of the signal intensity. When noise is additive, an observed image can be described as

$$I_v(x,y) = I(x,y) + V(x,y) \tag{1}$$

where I_v is the observed image with noise, I is the true signal (image), and V is the noise component. Many additive noise models exist and the following are some common additive noise models with their Probability Density Function (PDF) [11].
Gaussian noise provides a good model of noise in many imaging systems. Generally, we consider the normal distribution with arbitrary center μ, and variance σ^2. The PDF for such distribution is given by the formula

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

where the parameter μ is called the mean, and it determines the location of the peak of the density function, parameter σ is called standard deviation, and σ^2 is variance of the distribution.
Laplacian noise are also called as biexponential noise and its PDF is represented by,

$$f(x)=\frac{1}{\sqrt{2\sigma}}\ e^{-\frac{\sqrt{2}|x|}{\sigma}} \tag{3}$$

Uniform noise is not often encountered in real-world imaging systems, but provides a useful comparison with Gaussian noise. The PDF of uniform distribution is given by

$$f(x)=\begin{cases}\dfrac{1}{\sigma^2\sqrt{3}} & for\ |x| \le \sigma\sqrt{3} \\ 0 & else\end{cases} \tag{4}$$

Multiplicative Noise

When noise introduces is multiplicative effect, an observed image can be described as

$$I_{xv}(x,y)=I(x,y)H(x,y) \tag{5}$$

where I_{xv} is the observed image with noise, I is the true signal (image), and H is the multiplicative noise component.

When this noise is applied to a brighter area of an image, it presents a magnified view and a higher random variation in pixel intensity is observed. On the other hand, when this noise is applied to a darker region in the image, the random variation observed is not that much as compared to that observed in the brighter areas. Thus, this type of noise is signal dependent and distorts the image in large magnitude and is often called as the speckle noise [16].

Normally data-dependent noises arise when monochromatic radiation is scattered from a surface whose roughness is of the order of a wavelength, causing wave interference which results in image speckle. It is possible to analyze this noise with multiplicative or non-linear models. These models are mathematically more complicated and hence if possible, the speckle noise is mostly assumed to be data independent. The following is the PDF of the multiplicative (speckle) noise with Rayleigh distributions [17]:

$$f(x)=\begin{cases}\dfrac{2}{b}(x-a)e^{-\frac{(x-a)^2}{b}} & for\ x \ge a \\ 0 & for\ x<a\end{cases} \tag{6}$$

where the parameters are such that $a > 0$, b is a positive integer. The mean and variance of this PDF are given by equation 7 and 8.

$$\mu = a+\sqrt{\frac{\pi b}{4}} \tag{7}$$

$$\sigma^2=\frac{b\,(4-\pi)}{4} \tag{8}$$

Quantization Noise

Quantization noise [18] is the quantization error introduced by the process of quantization in the analog-to-digital conversion (ADC) in telecommunication systems and signal processing applications. It is a rounding error between the analog input voltage to the ADC and the output digitized value. The noise is non-linear and signal-dependent in nature. It can be modeled in several different ways.

In image processing, the noise caused by quantizing the pixels of a sensed image to a number of discrete levels is known as quantization noise. It has an approximately uniform distribution, and can be signal dependent, though it will be signal independent if other noise sources are big enough to cause dithering, or if dithering is explicitly applied. Quantization of number of discrete levels is important for displaying images on devices that support a limited number of colors and for efficiently compressing certain kinds of images. The human eye is fairly good at seeing small differences in brightness over a relatively large area, but not so good at distinguishing the exact strength of a high frequency brightness variation. This fact allows one to get away with a greatly

reduced amount of information in the high frequency components. This is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer. As a result of this, it is typically the case that many of the higher frequency components are rounded to zero, and many of the rest become small positive or negative numbers. Losses occur due to this process is termed as quantization noise.

Impulsive Noise

Impulsive noise is sometimes as called salt-and-pepper noise or spike noise [17]. An image containing salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions. This type of noise can be caused by dead pixels, analog-to-digital converter errors, and bit errors in transmission. It represents itself as randomly occurring white and black pixels.
Bipolar impulse noise follows the following distribution

$$f(x) = \begin{cases} f_a & for\ x = a \\ f_b & for\ x = b \\ 0 & otherwise \end{cases} \tag{9}$$

In this equation, if f_a or f_b is zero, we have unipolar impulse noise. If both are nonzero and almost equal, it is called salt-and-pepper noise. Impulsive noises can be positive and / or negative. It is often very large and can go out of the range of the image. It appears as black and white dots, or saturated peaks.

3. FORMATION OF COMBINED GLOBAL AND LOCAL PRESERVING FEATURES (CGLPF)

Earlier works based on PCA [14] or LDA [19] suffer from not preserving the local manifold of the face structure whereas the research works on LPP [4] lacks to preserve global features of face images. Some papers [1, 20] uses the combination of both PCA and LPP, captures only the most expressive features whereas our proposed work uses the combination LDA and the distance preserving spectral method LPP, that captures the most discriminative features which plays a major role in face recognition. Also those works that uses PCA captures the variation in the samples without considering the variance among the subjects. Hence in our proposed work, for the first time up to our knowledge, we employ the combination of global feature extraction technique LDA and local feature extraction technique LPP to achieve a high quality feature set called Combined Global and Local Preserving Features (CGLPF) that captures the discriminate features among the samples considering the different classes in the subjects which produces the considerable improved results in facial image representation and recognition.
The proposed combined approach that combines global feature preservation technique LDA and local feature preservation technique LPP to form the high quality feature set CGLPF is described in this section. Actually, the CGLPF method is to project face data to an LDA space for preserving the global information and then projecting to Locality Preserving Projection (LPP) space by using the distance preserving spectral methods, to add the local neighbourhood manifold information which may not be interested by LDA.

Preserving the Global Features

The mathematical operations involved in LDA, the global feature preservation technique is analyzed here. The fundamental operations are:

1. The data sets and the test sets are formulated from the patterns which are to be classified in the original space.
2. The mean of each data set μ_i and the mean of entire data set μ are computed.

$$\mu = \sum_i p_i \mu_i \tag{10}$$

where p_i is priori probabilities of the classes.

3. Within-class scatter S_w and the between-class scatter S_b are computed using:

$$S_w = \sum_j p_j * (cov_j) \tag{11}$$

$$S_b = \sum_j (x_j - \mu)(x_j - \mu) \tag{12}$$

where cov_j the expected covariance of each class is computed as:

$$cov_j = \prod_i (x_j - \mu_i) \tag{13}$$

Note that S_b can be thought of as the covariance of data set whose members are the mean vectors of each class. The optimizing criterion in LDA is calculated as the ratio of between-class scatter to the within-class scatter. The solution obtained by maximizing this criterion defines the axes of the transformed space.

The LDA can be a class dependent or class independent type. The class dependent LDA requires L-class L separate optimizing criterion for each class denoted by C_1, C_2, …, C_L and that are computed using:

$$C_j = (cov_j)^{-1} S_b \tag{14}$$

4. The transformation space for LDA, W_{LDA} is found as the Eigen vector matrix of the different criteria defined in the equation 14.

Adding Local Features

The local features are added to the preserved global features in order to increase the robustness of our technique against various noises. Actually the local features preserving technique seeks to preserve the intrinsic geometry of the data and local structure. The following are the steps to be carried out to obtain the Laplacian transformation matrix W_{LPP}, which we use to preserve the local features.

1. **Constructing the nearest-neighbor graph:** Let G denote a graph with k nodes. The i^{th} node corresponds to the face image x_i. We put an edge between nodes i and j if x_i and x_j are "close," i.e., x_j is among k nearest neighbors of x_i, or x_i is among k nearest neighbors of x_j. The constructed nearest neighbor graph is an approximation of the local manifold structure, which will be used by the distance preserving spectral method to add the local manifold structure information to the feature set.
2. **Choosing the weights:** The weight matrix S of graph G models the face manifold structure by preserving local structure. If node i and j are connected, put

$$S_{ij} = e^{\frac{-\| x_i - x_j \|^2}{t}} \tag{15}$$

where t is a suitable constant. Otherwise, put $S_{ij} = 0$.

3. **Eigen map:** The transformation matrix W_{LPP} that minimizes the objective function is given by the minimum Eigen value solution to the generalized Eigen value problem. The detailed study about LPP and Laplace Beltrami operator is found in [1, 21]. The Eigen vectors and Eigen values for the generalized eigenvector problem are computed using equation 16.

$$XLX^T W_{LPP} = \lambda XDX^T W_{LPP} \tag{16}$$

where D is a diagonal matrix whose entries are column or row sums of S, $D_{ii} = \Sigma_j S_{ji}$, $L = D - S$ is the Laplacian matrix. The i^{th} row of matrix X is x_i. Let $W_{LPP} = w_0, w_1, ..., w_{k-1}$ be the solutions of the above equation, ordered according to their Eigen values, $0 \le \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{k-1}$. These Eigen values are equal to or greater than zero because the matrices XLX^T and XDX^T are both symmetric and positive semi-definite. Note that the two matrices XLX^T

and XDX^T are both symmetric and positive semi-definite since the Laplacian matrix L and the diagonal matrix D are both symmetric and positive semi-definite.

By considering the transformation space W_{LDA} and W_{LPP}, the embedding is done as follows:

$$x \rightarrow y = W^T x,$$
$$W = W_{LDA} W_{LPP},$$
$$W_{LPP} = [w_0, w_1, ..., w_{k-1}] \quad (17)$$

where y is a k-dimensional vector, W_{LDA}, W_{LPP} and W are the transformation matrices of LDA, LPP and CGLPF algorithms respectively.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Real world signals usually contain departures from the ideal signal that would be produced by the model of signal production process. Such departures are referred to as noise. Noise arises as a result of unmodeled or unmodelable processes going on in the production and capture of the real signal. It is not part of the ideal signal and may be caused by a wide range of sources, e.g. variations in the detector sensitivity, environmental variations, the discrete nature of radiation, transmission or quantization errors, etc. These noises are the tough challengers in affecting the performance of many biometric techniques. In this work, we introduce different types of noises at varied specifications and analyze the robustness performance of the CGLPF feature set comparing with the conventional existing techniques such as PCA, LDA and LPP.

For our experiments, the facial images from the facial image database ORL are used. The ORL database contains a total of 400 images containing 40 subjects each with 10 images that differ in poses, expressions and lighting conditions. Figure 1 shows the sample images used in our experiments collected from ORL face database. In our experiments, we have used common types of noises namely, Gaussian additive noise, speckle multiplicative noise, quantization noise, and salt and pepper impulsive noise that affect the biometric image processing applications. In order to show the robustness of our CGLPF based face recognition method, these noises are introduced in the ORL database face images before applying the CGLPF algorithm. The ORL face database images with noises are shown in figure 2.

Figure 1: The sample set of images collected from ORL database

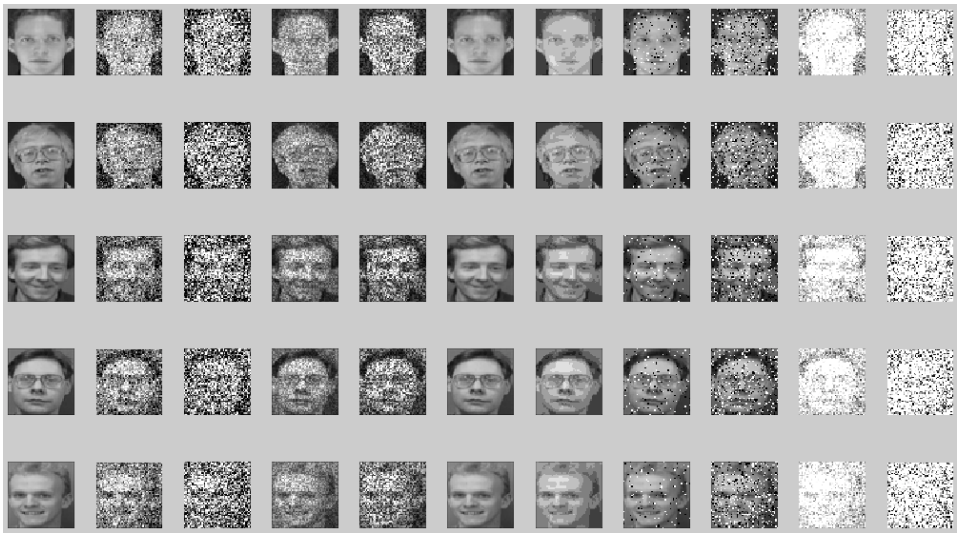


Figure 2: The sample set of noisy images

The first column of figure 2 shows the original image set without noise. The second and third columns show the images affected by Gaussian noise with mean 0.05 variance 0.05, and mean 0.05 variance 0.2 respectively. Similarly fourth and fifth columns show the image with speckle noise with variance 0.05 and 0.2 respectively. Quantization noise image with 1 bit and 6 bit quantization error are shown in column 6 and 7. Column 8 and 9 show the image with salt and pepper noise with variance of 0.05 and 0.2 respectively. Column 10 and 11 indicate the images affected by Gaussian noise with mean 0.5, variance 0.05 and mean 0.75, variance 0.5 respectively. It is evident from the figure that when the noise level increases, the face images get affected more and sometimes is not visible. Hence in our experiments, we have considered mean and variance varying from 0 to 0.2 only.

Any biometric authentication tool has some set of images called as prototype images also known as authenticated images, and another set of images which are given as input for the purpose of probing. The tool has to decide whether the input probe image is accepted or not by verifying the similarities of probe image and any matching prototype image without considering noises present, variations in poses, lighting conditions or illuminations. To start with, the probing image set is formed by applying the Gaussian noise with mean and variance equal to 0.05 on all the 400 images of the ORL face database. All the 400 images in the ORL database without adding any noise are taken as the prototype image set. Hence we got 400 images in prototype set (40 subjects X 10 poses) and 400 images in probe set (40 subjects X 10 poses). The CGLPF feature set is formed by applying the CGLPF technique on both the sets and the signatures are used in experimental phase.

In the experimental phase, we take the first image of the first subject from the prototype image set as the query image and the top matching ten images are found from a set of all 400 probe images. If the top matching images lie in the same row (subject) of the prototype query image, then it is treated as a correct recognition. The number of correct recognized images for each query image in the prototype image set is calculated and the results are shown in figure 3 for Gaussian noise with mean 0.05 and variance 0.05.

The same procedure is repeated by using PCA, LDA and LPP method and the results are depicted in figures 4, 5, and 6 respectively. Figure 7 shows the comparison of overall percentage of recognition using CGLPF, PCA, LDA and LPP. It can be noted from this figure that, the CGLPF outperform the other existing techniques like PCA, LDA and LPP in the Gaussian noisy environment with mean and variance equal to 0.05.

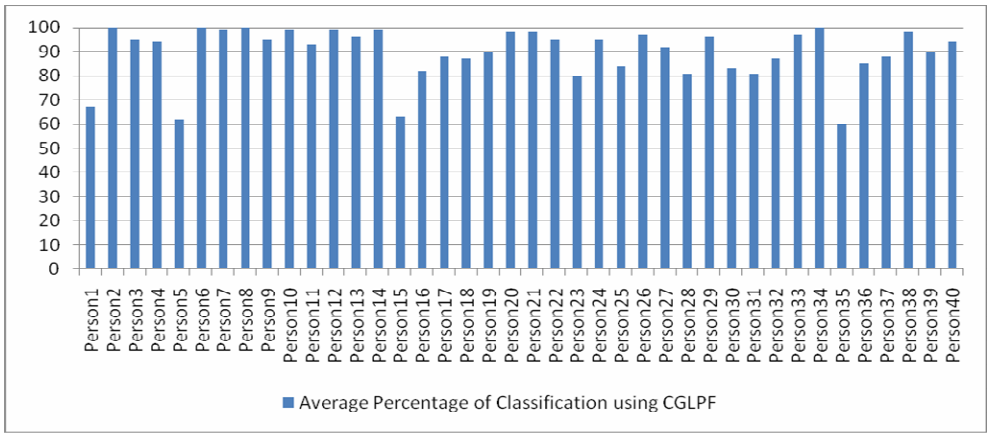


Fig.3.The average percentage of correct recognition obtained using CGLPF with Gaussian noise having mean 0.05 and variance 0.05

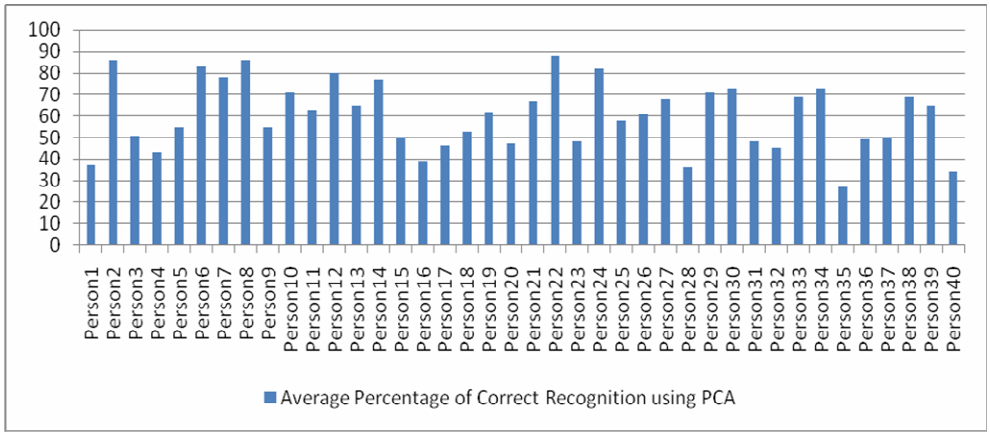


Fig.4.The average percentage of correct recognition obtained using PCA with Gaussian noise having mean 0.05 and variance 0.05

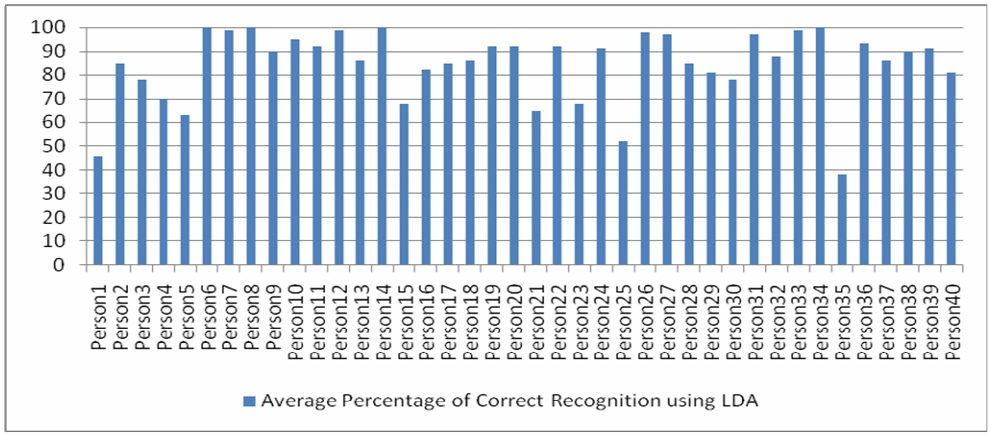


Fig.5.The average percentage of correct recognition obtained using LDA with Gaussian noise having mean 0.05 and variance 0.05

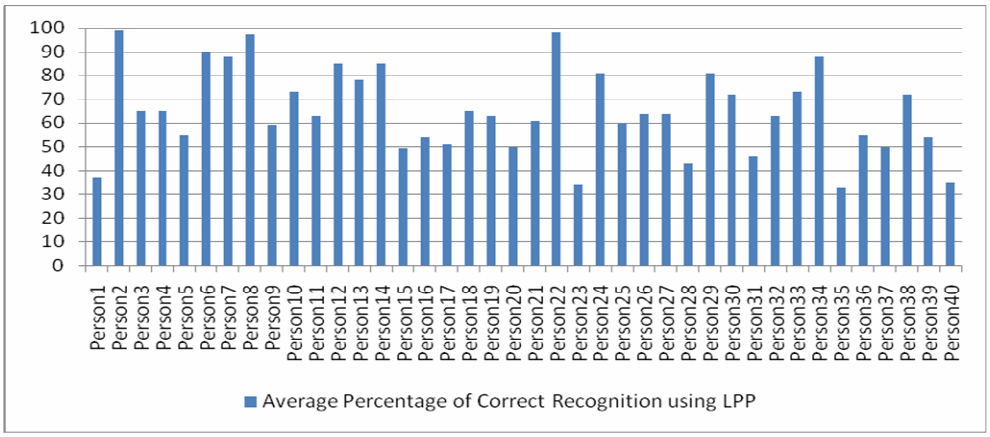


Fig.6.The average percentage of correct recognition obtained using LPP with Gaussian noise having mean 0.05 and variance 0.05

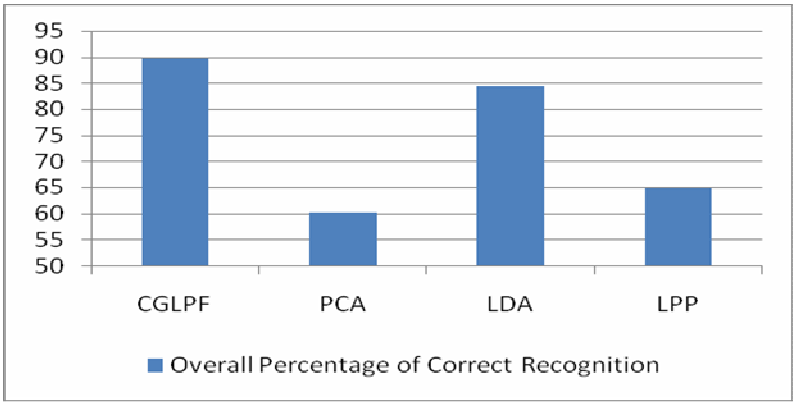


Fig.7. Comparison of overall percentage of correct recognition using CGLPF, PCA, LDA and LPP with Gaussian noise having mean 0.05 and variance 0.05

In the second part of our experiments, various other noises such as speckle, quantization and salt and pepper noises are applied by varying their respective parameters like mean and / or variance or quantization bits, in the probe images and various features of CGLPF, PCA, LDA and LPP algorithms are extracted. During the testing phase, the prototype images are taken one by one and the same features are extracted from it. The top ten matching images are taken and the numbers of correct matching images are counted. The overall percentage of correct recognition results obtained are tabulated in Table 1 for various noises with mean ranging from 0.05 to 0.2 and variance from 0.05 to 0.2. For most of the cases, our CGLPF algorithm performs better than other conventional techniques and it shows the high robustness of our proposed algorithm. For some cases, the LDA algorithm shows slightly improved results and it is observed that such cases use low variance value noises. In general, the high variance among the pixels increases the discrimination features among the local neighborhood pixels. Also the low variance exhibits the discrimination features among the global structure of the image. Hence when the variance becomes high, the added local features in the CGLPF method gives better results than the LDA which uses only the global structure information. Further, if the variance is low i.e., when the images possess high discrimination information in its global structure than local neighborhood, our CGLPF algorithm utilizes the global information preserved in it to produce good results.

Noise Details	Techniques			
	CGLPF	PCA	LDA	LPP
Gaussian Mean = 0, Variance = 0.05	90.9	64.75	90.875	68.45
Gaussian Mean = 0, Variance = 0.1	80.4	56.075	65.625	56.325
Gaussian Mean = 0, Variance = 0.15	76.7	44.05	63.35	40.675
Gaussian Mean = 0, Variance = 0.2	74.175	39.05	45.075	27.975
Gaussian Mean = 0.05, Variance = 0	92.6	65.575	93.1	74.55
Gaussian Mean = 0.05, Variance = 0.05	89.675	60.2	84.45	64.95
Gaussian Mean = 0.05, Variance = 0.1	74.325	51.725	67.5	48.75
Gaussian Mean = 0.05, Variance = 0.15	68.15	42.425	62.8	41
Gaussian Mean = 0.05, Variance = 0.2	60.175	37.525	44.625	30.7
Gaussian Mean = 0.1, Variance = 0	83.05	51.35	79.2	62.55
Gaussian Mean = 0.1, Variance = 0.05	80.075	49	76.15	52
Gaussian Mean = 0.1, Variance = 0.1	59.225	44.75	55.575	39.125
Gaussian Mean = 0.1, Variance = 0.15	58.275	39.875	51.675	33.775
Gaussian Mean = 0.1, Variance = 0.2	58.025	33.925	39.35	25.7
Gaussian Mean = 0.15, Variance = 0	59.05	30.725	55.5	41.725
Gaussian Mean = 0.15, Variance = 0.05	48.15	34.2	46.625	38.8
Gaussian Mean = 0.15, Variance = 0.1	46.95	34.775	46.45	30.2
Gaussian Mean = 0.15, Variance = 0.15	51.125	34.375	43	26.175
Gaussian Mean = 0.15, Variance = 0.2	50.775	27.925	34.975	23.525
Gaussian Mean = 0.2, Variance = 0	35.375	12.9	27.1	25.925
Gaussian Mean = 0.2, Variance = 0.05	44.475	21.75	37.65	23.3
Gaussian Mean = 0.2, Variance = 0.1	46.7	26.575	33.75	18.05
Gaussian Mean = 0.2, Variance = 0.15	41.35	26.675	30.4	23.025
Gaussian Mean = 0.2, Variance = 0.2	32.975	22.625	26.95	16.9
Speckle Variance = 0.05	95.875	68.5	96.925	74.5
Speckle Variance = 0.1	94	66.125	94.175	71.75
Speckle Variance = 0.15	93.425	63.725	90.2	70.175
Speckle Variance = 0.2	85.3	61.575	83.625	68.125
Quantization Bits Quantized = 1	96.925	69.35	93.05	77.3
Quantization Bits Quantized = 2	95.875	69.225	92.9	77.025
Quantization Bits Quantized = 3	94.25	68.825	92.825	77.025
Quantization Bits Quantized = 4	94	67.9	91.85	76.15
Salt & Pepper Variance = 0.05	95.8	66.9	94.725	74.525
Salt & Pepper Variance = 0.1	94.275	64.775	90.7	71.3
Salt & Pepper Variance = 0.15	92.825	60.075	80.125	67.275
Salt & Pepper Variance = 0.2	87.725	56.775	76.725	60.6

TABLE 1: Comparison of overall percentage of correct recognition obtained using CGLPF, PCA, LDA, and LPP under different noises with mean and variance ranging from 0 to 0.2 or quantization bits from 1 to 4.

Related to time complexity, it is the nature that the time complexity is increasing when using the combined schemes compared to using the techniques individually. But in our proposed method, the training is done offline and the testing is done in the real time or online. In the online phase, it is only going to project the testing image into the CGLPF feature set which is having only lower dimensions compared to the cases when the techniques are used individually. Hence when we employ our method in real time applications, there is no delay in the online and the offline delay does not cause any considerations in the real time image processing.

5. CONCLUSIONS

The robustness of CGLPF algorithm that combines the global and local information preserving features has been analyzed under various noisy environments such as Gaussian, speckle, quantization, and salt and pepper noise using ORL facial image database. In the feature set created using Laplacian faces in earlier papers, they use the PCA algorithm, only for reducing the dimension of the input image space whereas we use LDA algorithm for preserving the discriminating features in the global structure. Thus CGLPF feature set created using the combined approach retains both the global information and local information, in order to make the face recognition insensitive to most of the noises.

It is also observed that our proposed CGLPF algorithm shows the good robustness under different types of noisy conditions with respect to the percentage of correct recognition and in general it is superior to the conventional algorithms such as PCA, LDA and LPP. In our combined feature set, the preserved global features help to provide better robustness when the variance among the pixel intensities is high, while local feature preserved algorithm LDA shows better robustness when the variance is low. Therefore, the CGLPF feature set obtained through the combined approach would be an attractive choice for many facial related image applications under noiseless as well as noisy environments.

6. REFERENCES

1. X. He, S. Yan, Y. Hu, P.Niyogi, H. Zhang, '*Face recognition using Laplacian faces*', IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 27, no. 3,328–340, 2005.
2. K.J. Karande, S.N. Talbar, '*Independent Component Analysis of Edge Information for Face Recognition*', International Journal of Image Processing vol.3, issue 3, 120-130, 2009.
3. P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, '*Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*', IEEE Transactions on Pattern Analysis and Machine Intelligence vol.19, no.7, 711-720, 1997.
4. M. Belkin, P. Niyogi, '*Laplacian eigenmaps and spectral techniques for embedding and clustering*', Proceedings of Conference on Advances in Neural Information Processing System, 2001.
5. M. Belkin, P. Niyogi, '*Using manifold structure for partially labeled classification*', Proceedings of Conference on Advances in Neural Information Processing System, 2002.
6. S A Angadi, M. M. Kodabagi, '*A Texture Based Methodology for Text Region Extraction from Low Resolution Natural Scene Images*', International Journal of Image Processing vol.3, issue 5, 229-245, 2009.
7. C. Panda, S. Patnaik, '*Filtering Corrupted Image and Edge Detection in Restored Grayscale Image Using Derivative Filters*', International Journal of Image Processing vol.3, issue 3, 105-119, 2009.
8. Y. Chang, C. Hu, M. Turk, '*Manifold of facial expression*', Proceedings of IEEE International Workshop on Pattern Analysis, 2003.
9. P.Y. Simard, H.S. Malvar, '*An efficient binary image activity detector based on connected components*', Proceedings of. IEEE International Conference on Acoustics, Speech, and Signal Processing, 229–232, 2004.
10. L. Beaurepaire, K.Chehdi, B.Vozel, '*Identification of the nature of the noise and estimation of its statistical parameters by analysis of local histograms*', Proceedings of ICASSP-97, Munich, 1997.

11. Noise Models, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/VELDHUIZEN/node11.html
12. Image Noise, *http://en.wikipedia.org/wiki/Image_noise*
13. H.S.M. Al-Khaffaf , A.Z. Talib, R. Abdul Salam, '*A Study on the effects of noise level, cleaning method, and vectorization software on the quality of vector data*', Lecture Notes in Computer Science 299-309.
14. M. Turk, A. Pentland, '*Eigen Faces for Recognition*', Journal on Cognitive Neuroscience, 71-86, 1991.
15. K. Ruba Soundar, K. Murugesan, '*Preserving Global and Local Information – A Combined Approach for Recognizing Face Images*', International Journal of Pattern Recognition and Artificial Intelligence, accepted for publication.
16. Speckle Noise, *http://en.wikipedia.org/wiki/Speckle_noise*
17. Rafael C. Gonzalez, Richard E. Woods, '*Digital Image Processing*'. Pearson Prenctice Hall, (2007).
18. B. Widrow, I. Kollár, '*Quantization Noise: Roundoff Error in Digital Computation*', Signal Processing, Control, and Communications, Cambridge University Press, Cambridge, UK, 778-787, 2008.
19. W. Zhao, R. Chellappa, P.J. Phillips, '*Subspace linear discriminant analysis for face recognition'*, Technical Report CAR-TR-914, Center for Automation Research, Univ. of Maryland, 1999.
20. X. He, P. Niyogi, '*Locality preserving projections*', Proceedings of Conference on Advances in Neural Information Processing Systems, 2003.
21. A. Jose, Diaz-Garcia, '*Derivation of the Laplace-Beltrami operator for the zonal polynomials of positive definite hermitian matrix argument*', Applied Mathematics Sciences, Vol.1, no.4, 191-200, 2007.

Repeat-Frame Selection Algorithm for Frame Rate Video Transcoding

Yi-Wei Lin m9723030@ems.ndhu.edu.tw
Gwo-Long Li m9323004@ems.ndhu.edu.tw
Mei-Juan Chen cmj@mail.ndhu.edu.tw
Department of Electrical Engineering
National Dong-Hwa University
Hualien, 97401 Taiwan, R.O.C.

Chia-Hung Yeh*(Corresponding author) yeh@mail.ee.nsysu.edu.tw
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, 80424 Taiwan, R.O.C.

Shu-Fen Huang m9823002@ems.ndhu.edu.tw
Department of Electrical Engineering
National Dong-Hwa University
Hualien, 97401 Taiwan, R.O.C.

Abstract

To realize frame rate transcoding, the forward frame repeat mechanism is usually adopted to compensate the skipped frames in a video decoder for end-device. However, based on our observation, it is unsuitable for repeating all skipped frames only in the forward direction and sometimes the backward repeat may provide better results. To deal with this issue, we propose a new reference frame selection method to determine the direction of repeat-frame for skipped Predictive (P) and Bidirectional (B) frames. For P-frame, the non-zero transformed coefficients and the magnitude of motion vectors are taken into consideration to determine the use of forward or backward repeat. For B-frame, the magnitude of motion vector and its corresponding reference directions of the blocks in B-frame are selected as the decision criteria. Experimental results show that the proposed method provides 1.34 dB and 1.31 dB PSNR improvements in average for P and B frames, respectively, compared with forward frame repeat.

Keywords: Transcoding, Temporal transcoding, Frame-rate transcoding, Frame skipping, Forward/Backward repeat.

1. INTRODUCTION

In recent years, the applications of multimedia [1]-[4] are rising and popular. One of applications, video transcoding becomes an important issue in video communication with the development of the network transmission. Video transcoder converts videos into different qualities, frame rates, resolutions, even the coding standards [5]-[8] to fit the network variation. The concept of video transcoding [9] is shown in FIGURE 1. When network bandwidth is insufficient, three kinds of methods can be used to convert a bitstream into different bitrates and they are quality transcoding, spatial transcoding and temporal transcoding that is also named as frame rate transcoding. For quality transcoding [10]-[12], the quantization parameter (QP) is adjusted in encoder to fit target bitrate under the bitrate constraint. In addition, another way to achieve video transcoding is to adjust the spatial resolution of a sequence for transmission purpose [13]-[19]. In spatial transcoding, the shrunk sequence saves the bitrate and the decoder recovers the sequence to the original size when receiving data. However, the way to choose the down-scaling and up-scaling method is still a great challenging issue. Sometimes, we adjust the frame rate of a sequence to fit the target bitrate.

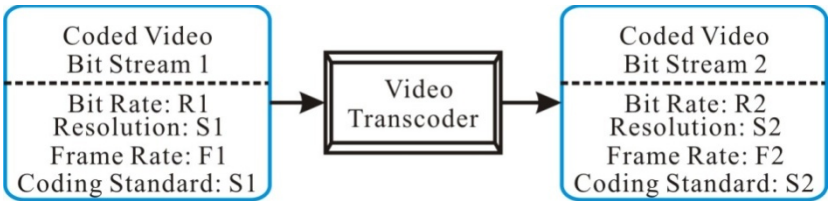
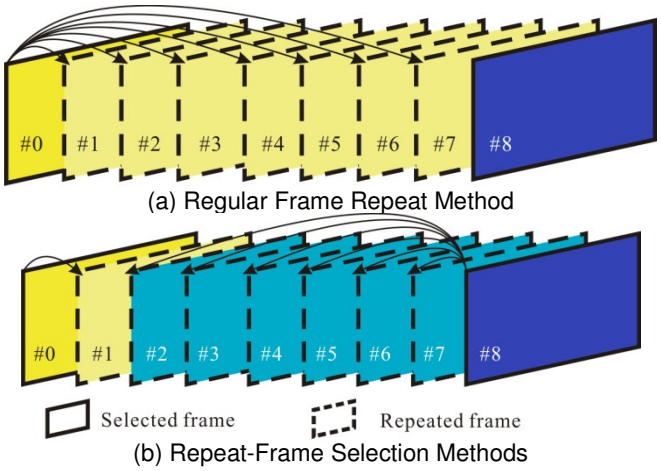


FIGURE 1: Concept of Video Transcoding

Many methods have been proposed for temporal transcoding and most of them focus on two directions, frame rate decision and key frame selection [20]-[28]. That is, we should decide the acceptable frame rate according to the current bandwidth and select the most significant frames in a group of picture (GOP). After that, a traditional video decoder repeats the previous frame to compensate skipped frames shown in FIGURE 2 (a) for GOP=8 and we call it as "Regular Forward Repeat Method" (RFRM). In lots of experiments, we observe that repeating all skipped frames in the forward direction seems not appropriate and backward repeat as shown in FIGURE 2 (b) may have better results.

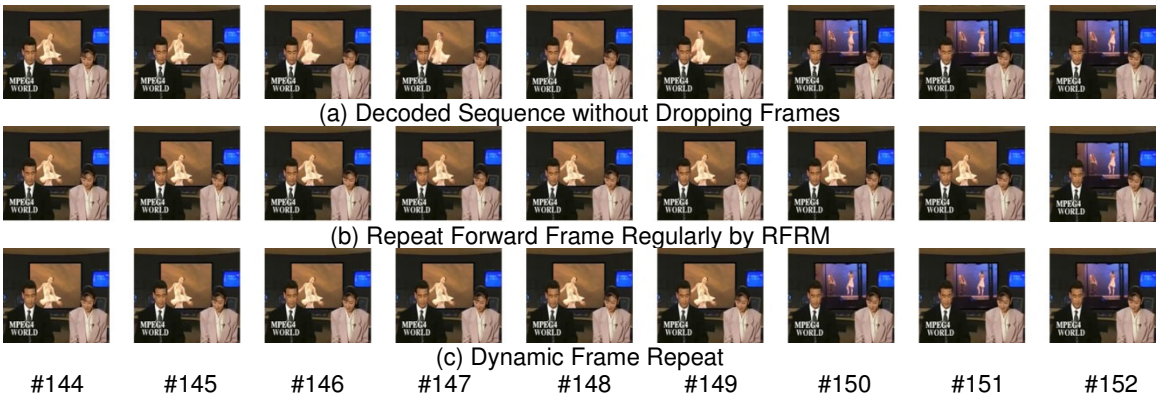


(a) Regular Frame Repeat Method

(b) Repeat-Frame Selection Methods

FIGURE 2: An Example for Sequence in GOP=8

FIGURE 3 shows an example of the benefit of backward frame repeat. FIGURE 3 (a) shows the decoded sequence without dropping frames and FIGURE 3 (b) is transcoded by frame rate descending with forward repeat, and FIGURE 3 (c) shows dynamic forward or backward frame repeat. From this example, it can be seen that the results after temporal transcoding can be improved significantly by considering both forward and backward repeats. Based on the observation mentioned above, this paper proposes a reference frame repeat method to determine the direction of repeat-frame for skipped frames during transcoding process. For P-frame, the number of nonzero transformed coefficients and the magnitude of motion vectors are jointly considered to determine the repeat direction. For B-frame, the prediction directions and the magnitude of motion vectors are combined to obtain the criteria for the repeat direction determination. This paper is organized as follows. In Section 2, the proposed "Repeat-Frame Selection Methods" (RFSM) is explicated. Section 3 presents the extensive experimental results to verify the efficiency of our methods. Finally, concluding remarks are given in Section 4.



(a) Decoded Sequence without Dropping Frames

(b) Repeat Forward Frame Regularly by RFRM

(c) Dynamic Frame Repeat

FIGURE 3: An Example of Different Methods in *News* Sequence for GOP=8

2. PROPOSED METHODS

FIGURE 4 shows the system diagram for the proposed system. We propose the RFSM to determine that the skipped frames should be repeated by either forward or backward direction. Instead of fully reconstructing pixel data, the proposed algorithm just employs the motion vector information, non-zero transformed coefficients and the prediction directions of B-frame partially decoded from the bitstream1 to determine the repeat direction. The Encoder2 in temporal transcoder will embed the results of forward/backward decision into Bitstream2 by inserting repeat direction into the headers. After receiving the Bitstream2, the Decoder2 can decompress the video bitstream and dynamic forward/backward frame repeat can be executed. The proposed methods for P-frame and B-frame cases are discussed in the following two subsections separately.

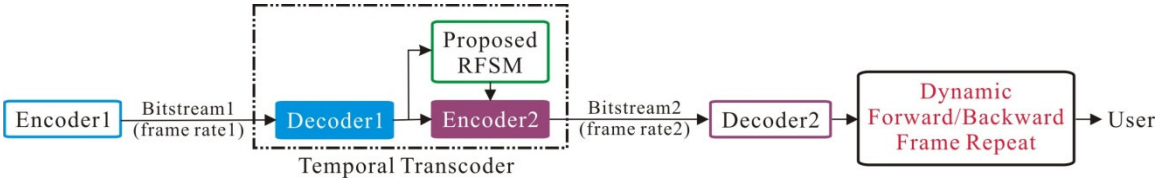


FIGURE 4: Diagram of Proposed Repeat-Frame Selection Method

2.1 P frames

For P-frame, since only the information of forward direction is available, the factors we consider are the magnitude of motion vectors and the number of non-zero transformed coefficients. Normally, the motion activity and the number of non-zero transformed coefficients indicate the property of the sequence and the complexity of the frame, respectively. From our observation, the high-motion sequence results in obvious amount of non-zero transformed coefficients. In our proposal, we define a selective factor (SF) as follows.

$$SF_P = \sum_{i=1}^{N} (NZcoeff_i \times MV_i), \quad (1)$$

$$MV_i = \sum_{k=1}^{M_i} (|MVX_k| + |MVY_k|), \quad (2)$$

where N refers to the number of the macroblock in one frame, M_i refers to the number of blocks in the i-th macroblock, MV_i refers to the sum of the motion vector magnitude in X and Y directions and $NZcoeff_i$ refers to the number of non-zero transformed coefficients in the i-th macroblock. After the SF_P of each skipped frame is calculated, we select the frame with the maximum SF_P as the separated frame which means two consecutive frames have higher motion activity variation shown in FIGURE 5. Finally, the frames after the separated frame (including the separated frame)

in a GOP are assigned as backward repeat. Take FIGURE 2(b) for example, the frame #2 is the separated frame in a GOP. In FIGURE 5, the frame #258 is the separated frame.

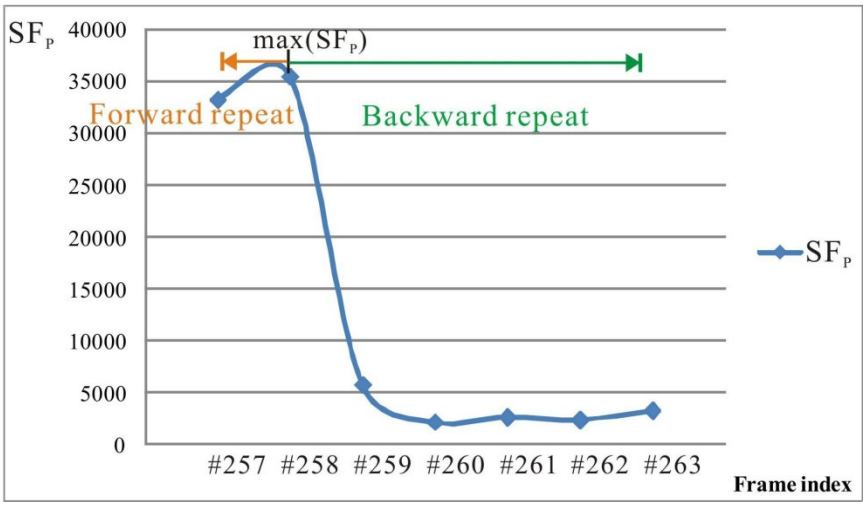


FIGURE 5: Separated Frame Determination of *Foreman* Sequence for P-Frame in GOP=8 case

2.2 B frames

For B-frame, we first decode the bitstream to obtain the magnitude of motion vectors and the prediction directions of each block. We avoid reconstructing pixel values in order to reduce computational complexity. It is well-accepted that the larger magnitude of motion vector implies the higher motion activity or scene change that the frame may contain. Therefore, the magnitude of motion vectors is selected as a factor and its prediction direction of decoded blocks is also included in the proposed method. If a block is encoded as forward prediction mode, it means that the most similar block is forward prediction rather than backward. Oppositely, backward prediction mode implies that the best match block can be found from the following frames. As a result, we take the magnitude of the motion vector in Forward or Backward prediction to be the factor of separated frame determination. SF_B is defined as the difference of two factors, $MV_{Forward}$ and $MV_{Backward}$, to stand for the motion tendency of forward and backward prediction directions in a frame.

$$SF_B = MV_{Backward} - MV_{Forward}\,, \tag{3}$$

$$MV_{Forward} = \sum_{u=1}^{U} (|MVX_{Forward\_u}| + |MVY_{Forward\_u}|)\,, \tag{4}$$

$$MV_{Backward} = \sum_{v=1}^{V} (|MVX_{Backward\_v}| + |MVY_{Backward\_v}|)\,, \tag{5}$$

where *MVX* and *MVY* represent the magnitude of motion vectors in *X* and *Y* directions for each block, respectively, and *U and V* are the numbers of blocks in forward and backward directions, respectively, of one frame. Finally, we select the frame with the maximum SF_B to be the separated frame in a GOP shown in FIGURE 6. Once the separated frame is determined, the frames after the separated frame (including separated frame) in a GOP are assigned as backward repeat.

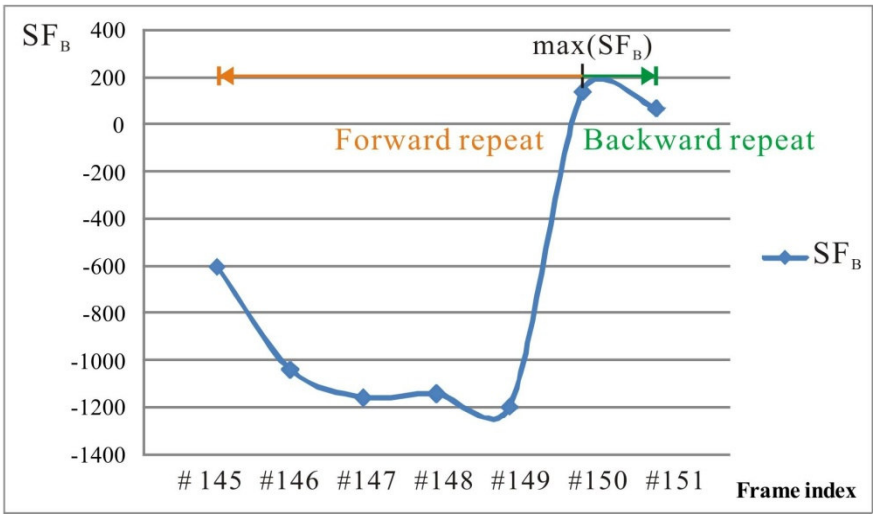


FIGURE 6: Separated Frame Determination of *News* Sequence for B-Frame in GOP=8 case

3. EXPERIMENTAL RESULTS

In this section, we compare the proposed method with *SAD-based frame repeat* method and RFRM in terms of subjective and objective qualities to demonstrate the efficiency of our method. *SAD-based* method decodes all frames to the pixel-domain and calculates the Sum of the Absolute Difference (SAD) between the forward and the backward reference frames. Afterwards, SAD is used to determine the forward and backward repeat. In *SAD-based* method, the skipped frame which first satisfies the condition of the backward SAD less than forward SAD will be selected as the separated frame. All methods are implemented on the H.264 JM 15.1 [29] reference software. The simulation setting is in the following. The test benchmark sequences include *CarPhone*, *Foreman*, *Mobile*, *News*, *Salesman* and *Silent* in QCIF resolution with 289 frames and the search range is 16. The coding structures are IPPP in P-frame and IBBBP in B-frame for GOP=4 and so on.

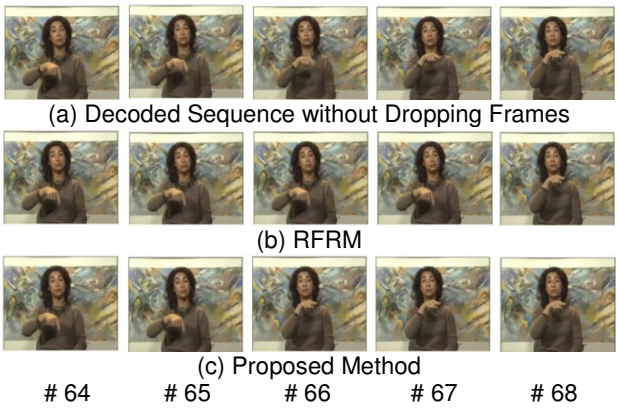


FIGURE 7: Subjective Comparison of *Silent* Sequence for P-Frame in GOP=4

FIGURE 7 to FIGURE 11 show the subjective quality comparisons for P-frame and B-frame of RFRM and our method. Take FIGURE 8 as example, the worker's hand waved out the scene in #258. Our proposed method selects the proper frame to repeat. It is evident that our proposed algorithm is very similar to the original sequence.

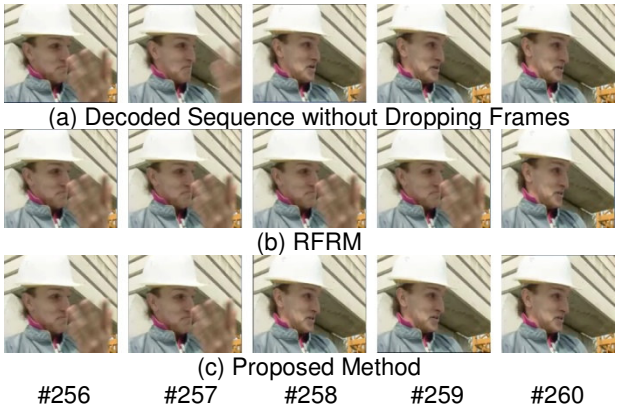


FIGURE 8: Subjective Comparison of *Foreman* Sequence for B-Frame in GOP=4

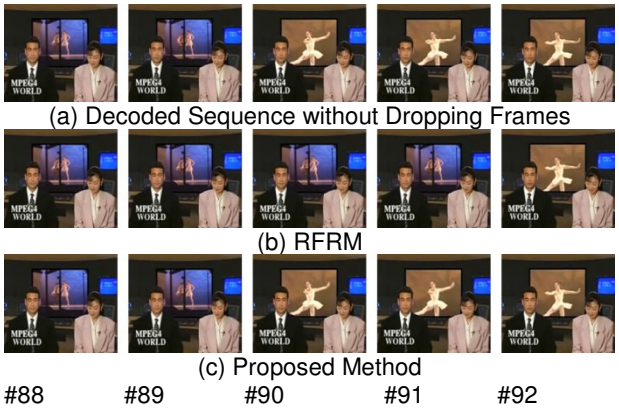


FIGURE 9: Subjective Comparison of *News* Sequence for B-Frame in GOP=4

FIGURE 10: Subjective Comparison of *CarPhone* Sequence for P-Frame in GOP=6

FIGURE 11: Subjective Comparison of *Foreman* Sequence for P-Frame in GOP=8

	GOP 4				GOP 6				GOP 8			
	SAD-based method	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*
CarPhone	24.04	22.70	23.63	0.93	25.72	24.15	25.27	1.12	26.54	24.45	25.86	1.41
Foreman	20.93	19.03	20.34	1.31	21.92	19.63	21.24	1.61	22.21	19.54	21.76	2.22
Mobile	17.87	16.04	17.22	1.18	18.43	16.18	17.69	1.51	18.11	15.95	17.11	1.16
News	23.16	21.85	22.66	0.81	25.18	23.61	24.30	0.69	24.87	23.21	24.22	1.01
Salesman	28.25	26.55	27.88	1.33	30.35	28.41	29.84	1.43	30.95	28.83	30.70	1.87
Silent	24.27	22.84	24.09	1.25	26.19	24.40	25.87	1.47	26.72	24.69	26.46	1.77
Average	**23.09**	**21.50**	**22.64**	**1.14**	**24.63**	**22.73**	**24.04**	**1.31**	**24.90**	**22.78**	**24.35**	**1.57**

TABLE 1: PSNR Comparison for P-Frame (dB)

	GOP 4				GOP 6				GOP 8			
	SAD-based method	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*
CarPhone	24.25	22.79	23.95	1.16	25.95	24.2	25.60	1.40	26.68	24.51	26.34	1.83
Foreman	21.04	19.08	20.77	1.69	22.00	19.67	21.58	1.91	22.33	19.60	21.98	2.38
Mobile	18.10	16.21	16.87	0.66	18.63	16.32	16.94	0.62	18.30	16.08	16.47	0.39
News	22.66	21.91	22.42	0.51	25.01	23.68	24.04	0.36	25.04	23.31	24.63	1.32
Salesman	28.09	26.95	28.15	1.20	30.47	28.53	30.28	1.75	30.82	28.70	30.59	1.89
Silent	24.77	23.19	24.46	1.27	26.73	24.78	26.21	1.43	26.85	24.69	26.42	1.73
Average	**23.15**	**21.69**	**22.77**	**1.08**	**24.80**	**22.86**	**24.11**	**1.25**	**25.00**	**22.82**	**24.41**	**1.59**

TABLE 2: PSNR Comparison for B-Frame (dB)

	P Frame		B Frame	
	RFRM	***Proposed***	***RFRM***	***Proposed***
GOP 4	75.1%	74.5%	72.6%	72.4%
GOP 6	82.9%	82.3%	81.1%	80.7%
GOP 8	87.1%	86.4%	85.2%	84.9%

TABLE 3: Decreased Computational Complexity compared with *SAD-based* Method

TABLE 1 and TABLE 2 show the PSNR comparisons for P-frame and B-frame. In those tables, we found that the PSNR improvements of our method is 1.14 dB, 1.31 dB and 1.57 dB for GOP=4, GOP=6 and GOP=8, respectively, when compared with RFRM method for P-frame case, and 1.08 dB, 1.25 dB and 1.59 dB for GOP=4, GOP=6 and GOP=8, respectively, when compared with *RFRM* method for B-frame case. Furthermore, when compared with *RFRM* method, the average PSNR improvements of all GOP sizes are 1.34 dB and 1.31 dB for P-frame and B-frame cases, respectively. TABLE 3 shows the decreased computational complexity, which is measured by CPU time, compared with *SAD-based* method. *SAD-based* method needs to decode bitstream to reconstruct pixel values for SADs calculation while RFRM and our proposal need not, which save much encoding time. Our algorithms only utilize the magnitude of motion vectors, non-zero transformed coefficients and motion compensation directions to determine frame repeat direction. Although our proposed method increases the computational complexity slightly when compared with RFRM, the quality can be increased significantly.

4. CONCLUSIONS

In this paper, in place of the traditional regular forward repeat method for frame-rate transcoding, we propose efficient algorithms which can dynamically select suitable frame to repeat. Our experimental results show that the proposed method has significant PSNR improvements

compared with traditional forward repeat. The proposed method can select the proper frame to repeat and achieve better subjective quality.

5. REFERENCES

1. N. A. Lili and K. Fatimah. "*Content modelling for human action detection via multidimensional approach*". International Journal of Image Processing, 3(1):17-30, 2009
2. F. Rohmad and Abdul Azim Abd Ghani. "*Empirical evaluation of decomposition strategy for wavelet video compression*". International Journal of Image Processing, 3(1):31-54, 2009
3. Y. Tabii and R.O.H. Thami. "*A framework for soccer video processing and analysis based on enhanced algorithm for dominant color extraction*". International Journal of Image Processing, 3(4):131-142, 2009
4. M.C. Chi, C.H. Yeh and M.J. Chen. "*Robust region-of-interest determination based on user attention model through visual rhythm analysis*". IEEE Transactions on Circuits and Systems for Video Technology, 19(7):1025-1038, 2009
5. H.F. Shen, X.Y. Sun and F. Wu. "*Fast H.264/MPEG-4 AVC transcoding using power-spectrum based rate-distortion optimization*". IEEE Transactions on Circuits and Systems for Video Technology, 18(6):746-755, 2008
6. A. Dziri, A. Diallo, M. Kieffer and P. Duhamel. "*P-picture based H.264 AVC to H.264 SVC temporal transcoding*". In Proceedings of IEEE International Wireless Communications and Mobile Computing Conference, 2008
7. X. Jing, W.C. Siu, L.P. Chau and A.G. Constantinides. "*Fast intra mode decision algorithm for H.263 to H.264/AVC transcoding*". In Proceedings of IEEE International Conference on Neural Networks and Signal Processing, 2008
8. M. Pantoja and N. Ling. "*Adaptive transform size and frame-field selection for efficient VC-1 to H.264 high profile transcoding*". In Proceedings of IEEE International Symposium on Circuits and Systems, 2009
9. J. Xin, C.W. Lin and M.T. Sun. "*Digital video transcoding*". Proceedings of the IEEE, 93(1):84-97, 2005
10. X. Xiu, L. Zhuo and L. Shen. "*A H.264 bit rate transcoding scheme based on PID controller*". In Proceedings of IEEE International Symposium on Communications and Information Technology, 2005
11. X.Y Wang, Y. Zhang, H.L. Li and W.L. Zhu. "*Adaptive rate control for dynamic bandwidth in video transcoding*". In Proceedings of IEEE International Conference on Communication Systems, 2008

12. Y.M. Zhou, Y. Sun, Z.D. Feng and S.X. Sun. "*New rate-complexity-quantization modeling and efficient rate control for H.264/AVC*". In Proceedings of 2008 IEEE International Conference on Multimedia and Expo, 2008
13. M.J. Chen, M.C. Chu and S.Y. Lo. "*Motion vector composition algorithm for spatial scalability in compressed video*". IEEE Transactions on Consumer Electronics, 47(3):319-325, 2001
14. P.Z., Y. Lu, Q. Huang and W. Gao. "*Mode mapping method for H.264/AVC spatial downscaling transcoding*". In Proceedings of IEEE International Conference on Image Processing, 2004
15. H. Sun and Y.P. Tan. "*Arbitrary downsizing video transcoding using H.26L standard*". In Proceedings of IEEE International Conference on Image Processing, 2003
16. H. Sun and Y.P. Tan. "*Fast motion re-estimation for arbitrary downsizing video transcoding using H.264/AVC standard*". IEEE Transactions on Consumer Electronics, 50(3):887-894, 2004
17. M. Pantoja and N. Ling. "*Transcoding with quality enhancement and irregular sampling*". In Proceedings of IEEE International Conference on Image Processing, 2008
18. H.Y. Shu and K. N. Ngan. "*Pre- and post-shift filtering for blocking removing in downsizing transcoding*". IEEE Transactions on Circuits and Systems for Video Technology, 19(6):882-886, 2009
19. X. Yu, E.H. Yang and H. Wand. "*Down-sampling design in DCT domain with arbitrary ratio for image/video transcoding*". IEEE Transactions on Image Processing, 18(1):75-89, 2009
20. M.A. Bonuccelli, F. Lonetti and F. Martelli, "*Temporal transcoding for mobile video communication*". In Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous System: Networking and Service, 2005
21. Y.H. Ho, W.R. Chen and C.W. Lin. "*A rate-constrained key-frame extraction scheme for channel-aware video streaming*". In Proceedings of IEEE International Conference on Image Processing, 2004
22. C.T. Hsu, C.H. Yeh, C.Y. Chen and M.J. Chen. "*Arbitrary frame rate transcoding through temporal and spatial complexity*". IEEE Transactions on Broadcasting, 55(4):767-775, 2009
23. J.N. Youn, M.T. Sun and C.W. Lin. "*Motion vector refinement for high-performance transcoding*". IEEE Transactions on Multimedia, 1(1):30-40, 1999
24. M.J. Chen, M.C. Chu and C.W. Pan. "*Efficient motion-estimation algorithm for reduced frame-rate video transcoder*". IEEE Transactions on Circuits and Systems for Video Technology, 12(4): 269-275, 2002

25. I.H. Shin, Y.L. Lee and H.W. Park. "*Motion estimation for frame-rate reduction in H.264 transcoding*". In Proceedings of 2nd IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, 2004
26. C.T. Hsu, C.H. Yeh and M.J. Chen. "*Effective frame rate decision by Lagrange optimization for frame skipping video transcoding*". In Proceedings of International Symposium on Visual Computing, 2008
27. K.C. Yang, G. Dane and K. El-Maleh. "*Temporal quality evaluation for enhancing compressed video*". In Proceedings of IEEE International Conference on Computer Communications and Networks, 2007
28. V. Chander, A. Reddy, S. Gaurav, N. Khanwalkar, M. Kakhani and S. Tapaswi. "*Fast and high quality temporal transcoding architecture in the DCT domain for adaptive video content delivery*". In Proceedings of IEEE International Conference Computer Engineering and Technology, 2009
29. Joint Video Team software JM15.1 http://iphome.hhi.de/suehring/tml/download/

Water-Body Area Extraction from High Resolution Satellite Images-An Introduction, Review, and Comparison

Rajiv Kumar Nath rajivknath@gmail.com
Research Scholar, Department of Civil Engineering
IIT Delhi
New Delhi, 110016, India

S K Deb skdeb@hotmail.com
Assistant Professor, Department of Civil Engineering
IIT Delhi
New Delhi, 110016, India

Abstract

Water resources play an important role in environmental, transportation and region planning, natural disaster, industrial and agricultural production and so on. Surveying of water-bodies and delineate its features properly is very first step for any planning, especially for places like India, where the land-cover is dominated by water-bodies. Recording images, such as from satellite, sometimes does not reflect the distinguished characteristics of water with non-water features, e.g. shadows of super structures. Image of water body is confused easily with the shadow of skyscraper, since calm water surface induces mirror reflection when it gives birth to echo wave. Water transport is cheapest. Developing/poor countries like India will be benifitted if water transport is encouraged. In water transport, the link should be made between various land masses, including building blocks, through proper navigational system. Hence there should be clear distinction between calm water and the shadows of buildings. Over the past decade, a significant amount of research been conducted to extract the water body information from various multi-resolution satellite images. The objective of this paper is to review methodologies applied for water body extraction using satellite remote sensing. The Geographic Information System (GIS) and the Global Positioning System (GPS) have also been discussed as they are closely linked with Remote Sensing. Initially, studies on water body detection are treated. Methodological issues related to the use of these methods were analysed followed by summaries. Results from empirical studies, applying water-body extraction techniques are collected and discussed. Important issues for future research are also identified and discussed.

Keywords: Feature extraction, multi-resolution satellite image, remote sensing, and water body.

1. INTRODUCTION

Watershed is a region (or area) delineated with a well-defined topographic boundary and water outlet. It is a geographic region within which hydrological conditions are such that water becomes concentrated within a particular location, for example, ocean, sea Lake, a river, or a reservoir, by which the watershed is drained. Within the topographic boundary or a water divide, watershed comprises a complex of soils, landforms, vegetations, landform and land uses. The terms watershed, catchment, and basins are often considered synonyms [1]. Remote sensing, defined as the science of using an instrument for measuring a target and its properties from a remote location, without a physical connection between the measuring instrument and the target, which is to be featured. Typically, the measurements are performed through various techniques. Those techniques are electromagnetic radiation (e.g. ultra-violet, visible light, reflective, thermal infrared, microwaves, etc.). The instrument records the radiation reflected or emitted by the target and its properties are then inferred from the measured signal.

One of the advantages of remote sensing is that the measurements can be performed from a great distance (several hundred or even several thousand kilometers in the case of satellite sensors), which means that large areas on ground can be covered easily. With satellite instruments it is also possible to observe, a target repeatedly; in some cases every day or even several times per day.

Classification is a widely studied issue in remote sensing image processing. The common application ranges from land use analysis to change detection. Among the classes of interest, urban areas, farmland, forest, and river/lake areas are traditionally selected. The observation of water body from remote sensing images, is of particular importance during these recent years for two main reasons: (i) there is a world-wide an important need to assess existing water resource and water resource changes –because of the increasing water scarcity and related problems; (ii) the so-called “climate change” affects directly and is directly affected by water cycling; (iii) study of water bodies may help to develop water transport route, either by using existing one directly or connecting the existing one by preparing canals to develop a longer water route; (iv) timely information of water increase in hills and mountains may help to develop some strategy to restrict flood calamities. Remote sensing and its allied techniques such as geographic information system have a pervasive impact on the conduct of practical work. The application of these are in business, ecology, engineering, forestry, geography, geology, urban and regional planning, water resources management, transportation engineering or environmental science Remote sensing data provides a mean to observe and analyze some of the related phenomena, such as flood disasters and land use change. There exist a close interaction among the related areas of remote sensing, GIS, GPS, digital image processing and environmental, transportation and regional medelling.

The ability to map open surface water is an integral part to many hydrologic and agricultural models, wildlife management programmes, and recreational and natural resource studies. The study of X-band *HH* polarized airborne Synthetic Aperture Radar (SAR) imagery to examine the potential of SAR data to map open fresh water areas extant on 1:100000 USGS topographic maps [2] and SAR image based on technique of imaging in different directions and object-oriented [3]. The remote sensing- GIS techniques used for identification of various land-use classes on satellite imagery and enhanced products and identification of time-sequential changes in land-use patterns [4]. A new model based on EOS/MOSDIS model which can segment the water body and extract by the criteria of NDWI<-0.1 or NDVI<0.04 & (CH4-CH5)>2.0 [5]. The decision tree and programming method is used for extracting water body information from flood affected region [6],[7]; semi-automated change detection approach is used for extracting water feature form satellite image [8],[9],[10],[11]; an automatic extraction method is used for extracting water body from IKONOS and other high resolution satellite image [12],[13],[14]; Thresholding and multivariate regression method [15], A conceptual clustering technique and dynamic Thresholding [16], an original entropy based method [17]. The water body can be extracted by

classification; unsupervised classification [18]; The Euclidean Classifier and the Eigenvector Classifier [19]; The SVM with One-Against-One (1A1) and One-Against-All (1AA) techniques is used for land cover mapping [20] ; A supervised classification algorithm [21],[22] of remote sensing satellite image that uses the average fuzzy intra cluster distance within the Bayesian algorithm [23],[24]; sometimes combination of supervised and unsupervised classification is used also called automated [25], The edge detection algorithm [26], the data fusion technique is used to characterize and delineate 1993 flood damage in the Midwest of St. Luis, USA [27]; a remote sensing and Geographical Information System (GIS) to estimate and hindcast water quality changes using historical land use data for a watershed in eastern England [28],[29]. Some researches focused on the water quality of the specific water body, in this; first, we extract the water body then assess the water quality [30], [31], [32].

2. SATELLITES AND SENSORS APPLIED IN WATER BODY EXTRACTION

A large number of earth observation satellites has orbited, and is orbiting our planet to provide frequent imagery of its surface. From these satellites, many can potentially provide useful information for assessing erosion, although less has actually been used for this purpose. This section provides a brief overview of the space borne sensors applied in water-body extraction studies. The sensors can be divided in those measuring reflection of sunlight in the visible and infrared part of the electromagnetic spectrum and thermal infrared radiance (optical systems), and those actively transmitting microwave pulses and recording the received signal (imaging radars).

Optical satellite systems are most frequently been applied in water body extraction research. The parts of the electromagnetic spectrum covered by these sensors include the visible and near-infrared (VNIR) ranging from 0.4 to 1.3 μm, the shortwave infrared (SWIR) between 1.3 and 3.0 μm, the thermal infrared (TIR) from 3.0 to 15.0 μm and the long-wavelength infrared (LWIR) from (7-14 μm). Table 1 summarizes sensor characteristics of the systems used [33], [34], [35].

Satellite	Sensors	Operation Time	Spatial resolution	# spectral bands	Spectral domain
Landsat-1,2,3	MSS	1972-1983	80m	4	VNIR
NOAA/ TIROS	AVHRR	1978-present	1001m	5	VNIR, SWIR, TIR
Nimbus-7	CZCS	1978-1986	825m	6	VNIR
Landsat-4,5	TM	1982-1999	30m 120m	6 1	VNIR, SWIR TIR
SPOT-1,2,3	HRV	1986-present	10m 20m	1 3	VNIR VNIR
IRS-1A,1B	LISS-1 LISS-2	1988--1999	72.5m 36.25m	4 4	VNIR VNIR
IRS-1C,1D	PAN LISS-3	1995-present	5.8m 23.5m 70m	1 3 1	VNIR VNIR SWIR
SPOT-4	HRVIR	1998-present	10m 20m	1 4	VIS VNIR, SWIR
IKONOS	Panchromatic Multispectral	1999-present	1m 4m	1 4	VNIR VNIR
Landsat-7	ETM	1999-present	15m 30m 60m	1 6 1	VNIR VNIR TIR
Terra	ASTER MODIS	1999-present 1999-present	15m 30m 90m 250m 500m 1000m	3 6 5 2 5 29	VNIR SWIR TIR VIS NIR SWIR/ MWIR LWIR
Quick Bird	Panchromatic Multispectral	2001-present	0.61m 4m	1 4	VNIR VNIR
SPOT-5	Panchromatic Multispectral	2002- present	5m 10 10m 20m	1 4	VNIR NIR SWIR
WorldView-1	Panchromatic	2007-present	0.55m	1	VIR NIR
GEOEYE-1	Pan-sharpened Panchromatic Multispectral	2008-present	0.41m 0.41m 1.65m	3 1 4	VIR NIR

TAB

LE 1: Overview of optical satellite sensors applied in water body extraction

Landsat is still among the widest used satellites, partly because it has the longest time series of data of currently available satellites. The first satellites of the Landsat family were equipped with the Multispectral Scanner (MSS), having four bands at 80-m resolution. AVHRR (Advanced Very

High Resolution Radiometer) has five bands in 1.1-km resolution and has been flown on many platforms, including TIROS-N (Television Infrared Observation System) and several NOAA-satellites (National Oceanic and Atmospheric Administration).

Later Landsat satellites had the Thematic Mapper (TM) sensors onboard with improved resolution and more spectral bands. The SPOT series of satellites started acquiring data in 1986 with the HRV-sensor (High Resolution Visible). The HRV-sensor has a 10-m panchromatic mode and a three band 20-m resolution multispectral mode. The Indian Remote Sensing Satellites (IRS) 1A and 1B both have two sensors called LISS-1 and LISS-2 (Linear Imaging and Self-Scanning Sensor), which are identical except for a two times higher spatial resolution on LISS-2. IRS 1C and 1D also have an identical payload being a 5.8-m resolution panchromatic camera (PAN) and a 23.5-m resolution multispectral sensor called LISS-3. SPOT-4 flew the HRVIR-sensor (High Resolution Visible Infrared) on which a SWIR band was added. IKONOS and QuickBird are both high-resolution satellites, with a spatial resolution in panchromatic mode of 0.61 and 1.00 m respectively, and 2.44 and 4.00 m in multispectral mode. The start of space borne imaging radar instruments was in 1978 with the SAR (synthetic aperture radar) onboard SEASAT, operating in L-band (23.5-cm wavelength) during 105 days only. For erosion studies, only five SAR sensors have been applied, which were flown on ERS-1 and 2, JERS-1, RADARSAT-1, and ENVISAT respectively. In 1991, ERS-1 launched with the Active Microwave Instrument (AMI) onboard operating in C-band (5.7-cm wavelength). The SAR image mode of AMI acquired data at 30-m resolution. ERS-2 flies the same instrument and has been operational from 1995 to till present. JERS-1 (Japanese Earth Resources Satellite) flew an 18-m resolution L-band SAR (23.5-cm wavelength), recording data from 1992 to 1998. RADARSAT-1 has acquired C-band SAR data since 1995 and has the possibility of using a variety of incidence angles (between 20^{o} and 49^{o}) and different resolutions (between 10 and 100 m). The Advanced SAR (ASAR) onboard ENVISAT, launched in 2002, also has the possibility of using several incidence angles (between 15^{o} and 45^{o}). Besides, the C-band SAR can transmit and receive radar pulses both in horizontal and vertical polarization, which refers to the plane in which the electromagnetic wave is propagating. Spatial resolutions of ASAR are approximately 30 m, 150 m, or 1 km, depending on the mode used. Landsat satellites had the enhanced TM (ETM) sensors onboard with improved resolution and more spectral bands. ASTER (Advanced Space borne Thermal Emission and Reflection Radiometer) is one of the sensors onboard the Terra satellite. It has 14 spectral bands of which several are situated in the SWIR and TIR regions. One near infrared (NIR) band looks both nadir and backward creating stereo-view from a single pass. MODIS (Moderate Resolution Imaging Spectroradiometer) is one of the sensor onboard the Terra (EOS AM) and Aqua (EOS PM) satellites. It has five bands near infrared. It has 29 bands of which several are situated in the SWIR/MWIR and LWIR regions.

3.0 OVERVIEW OF EXISTING METHODS

3.1 Feature extraction method

1) The Entropy Based Water Body extraction method has been tested on ERS SAR amplitude data, SPOT HRV, and LANDSAT-7 ETM+ panchromatic images. Figure 1, shows the water areas extraction in Hubei Province of China about LANDSAT-7 ETM+ image and Figure 2 shows an example of extraction process from ERS SAR amplitude image on Poyang Lake of China. Because of the speckle effect in SAR images, the method works better for optical images than that for SAR images. The details of the water area border smoothed slightly in the course of post-processing.

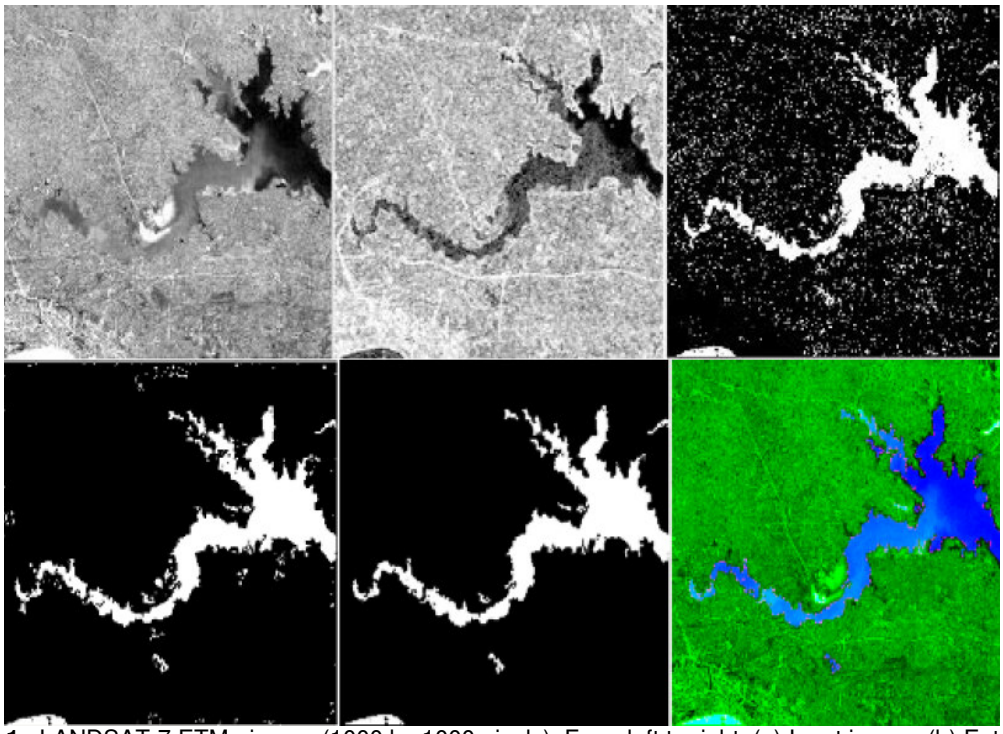


Figure 1: LANDSAT-7 ETM+ image (1000 by 1000 pixels). From left to right: (a) Input image, (b) Entropy data from step 2, (c) Segmented result From step 3, d) Post-processing in step 4, (e) Extracted Water body, and (f) Overlaid with the input image.

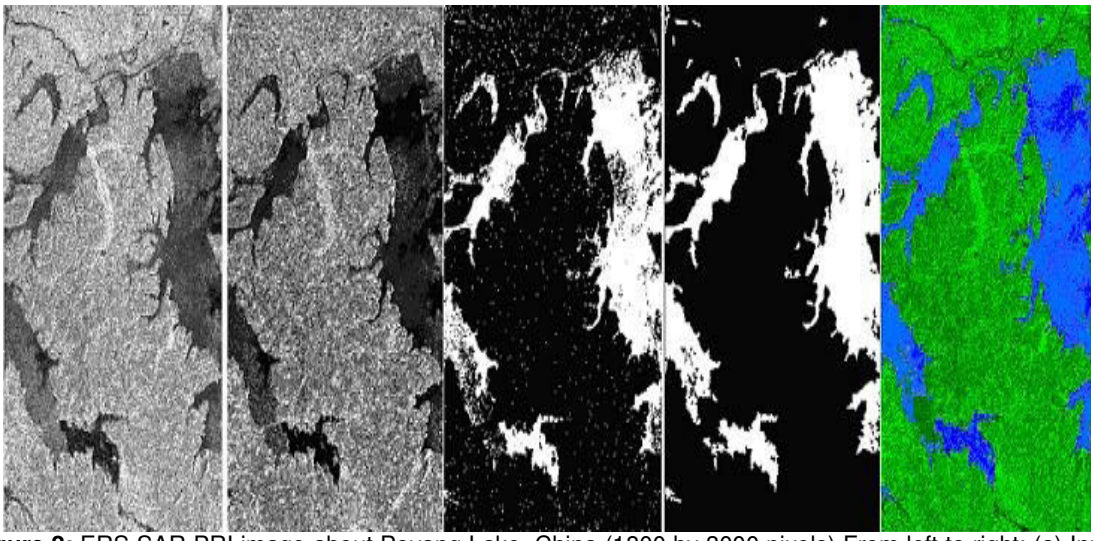


Figure 2: ERS SAR PRI image about Poyang Lake, China (1800 by 3000 pixels) From left to right: (a) Input Image, (b) Entropy Image, (c) Segmented results, (d) After post-processing, and (e) Overlaid with the input image

2) In general, images have the following features – color, texture, shape, edge, shadows, temporal details etc. The most promising features were color, texture, and edge. These features are extracted individually from the satellite and combined to get the final extracted image.

3) The mean shift algorithm is a powerful technique for image segmentation. The algorithm recursively moves to the kernel smoothed centroid for every data point. The quadratic computational complexity of the algorithm is a significant barrier to the scalability of this algorithm

to practical applications. The fast Gauss transform (FGT) has successfully accelerated the kernel density estimation to linear running time for low-dimensional problems. Unfortunately, the cost of a direct extension of the FGT to higher-dimensional problems grows exponentially with dimension, making it impractical for dimensions above three [36], [37]. An image segmented into homogeneous regions by mean shift segmentation. Then, the major water body, identified and an initial shoreline generated. The final shoreline obtained by local refinement within the boundaries of the candidate regions adjacent to the initial shoreline.

4) Skeletonization is the process of peeling off of a pattern as many pixels as possible without affecting the general shape of the pattern [38], [39]. In other words, after pixels have been peeled off, the pattern should still be recognized. The skeleton hence obtained must have the following properties: 1) as thin as possible; 2) connected; and 3) centered. The water-body feature extracted from satellite imagery with a combination of two processes. This process includes the boundary extraction and skeletonization from color imagery using a color image segmentation algorithm, a crust extraction algorithm, and new skeleton extraction algorithm.

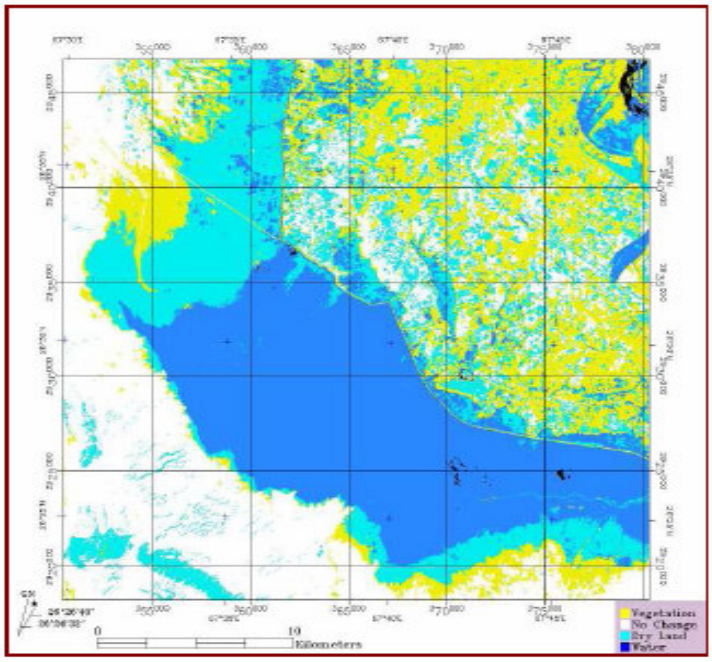


Figure 3: Result of Change detection 1992/2003

3.2 Supervised and Unsupervised Classification method

Advances in sensor technology for Earth observation make it possible to collect multispectral data in much higher dimensionality. In addition, multisource data also will provide high dimensional data. Such high dimensional data will have several impacts on processing technology: (1) it will be possible to classify more classes; (2) more processing power will be needed to process such high dimensional data (3) with large increases in dimensionality and the number of classes, processing time will increase significantly.

The analysis of remotely sensed data is usually done by machine oriented pattern recognition techniques. One of the most widely used pattern recognition techniques is classification based on maximum likelihood (ML) assuming Gaussian distributions of classes. A problem of Gaussian ML classification takes long processing time. The long processing time leads to long computational time and as a result computational cost rises. This computational cost may become an important problem if the remotely sensed data of a large area is to be analyzed or if the processing hardware is more modest in its capabilities. The advent of the future sensors will aggravate this

problem. Hence, attention should be paid to extract detailed information from high dimensional data while reducing processing time considerably [40].

There are various types of supervised classification method are used to classify the water body from high-resolution satellite images.

1) The supervised classification is used to classify the satellite image of years (1992 and 2003) in three different classes namely blue color for water, green for the vegetation and aqua for dry land and these results are compared to find out the change in the Mancher Lake of Pakistan . Report shows that number of points (Npts) selected for the sample region on the image and percentage (Pct) show the area of water, vegetation, and dry land that is shown in table 2 and figure 4.

For 1992

Class Name	Npts	Pct (%)
Unclassified	[0]	0.00
Vegetation	[242754]	23.119
Water	[530251]	35.049
Dry Land	[665373]	43.980

For 2003

Class Name	Npts	Pct (%)
Unclassified	[0]	0.00
Vegetation	[317276]	20.971
Water	[202640]	19.299
Dry Land	[604606]	57.582

Table 2: The percentage of water and other classified data

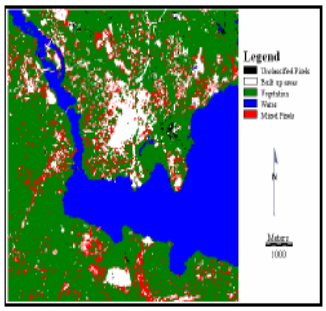


Figure 4a 1A1 Linear

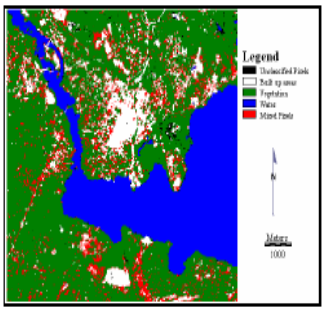


Figure 4b 1AA Linear

2) Support Vector Machine with One-Against-One (1A1) and One-Against-All (1AA) techniques is used for land cover mapping of the Landsat Scene located at the source of River Nile in Jinja, Uganda. The bands used in this research consisted of Landsat's optical bands i.e. bands 1, 2, 3, 4, 5, and 7. The classes of interest were built up area, vegetation, and water. Table 3 gives a summary of the unclassified and mixed pixels resulting from 1A1 and 1AA classification. From Table 3 it is evident that the 1AA approach to multiclass classification has exhibited a higher propensity for unclassified and mixed pixels than the 1A1 approach. From Table 4, all accuracies would be classified as yielding very strong correlation with ground truth data. The individual performance of the SVM classifiers however show that classification accuracy reduced for the linear and RBF classifiers stayed the same for the polynomial and increased for the quadratic classifier.

Classifier	Type	1A1	1AA
Linear	Unclassified Pixels	16	700
	Mixed Pixels	0	9048
Quadratic	Unclassified Pixels	142	5952
	Mixed Pixels	0	537
Polynomial	Unclassified Pixels	69	336
	Mixed Pixels	0	2172
RBF	Unclassified Pixels	103	4645
	Mixed Pixels	0	0

TABLE 3: Summary of number of unclassified and mixed pixels

Further, analysis of these results shows that these differences are pretty much insignificant at the 95% confidence interval. It can therefore be concluded that whereas one can be certain of high classification results with the 1A1 approach, the 1AA yields approximately as good classification accuracies.

SVM	1A1	1AA	|Z|	Significance
Linear	1.00	0.95	0.06	Difference insignificant
Quadratic	0.88	0.94	-0.02	Difference insignificant
Polynomial	1.00	1.00	0.0	No difference
RBF	0.97	0.92	0.01	Difference insignificant

TABLE 4: Summary of number of unclassified and mixed pixels

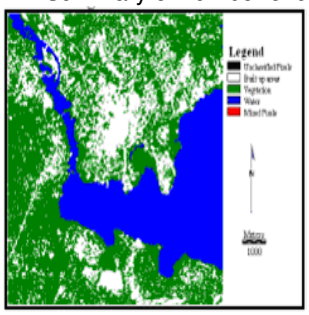


Figure 5a 1A1 Polynomial

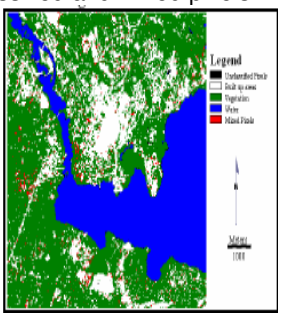


Figure 5b 1AA Polynomial

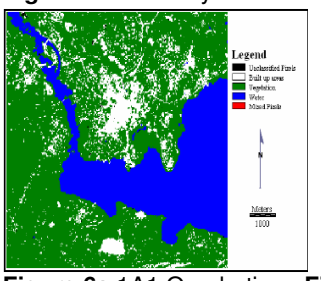


Figure 6a 1A1 Quadratic

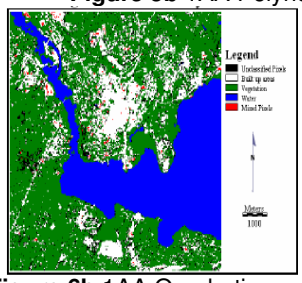


Figure 6b 1AA Quadratic

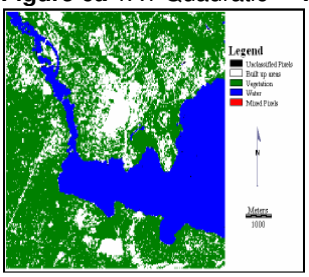


Figure 7a 1A1 RBF

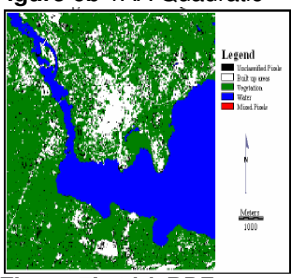


Figure 7b 1AA RBF

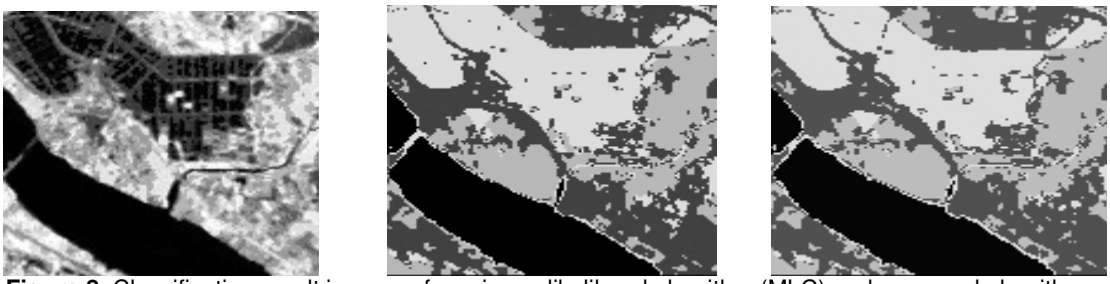
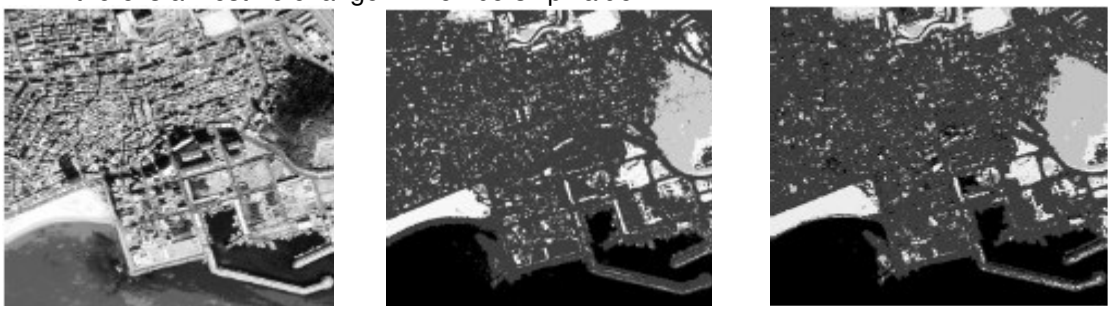


Figure 8: Classification result images of maximum likelihood algorithm (MLC) and proposed algorithm using Landsat TM satellite image

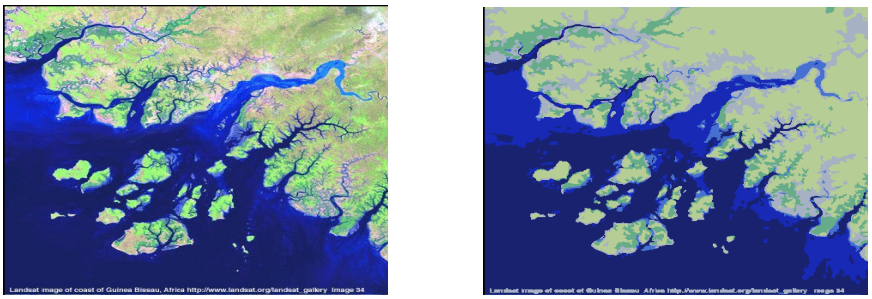
The author proposed a Learning Vector Quantization (LVQ) neural network method for automatic extraction of water bodies from Landsat 4 satellite image. In this work, Landsat Thematic Mapper(TM) sensor image of Mississippi river region of 1986 was used. It is a supervised classification method and aims to define the decision surface between competing classes. They compared their results with Tasseled Cap Transformation (TCT) and conventional rule based method. It observed that the result obtained by LVQ method is poor than rule based and TCC methods but the later two methods need human guidance while LVQ method is automatic [41].

3) The Bayesian supervised algorithm using the average intracluster distance within the fuzzy Gustafson-Kessel (GK) and Bayesian algorithm. The suggested algorithm uses the fuzzy GK algorithm in the form extended for the FCM. Different cluster distributions and sizes usually lead to sub optimal results with FCM. In order to adapt to different structures in data, GK algorithm used the covariance matrix to capture ellipsoidal properties of cluster. It makes classification of the remote sensing satellite image with multidimensional data possible. Fuzzy algorithm generally iterates the execution until there is almost no change in membership value.

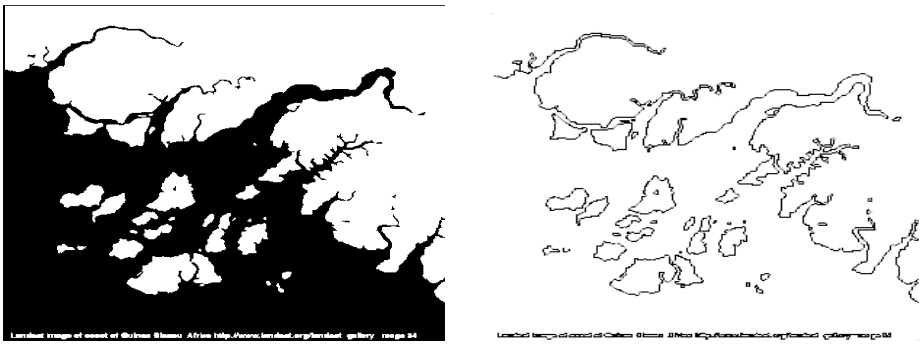


a) Original Image (b) MLC Image (c) Proposed Algorithm Image

Figure 9: Classification result images of maximum likelihood algorithm (MLC) and proposed algorithm using IKONOS satellite image.



(a) Satellite image of Guinea Bissau (b) Over-segmented image



(c) North Atlantic ocean (d) Extracted coastline

Figure 10: Feature boundary extraction from the satellite image of Guinea Bissau

Category		Forest	Water	Crop	Urban	Overall accuracy
Number of training pixels		1024	1024	1024	1024	4096
Classification methods	Maximum likelihood	90.43%	98.05%	92.48%	89.06%	92.50%
	FCM	85.15%	97.46%	85.83%	92.45%	90.22%
	Proposed method	91.25%	98.53%	93.45%	92.45%	93.92%

TABLE 5: The classification results by proposed algorithm, conventional maximum likelihood and FCM algorithm from Landsat TM resolution satellite image

Category		Forest	Water	Soil	Urban	Overall accuracy
Number of training pixels		9249	11323	13785	19426	53783
Classification method	Maximum likelihood	9110	11285	12753	18080	95.94%
	FCM	8820	10951	12896	17094	93.42%
	Proposed method	9235	11291	13127	18586	97.62%

TABLE 6: The classification results by proposed algorithm, conventional maximum likelihood and FCM algorithm from IKONOS high resolution satellite image

4) The supervised classification technique using Gabor Filter for the textural attribute to the high resolution satellite image. The author proposed a wavelet transform and Gabor filter based texture analysis for the recognition of water bodies from satellite images including other object on Earth surface [42]. The authors proposed two approaches namely pixel by pixel classification technique approach and object oriented image analysis for classification of water bodies and other land cover in a satellite image [43]. The authors proposed a mathematical morphological analysis approach for detecting water bodies form satellite image. They also suggest chromaticity analysis for removal of atmospheric differences between images [44].

5) The spectral, spatial, and textural features for each region are generated from the thresholded image by dynamic thresholding. Then given these features as attributes, an unsupervised machine learning methodology called conceptual clustering(COBWEB/3) is used to cluster the regions found in the image into N classes—thus, determining the number of classes in the image automatically. This technique is applied successfully to ERS-1 synthetic aperture radar (SAR), Landsat Thematic Mapper (TM), and NOAA advanced very high-resolution radiometer (AVHRR) data of natural scenes. Fig. 11 shows an original SAR sea ice image that consists of packed ice with very dark, cutting linear structures (leads) and grayish regions (new ice or open water). Moreover, there are brighter, silky structures (possibly deformed first year ice) straining within the grayish regions. Therefore, there are essentially four classes in the image.

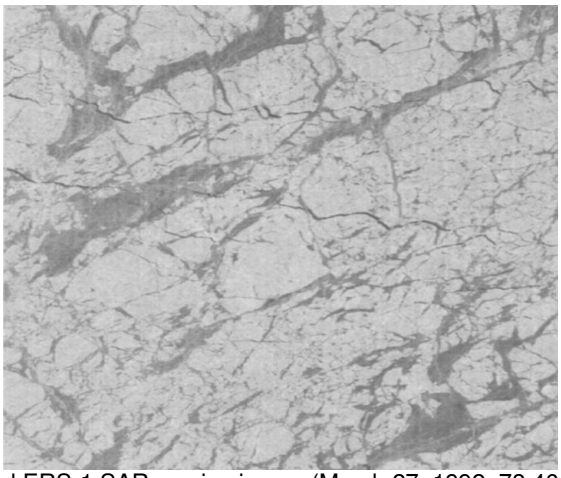


Figure 11: Original ERS-1 SAR sea ice image (March 27, 1992, 73.46 N, 156.19E).ESA

Fig. 12(a) shows the Yellow River plain, Shandong Peninsula, and the delta of Yangtze River at the south in China. It is the infrared band (0.725–1.10 m) of AVHRR, with a resolution of 1500 m/pixel. The image was a composite of a ten-day series, taken during September 1–10, 1992. In the image, the dark regions are bodies of water (sea, rivers, and lakes). To the west of the region lies the mountain range of Taihang. To the south of the region lies the mountain range of Dabie. Fig. 12(b) shows the segmentation results. The class labels are as follows:
1) black—water;
2) bright green—saline meadow;
3) orange—temperate coniferous forest and grassland;
4) dark green—warm temperate crops (rice) and deciduous coniferous forest;
5) yellow—scrub (mountains);
6) red—possibly broad-leaved deciduous forest.

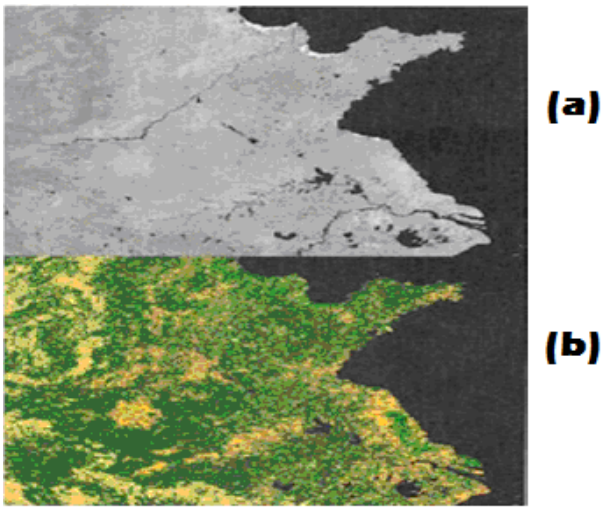


Figure 12: (a) Original AVHRR image. (b) The result of our segmentation: six classes

6) The unsupervised classification in 15–30 classes was used for distinguishing between land and water (Fig 13). The 381 AVHRR scenes selected from the cloud algorithm were classified, using all channels. Extending back in time, Landsat data prior to 1985 and for the year 1986 was used for estimating flooding independently for 8 months (November 1972, May–June 1979, May–August 1984, November 1986). The Okavango Delta covers 4 Landsat scenes, and for each of these dates at least 3 Landsat scenes were available. If needed, the data gap (i.e. the 4th quadrant) was filled by dates with similar flooding patterns to the other.

The size discrepancy in total flooded area between the AVHRR estimated floods against ATSR/Landsat estimated floods (columns 2 and 4 in Table 2) varies between 6 and 1351 km^2, averaging at 509 km^2, or 11%. The spatial discrepancy is given as the percentage of the AVHRR derived flooding falling inside the ATSR/Landsat derived flood (column 5 in Table7). This spatial accuracy varies between 63% and 89% (79% – 89% for full scenes).

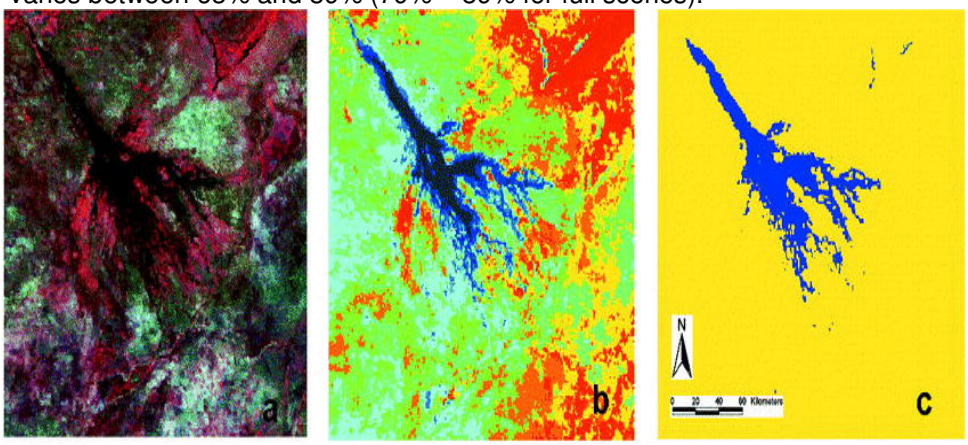


Figure 13: Classification steps a) original AVHRR scene (rgb 1, 2, 3) (date 25 August 1998); b) unsupervised classification in 10 classes; and c) water – land classification

Date	AVHRR(km^2)	Date	Reference Landsat (km^2)	AVHRR correct (%)
5 July 1994	7387	7 Jul/1 Aug 1994	7126	86
4 Dec 1994	(4891)	7 Dec/14 Dec 1994	(4926)	78(only part of image)
15 Feb 1995	(2539)	16 Feb 1995	(3785)	81(only part of image)
7 Oct 1999	6332	10 Oct/2 Nov 1999	6326	84
7 Apr 2000	7936	3 Apr/10 Apr 2000	7958	85
8 Sept 2000	8518	1 Sept/10 Sept/2 Nov 2000	8192	87
			Reference ASTR(km^2)	
25 Aug 1999	7226	30 Aug 1999(8 Sept 1999)	6902	87
3 Sept 1999	7042	2 Sept 1999(8 Sept 1999)	6992	89
19 Sept 1999	6562	18 Sept /21 Sept 1999	5675	79
28 Sept 1999	(6532)	24 Sept 1999	(6304)	85(only part of image)
7 Oct 1999	(5804)	4 Oct 1999	(4706)	73(only part of image)
16 Dec 1999	(5377)	16 Dec 1999	(4026)	63(partly cloudy)

TABLE 7: Accuracy evaluation result derived from cross tabulation of classification and reference data

3.3 Feature Based Classifier

The water-type classification process by applying statistical decision criteria to define class boundaries and assign pixels to a particular class. We have implemented two different feature-based classifiers, the Euclidean Distance Classifier, and the Eigenvector Classifier. The Euclidean Distance Classifier assigns each pixel $\mathbf{p}_j$ to a water type *i* based on the distance between that pixel and the centroid or mean of each class.

3.4 Data Fusion

1) After the feature extraction two change detection methods are applied: a) Image to Image and b) feature based. In image-to-image approach, the multi-temporal images this can be distinguish

between two approaches. An indirect image change detection, where the change analysis follows an image classification process. The comparison can be done by either differencing the two raster classified thematic layers or by extracting the boundaries of the thematic regions and conduct a vector (i.e., feature-based) change analysis. With this approach we overcome problems related to image acquisition conditions, such as different sensors, atmospheric and illumination conditions and viewing geometries. The accuracy of the detected changes is proportional to the accuracy of the image ortho rectification and of the classification results. In the second approach, image rationing, image differencing, image regression and Principal Component Analysis were used. While the feature based approach, the feature-based approach various functions of spatial analysis are used, such as layer union, layer intersection, buffer generation, and topological overlay.

Figure 14: Thresholding on Landsat 7 band 5

Figure 15: Extracted water bodies

2) A variety of satellite images of the 1993 flooding in the St. Louis area were evaluated and combined into timely data sets. The resulting maps were valuable for a variety of users to quickly locate both natural and man-made features, accurately and quantitatively determine the extent of the flooding, characterize flood effects and flood dynamics, and easily convey the results to a wide audience. Furthermore, the maps can continue to be used to help track changes over time, characterize the nature of the flooding, identify failures/weak points in the flood control systems, provide input into future flood plain analysis planning, and communicate details about the flooding clean-up work to both the general public and government planners.

(a)

(b)

(c) (d)

Figure 16: (a) Reference of normal Mississipsi, Illinois and Missouri River Channel; (b) July 29, 1993 SPOT Image of Flooded areas; (c) July 14, 1993 Satellite Radar Image Superimposed on Reference Map; (d) July 18, 1993 Band 4 Landsat TM Image of Flooded River System using Landsat TM satellite image.

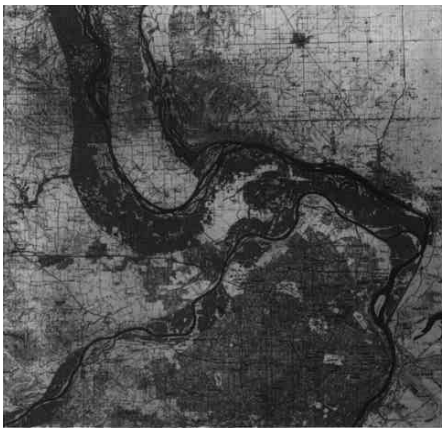


Figure 17: Image Showing Combined Data Sets

The authors proposed an algorithm DRAGON (Drainage Algorithm for Geospatial knowledge) which is a fusion method which is based on image processing and hydrologic modelling. The hydrologic modeling methodology based on modeling stream locations from DTED (Digital Terrain Elevation Data). Satellite imagery provides direct evidence of stream and lake locations, and used to complement and/or supersede stream locations derived from the DTED [45].

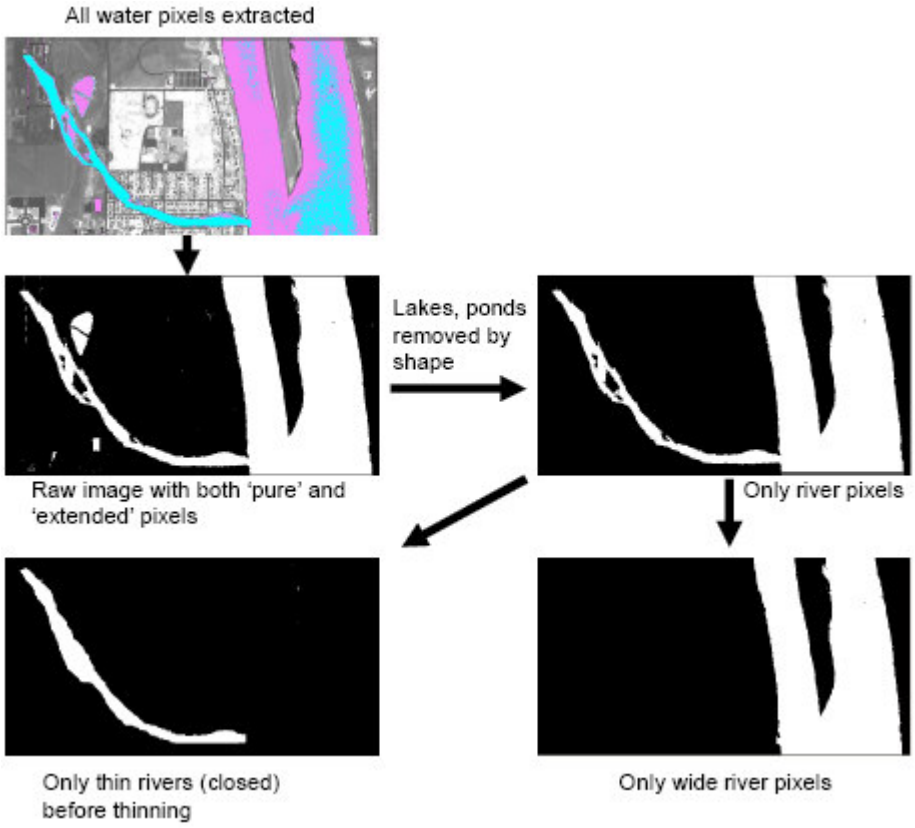


Figure18: Example of the piecewise refinement used by the DRAGON methodology to extract narrow (< 30m wide) and wide rivers

4. CHALLENGES, CONCLUSIONS, AND THE FUTURE

The distinction of colors between the shadows of tall buildings and calm water surface is still a challenge to the professionals. Therefore, it is difficult to get the exact information about water body in urban areas. To get the exact water in urban areas other similarity checks are required to be performed. Several algorithms were developed for extracting water body but none of them are accepted universally. Hence those are not applicable to various sensor images. Most of them are application specific.

In future the improvement in the water body extraction algorithm is expected, so that the system will be automated for handling all types of sensor images and it will be combined with other tools to provide better information for flood, availability of underground water. These aspects are critical issue in developing countries. Sometimes, it is tedious to collect the ground data manually.

Conclusions: The first part of this paper introduced the importance of water body information, the motivations of performing water feature extraction and the major difficulties in water body segmentation. The paper describes the different types of satellites and sensors used in acquiring satellite images for extracting water feature. Some of the results are discussed. Finally, an attempt has been made to conclude the current challenges as well as the future on water body extraction techniques.

5. REFERENCES

[1] Noble I. M. Apps, R. Houghton, D. Lashoff, W. Makundi, D. Murdiyarso, B. Murray, W. Sombroek, R. Valentini, R. Lal et al. 2000. Implications of different definitions and generic issues. In: "Land Use, Land Use Change and Forestry". IPCC Special Report, Washington, D.C., 377 pp.

[2] F. M. Henderson. "Environmental factors and the detection of open surface water areas with X-band radar imagery". International Journal of Remote Sensing, vol. 16(13), pp. 2423 – 2437, September 1995.

[3] Xie Chunxi , Zhang Jixian , Huang Guoman , Zhao Zheng and Wang Jiaoa. "Water body information extraction from high resolution airborne synthetic aperture radar image with technique of imaging in different directions and object-oriented". In Proceeding of the ISPRS Congress Silk Road for Information from Imagery, Beijing, 2008, pp. 165-168.

[4] A. Prakash and R. P. Gupta. "Land-use mapping and change detection in a coal mining area - a case study in the Jharia coalfield, India". International Journal of Remote Sensing, vol. 19(3), pp. 391 - 410, Feb 1998.

[5] Zhang Qiuwen, Wang Cheng, Shinohara Fumio, Yamaoka Tatsuo, "Automatic extraction of water body based on EOS/MODIS remotely sensed imagery". In Proceedings of the SPIE, Volume 6786, pp. 678642, 2007.

[6] HU Zhuowei , Gong Huili and Zhu Liying. "Fast flooding information extraction in emergency response of flood disaster". ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs, Urumchi, Xingjiang, China, August 28-29, 2007.

[7] Célia Gouveia and Carlos DaCamara. "Continuous mapping of the Alqueava region of Portugal using satellite imagery". In Proceeding of the EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 12 - 16 June 2006.

[8] Costas Armenik and Florin Savopol. "Image processing and GIS tools for feature and change extraction". In Proceeding of the ISPRS Congress Geo-Imagery Bridging Continents, Istanbul, Turkey, July 12-13, 2004, pp. 611-616.

[9] Cunjian Yang Cunjian Yang Rong He Siyuan Wang. "Extracting water-body from Beijing-1 micro-satellite image based on knowledge discovery". In the Proceeding of the IEEE International Geoscience & Remote Sensing Symposium, Boston, Massachusetts, U.S.A, July 6-11, 2008.

[10] A Gandhe, V Venkateswarlu and R N Gupta. "Extraction of coal under a surface water body – a Strata Control Investigation". Journal of Rock Mech. Rock Engg vol. 38 (5), pp. 399–410, 2005.

[11] Patricia G. Foschi, Deepak Kolippakkam, Huan Liu and Amit Mandvikar. "Feature extraction for image mining". In Proceeding of the Multimedia Information Systems Conf., pp. 103-109, 2002.

[12] Kaichang Di, Ruijin Ma, Jue Wang, Ron Li., "Coastal mapping and change detection using high-resolution IKONOS satellite imagery**". In** Proceedings of the 2003 annual national conference on Digital government research**,** Boston, MA, vol.130, pp 1 – 4,2003.

[13] Ojaswa Sharma, Darka Mioc and François Anton. “Feature extraction and simplification from color images based on color image segmentation and skeletonization using the quad-edge data structure”. In Proceeding of the 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2007.

[14] Kaichang Di, Ruijin Ma, Jue Wang, Ron Li. “Automatic shoreline extraction from high-resolution IKONOS satellite imagery”. In Proceedings of the 2003 annual national conference on Digital government research, Boston, MA, vol.130, pp 1 – 4,2003.

[15] Yuanzhi Zhang, Jouni T. Pulliainen, Sampsa S. Koponen, and Martti T. Hallikainen. “Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland”. IEEE Transaction on GEOSCIENCE AND REMOTE SENSING, 0196-2892, 2003.

[16] Leen-Kiat Soh and Costas Tsatsoulis. “Segmentation of satellite imagery of natural scenes using data mining”. IEEE Transactin On GeoScience and Remote Sensing, vol. 37(2), pp. 1086-1099, 2005.

[17] Zhang Zhaohui, Veronique Prinet and MA Songde. “Water body extraction from multi-source satellite images”. IEEE, 0-7803-7929-2/03, 2003.

[18] Jenny M. McCarthy, Thomas Gumbricht, Terence McCarthy, Philip Frost, Konrad Wessels, and Frank Seidel. “Flooding patterns of the Okavango wetland in Botswana between 1972 and 2000”. AMBIO: A Journal of the Human Environment, vol. 32(7), pp. 453-457, November, 2003.

[19] Linda V. Martin Traykovski and Heidi M. Sosik. “Optical classification of Northwest Atlantic water types based on satellite ocean color data”. Biology Department, MS 32, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

[20] Gidudu Anthony, Hulley Greg and Marwala Tshilidzi. “Classification of images using support vector machines”. arXiv: 0709.3967v1, Cornell University, Library, 2007.

[21] Habibullah U Abbasi, Mushtaq A Baluch and Abdul S Soomro. “Impact assessment on Mancher lake of water scarcity through remote sensing based study”. In Proceeding of GIS, Saudi Arabia.

[22] Ana Carolina Nicolosi da Rocha Gracioso, Fábio Fernando da Silva, Ana Cláudia Paris and Renata de Freitas Góes. “Gabor filter applied in supervised classification of remote sensing images”. In Symposium Proceeding of the SIBGRAPI 2005.

[23] Young-Joon Jeon, Jae-Gark Choi, and Jin-Il Kim. “A study on supervised classification of remote sensing satellite image by bayesian algorithm using average fuzzy intracluster distance”. R. K lette and J. Žunić (Eds.): IWCIA 2004, LNCS 3322, pp. 597–606, 2004.

[24] Alecu Corina, Oancea Simona, and Bryant Emily. “Multi-resolution analysis of MODIS and ASTER satellite data for water classification”. In Proceedings of the SPIE, the International Society for Optical Engineering, San Jose CA, ETATS-UNIS 2006.

[25] L.M. Fuller, T.R. Morgan, and S.S. Aichele. “Wetland delineation with IKONOS high-resolution satellite imagery, Fort Custer Training Center, Battle Creek, Michigan, 2005”. Scientific Investigations Report 2006–5051.

[26] Jean-Francois Cayula and Peter Cornillon. “Edge detection algorithm for SST algorithm”. Journal of Atmospheric and Oceanic Technology, vol. 9, pp 67-80, 1992.

[27] G M Petrie, G.E. Wukelic, C.S. Kimball,K.L. Steinmau and, D.E. Beaver. “Responsiveness of satellite remote sensing and image processing technologies for monitoring and evaluating 1993 Mississippi River flood development using ERS-1 SAR, LANDAST, and SPOT digital data”. In Proceeding of the ASPRS/ACSM, Reno, NV, 1994.

[28] Nandish M. Mattikalli and Keith S. Richards. “Estimation of surface water quality changes in response to land use change: application of the export coefficient model using remote sensing and geographical information system”. Journal of Environmental Management vol. 48, pp. 263–282, 1996.

[29] O S Mudenda and E Nkonde. “The man-made satellite; an instrument of opportunity”. In CD-ROM Proceeding of the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005) Bucharest, Romania, 4-7 May.

[30] Patrick Brezonik, Kevin D. Menken and Marvin Bauer. “Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM)”. Journal of Lake and Reservoir Management 21(4), pp. 373-382, 2005.

[31] Leif G. Olmanson, Steve M. Kloiber Patrick L. Brezonik and Marvin E. Bauer. “Use of satellite imagery for water clarity assessment of Minnesota’s 10,000 Lakes”. University of Minnesota.

[32] A. G. Dekker, T. J. Malthus, M. M. Wijnen and E. Seyhan. “Remote sensing as a tool for assessing water quality in Loosdrecht lakes”. Journal of Hydrobiologia, vol. 233, pp. 137-159, 1992.

[33] Panu Nuangjumnong , Ramphing Simking. “Automatic Extraction of Road and Water Surface from SPOT-5 Pan-Sharpened Image”. In Proceeding of the Conference Map Asia, 2009.

[34] Hafeez MM, Chemin Y, Van De Giesen, and Bouman B A M. “Field Evaporation in Central Luzon, Philippines, using different sensors: Landsat 7 ETM+, Terra Modis and Aster”. In Proceeding of the Symposium on Geospatial Theory, Processing and Applications, Ottawa, 2002.

[35] Magsud Mehdiyev, Ryuzo Yokoyama and Lal Samarakoon. “DETECTION OF WATER-COVERED AREAS BY USING MODIS IMAGERY”. In Proceeding of the GeoInfo Conference, ACRS2004, Chiang Mai, Thailand, October 25-30, 2004.

[36] C. Yang, R. Duraiswami, N. Gumerov and L. Davis. Improved Fast Gauss Transform and Efficient Kernel Density Estimation. In Proceeding of the IEEE International Conference on Computer Vision, pages 464-471, 2003.

[37] C. Yang, R. Duraiswami, D. DeMenthon and L. Davis. Mean-Shift Analysis Using Quasi-Newton Methods. In Proceeding of the IEEE International Conference on Image Processing, pages 447 - 450, vol.3, 2003.

[38] Janke, R., R. Murray, J. Uber, R. Bahadur, T. Taxon and W. Samuels, 2007. “Using TEVA to assess impact of model skeletonization on contaminant consequence assessment and sensor placement design”. In Proceeding of the World Environmental and Water Resources Congress, Tampa, FL, May 15-19.

[39] H. Sundar, D. Silver, N. Gagvani, S. Dickinson. “Skeleton Based Shape Matching and Retrieval”. In Proceedings of the Shape Modeling International 2003, p.130, May 12-15, 2003.

[40] Chulhee Lee, David Landgrebe. “Feature Extraction And Classification Algorithms For High Dimensional Data”. School of Electrical Engineering Purdue University West Lafayette, Indiana 47907-1285, TR-EE 93-1, January 1993.

[41] Kefei Wand and Yifeng Zhu. “Recognition of Water Bodies from Remotely Sensed Imagery by Using Neural Network”. UNIVERSITY OF NEBRASKA - LINCOLN, CSE873 COMPUTER VISION.

[42] Mariana Tsaneva, Doyno Petkov. “RECOGNITION OF OBJECTS ON THE EARTH'S SURFACE THROUGH TEXTURE ANALYSIS OF SATELLITE IMAGES”. In Proceeding of the Third Scientific Conference with International Participation SPACE, ECOLOGY, NANOTECHNOLOGY, SAFETY 27–29 June 2007, Varna, Bulgaria.

[43] MARIE-CATHERINE MOUCHOT, THOMAS ALFOLDI, DANIEL DE LISLE and GREG McCULLOUGH. “Monitoring the Water Bodies of the Mackenzie Delta by Remote Sensing Methods”. ARCTIC, VOL. 44, SUPP. 1 (1991), PP. 21-28.

[44] T Van de, W De Genst, F Canters, N Stephens, E Wolf, and M Binard. “Extraction of Land Use/ Land Cover-Related Information from very High Resolution Data in Urban and Suburban areas”. In proceeding of the 23rd EARSeL Annual Symposium on June 3, 2003.

[45] Gregg Petrie, Brain Moon, and Karen Steinmaus. “Semi-automated stream extraction at PNNL”. In proceeding of the Overwatch Geospatial Users Conference, 2008.

Reversible Data Hiding in the Spatial and Frequency Domains

Ching-Yu Yang *chingyu@npu.edu.tw*
Dept. of Computer Science and Information Engineering
National Penghu University
Penghu, 880, Taiwan

Wu-Chih Hu *wchu@npu.edu.tw*
Dept. of Computer Science and Information Engineering
National Penghu University
Penghu, 880, Taiwan

Abstract

Combinational lossless data hiding in the spatial and frequency domains is proposed. In the spatial domain, a secret message is embedded in a host medium using the min-max algorithm to generate a stego-image. Subsequently, the stego-image is decomposed into the frequency domain via the integer wavelet transform (IWT). Then, a watermark is hidden in the low-high (LH) and high-low (HL) subbands of the IWT domain using the coefficient-bias approach. Simulations confirm that the hidden data is successfully extracted and the host image is completely recovered. In addition, the perceptual quality of the mixed image generated by the proposed method is good. Moreover, the mixed images are robust against attacks such as JPEG2000, JPEG, brightness adjustment, and inversion.

Keywords: Reversible data hiding, IWT, Min-max algorithm, Coefficient-bias approach.

1. INTRODUCTION

A stable and efficient data switching network makes it easy for individuals and organizations to exchange (or share) their resources on the Internet. Business-to-business (B2B), business-to-consumer (B2C), and customer-to-customer (C2C) commerce are three popular services provided over the Internet. However, data can be eavesdropped on, illicitly tampered, or falsified during transmission. Most commercial parties (or organizations) utilize encryption to protect important (or private) data during transactions. However, confidential data can become insecure if a private key is exposed or stolen by a third party. Data hiding techniques are an alternative solution to data protection. Generally speaking, data hiding can be classified into fragile watermarking and robust watermarking [1-2]. Fragile watermarking approaches [3-5] have the capability of hiding a large amount of data in a host medium while obtaining good resultant perceived quality. However, the marked images generated by these approaches are vulnerable to manipulations. Robust watermarking schemes [6-8] that can resist image processing attacks have been presented. However, most of the schemes allow a limited payload size.

Host media are important objects, such as law enforcement, military maps, and medical images, so they must not be damaged after digital watermarking. Several researchers presented lossless watermarking techniques [9-16]. Tian [9] implemented the difference expansion (DE) technique for lossless data hiding. To obtain extra storage space, Tian employed the DE technique to explore redundancy in the image content. Simulations showed that both the hiding capacity limit and the perceptual quality of the marked images were among the best at that time. Alattar [10] extended Tian's algorithm with DE of vectors, instead of pairs, to improve hiding efficiency. Using a generalized integer transform, Alattar presented a

reversible watermarking algorithm, which has a very high-bit hiding capacity, along with high peak signal-to-noise ratio (PSNR) performance. Ni et al. [11] utilized the ideal of the zero (or the minimum) points of the histogram to embed data bits into a host medium. Although the average PSNR was 48.20 dB, the payload size was insufficient. Based on the idea of three-pixel block differences, Lin and Hsueh [12] suggested a high performance reversible hiding algorithm. The average (pure) payload was 1.79 bit per pixel (bpp), but the resultant PSNR was 22.06 dB. Lin et al. [13] presented a multilevel reversible data hiding scheme based on difference image histogram modification. By employing the peak point of a difference image with a multilevel hiding policy, the scheme allows a large number of embedded bits while maintaining good resultant perceptual quality. Using a location map, auxiliary information, and a novel LSB substitution, Hsiao et al. [14] employed a block-based reversible data hiding method. The average PSNR generated by the method was about 30 dB with an embedding rate of 1.02 bpp. Tseng and Chang [15] proposed a reversible watermarking algorithm using the idea of shiftable pixel pairs. The extended difference expansion algorithm has a great hiding capacity without producing noticeable distortion. Tsai et al. [16] utilized predictive coding and histogram shifting to further improve the performance of Ni et al.'s method. The technique has good hiding capability and resulting perceived quality for stego-images produced from medical images.

The above lossless data hiding schemes [9-16], which are conducted in the spatial domain, provide a large number of hiding bits. In the present study, we develop a reversible data hiding method based on the spatial and frequency domains that has the capability of resisting manipulations. The rest of the paper is organized as follows. The proposed min-max algorithm and coefficient-bias algorithm are described in Section 2. Section 3 presents the simulations. The conclusion is given in Section 4.

2. PROPOSED METHOD

In the proposed method, a secret message is first embedded in the spatial domain using the min-max algorithm, and then a watermark is hidden in the integer wavelet transform (IWT) domain [3] using the coefficient-bias approach. More specifically, the watermark is embedded in the low-high (LH) and high-low (HL) subbands of the L1 IWT domain. A schematic overview of the proposed method is shown in Fig. 1. Note that 'Secret Message' and 'Test-logo' as shown in Fig. 1 denote two various attributes of input data. However, it can be replaced by a single input message or a piece of icon. Notice as well the IIWT appeared in Fig. 1 stands for inverse integer wavelet transform.

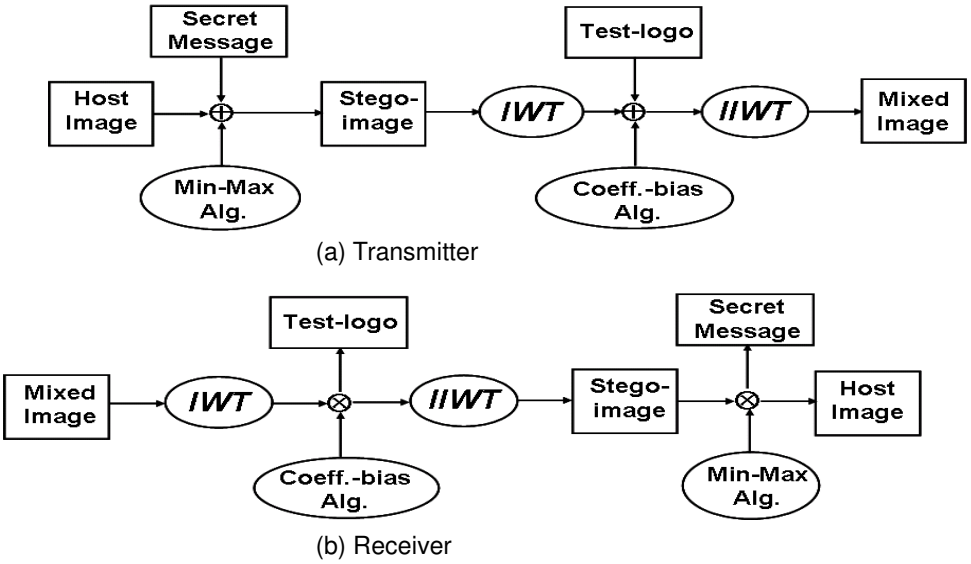


FIGURE 1: Block diagram of the proposed method. (a) Transmitter and (b) receiver.

2.1 Min-max algorithm

To provide extra storage space for hiding data bits, the proposed min-max algorithm was employed in the spatial domain. Without loss of generality, let $P = \{p_j\}_{j=0}^{(n\times n)-1}$ be an $n\times n$ nonoverlaping block divided from a host image. Let $p_{\min} = \arg Min\{p_j\}_{j=0}^{(n\times n)-1}$ and $p_{\max} = \arg Max\{p_j\}_{j=0}^{(n\times n)-1}$ be the minimum value and maximum value of the pixel in block *P*, respectively. Also let σ be a control parameter and *k* is a postive multiplier. The main steps of the min-max algorithm are as follows:

Step 1. Input a block *P* from a host image.

Step 2. Compute $p_{\min}$ and $p_{\max}$ of *P*.

Step 3. If $p_{\min} \geq 128$, then subtract $p_{\min}$ from p_j to obtain q_j; otherwise, subtract $p_{\max}$ from p_j to obtain q_j.

Step 4. If there exists a pixel $q_j \geq \sigma$, then add σ to q_j to obtain $\tilde{q}_j$. (The pixels are not qualified to carry bits.)

Step 5. If there exists $q_j < \sigma$, then multiply *k* by q_j to obtain $\hat{q}_j$. If an input bit is 1, add 1 to $\hat{q}_j$; otherwise, do nothing.

Step 6. If $p_{\min} \geq 128$, then add $p_{\min}$ to $\tilde{q}_j$ and $\hat{q}_j$, respectively; otherwise, subtract $p_{\max}$ from $\tilde{q}_j$ and $\hat{q}_j$, respectively. (The marked block contains the hidden bits.)

Step 7. Repeat Step 1 untill all data bits have been embedded in the block.

At the receiver, all of the data bits are sequentially extracted from the hidden block in a stego-image using a reverse procedure of the above algorithm. The host image can thus be completely recovered. Fig. 2 shows an example of bits being embedded using the min-max algorithm. In Fig. 2(a), we assume that the divided block has a size of 4×4 and that the input bit stream is "11101001011." *k* and σ are set at 2 and 5, respectively. Note that the minimum value, $p_{\min}$, and the maximum value, $p_{\max}$, of the block are 163 and 168, respectively. Step 3 of the algorithm produces a difference block, as shown in Fig. 2(b). To further alleivate distortion, the coefficients q_j which satisfy $q_j \geq \sigma$ are isolated from others in the block, as shown in Fig. 2(c), by adding σ *to* q_j via Step 4. The hidden block shown in Fig. 2(d) was obtained in Step 5. Finally, the marked block shown in Fig. 2(e) was generated in Step 6. Note that the mean square error (MSE) computed from the original block and the marked one is 8.69. An example of bit extraction is shown in Fig. 3. The figure shows a reverse procedure conducted on the marked block. The hidden bits are successfully extracted and the original block is completely recovered.

168	164	165	163
168	164	165	163
168	164	165	163
168	164	165	163

(a)

5	1	2	163
5	1	2	0
5	1	2	0
5	1	2	0

(b)

10	1	2	163
10	1	2	0
10	1	2	0
10	1	2	0

(c)

10	3	5	163
10	3	4	1
10	2	4	1
10	2	5	1

(d)

173	166	168	163
173	166	167	164
173	165	167	164
173	165	168	164

(e)

FIGURE 2: Example of bit embedding. (a) 4×4 block of the original block, (b) a difference block, (c) isolated-coefficients, (d) the hidden block, and (e) the marked block.

173	166	168	163
173	166	167	164
173	165	167	164
173	165	168	164

(a)

10	3	5	163
10	3	4	1
10	2	4	1
10	2	5	1

(b)

10	1	2	163
10	1	2	0
10	1	2	0
10	1	2	0

(c)

5	1	2	163
5	1	2	0
5	1	2	0
5	1	2	0

(d)

168	164	165	163
168	164	165	163
168	164	165	163
168	164	165	163

(e)

FIGURE 3: Example of bit extraction. (a) An input marked block, (b) coefficient subtraction, (c) bit extraction, (d) restored differecnce block, and (e) recovered original block.

2.2 Coefficient-bias approach

As described previously, the purpose of the coefficient-bias approach (with pixel adjustment) is to embed a watermark in the frequency domain. The details are given in the following three subsections.

2.2.1 Data embedding

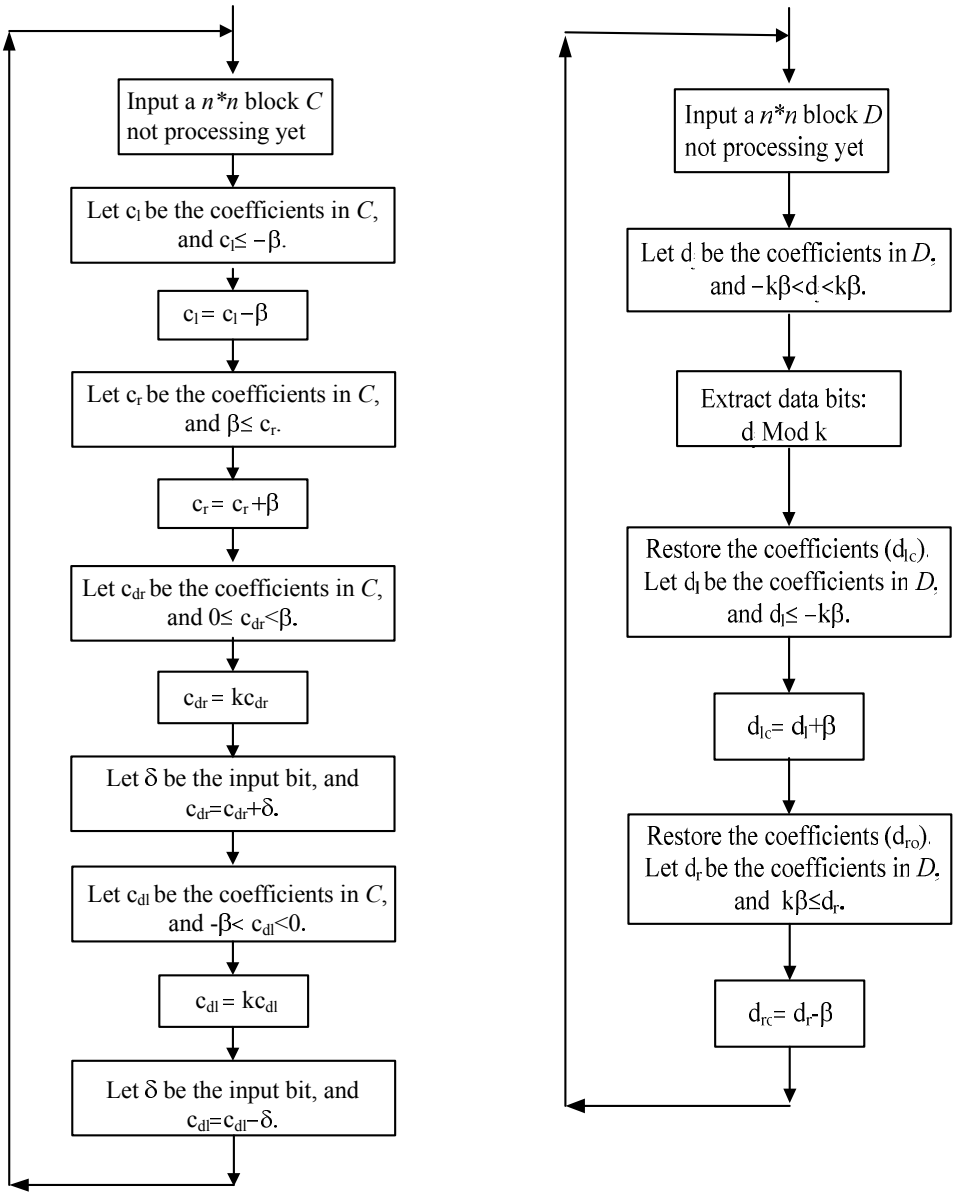
Decompose a stego-image into IWT domain. Input an *n*×*n* block $C = \{c_j\}_{j=0}^{(n\times n)-1}$ from the LH (or HL) of the IWT coefficient and δ be the input data. If there exists a coefficient $c_l \in C$ and $c_l \le -\beta,$ then subtract β from c_l. If there also exists a coefficient $c_r \in C$ and $\beta \le c_r,$ then add β to c_r. The payload provided by a host image is determined by the parameter β. Let $\hat{c}$ be the resultant coefficient of a IWT block. The above rules can be summarized as follows:

$$\hat{c} = \begin{cases} c_l - \beta, & \text{if } c_l \le -\beta; \\ \\ c_r + \beta, & \text{if } c_l \ge \beta. \end{cases} \tag{1}$$

After coefficient adjustments, data bits are ready to be embedded in blocks. Multiply coefficients $c_{dr} \in C$ which satisfy $0 \le c_{dr} < \beta$ by *k* to obtain $\hat{c}_{dr}$. *k* is an integer. Add δ to $\hat{c}_{dr}$. Then, multiply coefficients $c_{dl} \in C$ which satisfy $-\beta < c_{dl} < 0$ by *k* to obtain $\hat{c}_{dl}$. Subtract δ from $\hat{c}_{dl}$. Normally, to embed a data bit into each of the candidate coefficients, the value of *k* is set at 2. The procedure is repeated until all data bits have been processed.

2.2.2 Data extraction

At the receiver, a marked image is first decomposed into the IWT domain. Then, read in a block *D* of size *n*×*n* from the LH and HL subbands of IWT, respectively. If there exists a coefficient $d_j \in D,$ which satisfies $-k\beta < d_j < k\beta,$ divide d_j by *k*. The hidden bits can be obtained from the residual. Subsequently, restore the coefficients which were originally located between -β and β by dividing d_j by *k*. Then, restore the coefficients which were originally less than or equal to -β by adding $d_l,$ which satisfies $d_l \le -2\beta$, to β and restore the coefficients which were originally greater than or equal to β by subtracting $d_r,$ which satisfies $d_r \ge 2\beta$, from β. The procedure is repeated until all data bits are extracted. The coefficient-bias approach is summarized in Fig. 4.



(a) Encoding part (b) Decoding part

FIGURE 4: Flowchart of the proposed approach. (a) Encoder and (b) decoder.

2.2.3 Pixel adjustment

The aim of the pixel adjustment used here is to ensure lossless data hiding. To determine whether the goal of a successful recovery of a mixed image is achieved or not, a prior data extraction is performed before the mixed image transmitted to the receiver. More specifically, If a stego-image cannot be losslessly recovered from a mixed image, pixel adjustment is utilized. That is, if pixel p in a host medium satisfies either $p < \phi_1$ or $\phi_2 < p$, then the pixel is adjusted to a new value by adding or subtracting pixel-offset γ (the value of γ can be set to be the same as that of the parameter β). The new pixel $\hat{p}$ is obtained using:

$$\hat{p} = \begin{cases} p+\gamma & \text{if } p < \phi_1; \\ p-\gamma, & \text{if } \phi_2 < p. \end{cases} \quad (2)$$

The overhead information, which is used to record the position of each adjusted-pixel, can be

losslessly compressed [17] and out-of-band transmission to the receiver. The stego-images can be recovered completely at the receiver by a reverse pixel adjustment.

3. EXPERIMENTAL RESULTS

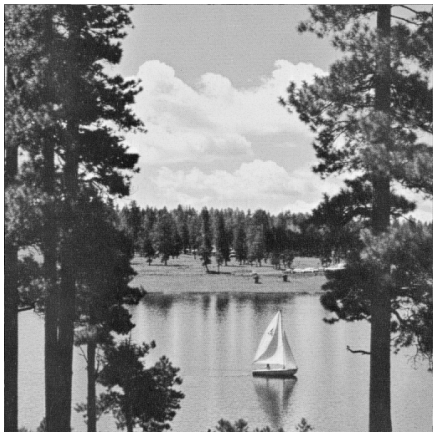
Several 512×512 gray-scale images were used as the host images. A quarter of the host image *Lena* was used as test data. The mixed images generated by embedding parts of the test data in the host images using the proposed method are shown in Fig. 5. The block size is 4×4. The control parameters σ and β were set to 3 and 8, respectively. The multiplier factor k used here is 2. Fig. 5 shows that the perceptual quality of the mixed images is good. Their hiding performance is listed in Table 1. Most of the images required no pixel adjustment during data embedding. 1-pixel and 10-pixel adjustments were required for the images Elaine and Sialboat, respectively. Note that the two sets of parameters (ϕ_1, ϕ_2) used in these two images were (1, 255) and (7, 255), respectively. The average PSNR is 34.59 dB with an embedding rate of 0.457 bpp. In addition, payload size generated by the proposed method in the spatial and frequency domain, respectively, is given in Table 2. It is obvious that payload size provided in IWT domain is about seven times larger than that provided in spatial domain. The trade-off between PSNR and hiding rate for the proposed method is shown in Fig. 6. To obtain higher PSNR performance with an embedding rate of less than 0.2 bpp, the value of σ should be set at 1, and that of β be set below 3. On the other hand, better bits-hiding capability is obtained when larger values of σ and β are used in the proposed method.

(a) Lena (b) Jet

(c) Peppers (d) Elaine

FIGURE 5: Mixed images generated by the proposed method. (a) Lena, (b) Jet, (c) Peppers, (d) Elaine, (e) Goldhill, and (f) Sailboat.

(e) Goldhill



(f) Sailboat

FIGURE 5: Continued.

Images	Embedding rate (bpp)	PSNR (dB)	No. of pixel-adj.
Lena	0.491	34.99	0
Jet	0.527	35.29	0
Peppers	0.487	35.00	0
Elaine	0.423	34.09	1
Goldhill	0.410	34.12	0
Sailboat	0.403	34.06	10

TABLE 1: Hiding performance for Fig. 5.

Images	IWT domain	Spatial domain	Total payload	PSNR
Lena	110,557	18,246	128,803	34.99
Jet	110,794	27,476	138,270	35.29
Peppers	112,724	14,866	127,590	35.00
Elaine	99,650	11,210	110.860	34.09
Goldhill	95,877	11,599	107,476	34.12
Sailboat	94,154	11,019	199,329	34.06

TABLE 2: Payload size generated by the proposed method in the spatial and IWT domain.

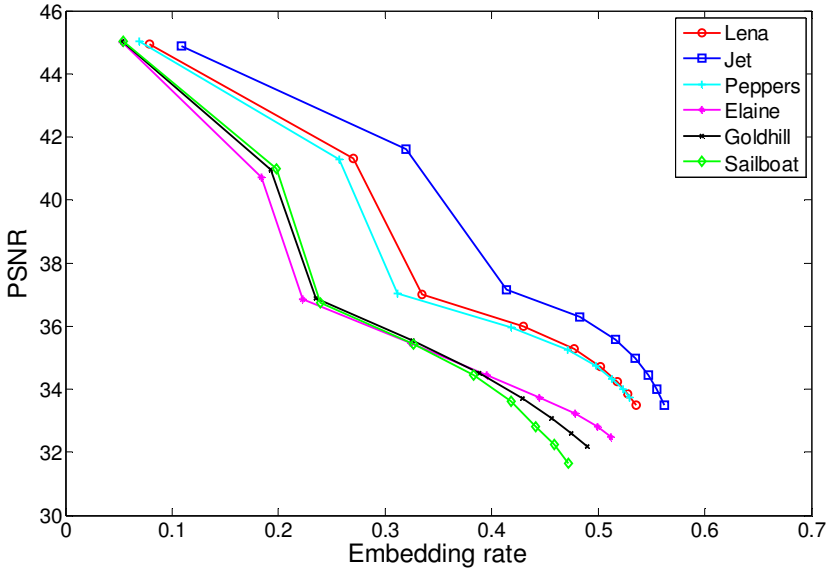


FIGURE 6: Trade-off between PSNR and hiding rate for the proposed method.

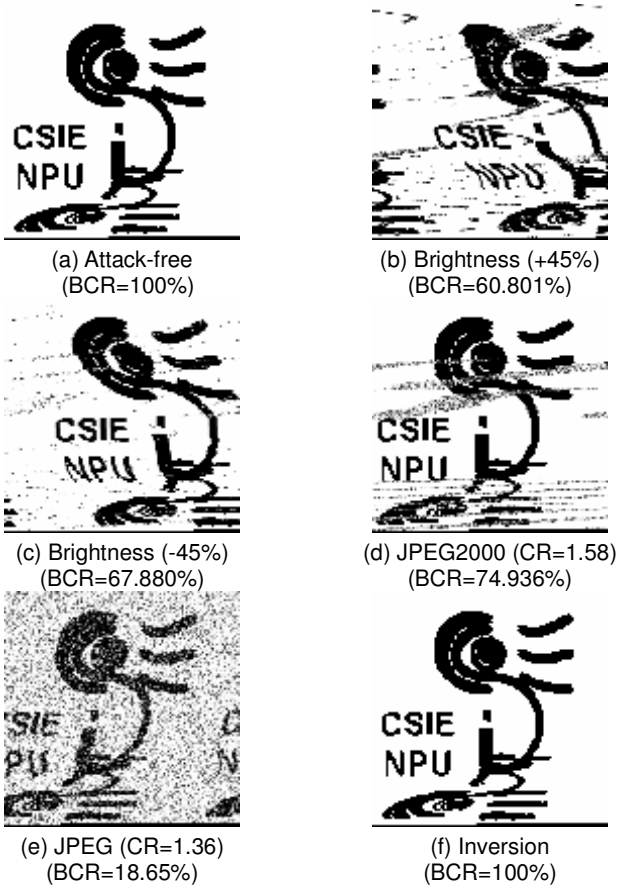
Performance comparison between our method and several lossless data hiding schemes [12-14] is listed in Table 3. It can be seen from Table 3 that the schemes (performed in the spatial domain) provide a large hiding capacity, but their average PSNR is about 30 dB. Since the perceived quality is not so good that it might be attarcted by the third parties. In other words, the resulting images generated by these scheme are vulnerable to attack. However, the resultant images generated by our method are more robust against attack than those generated by spatial-domain methods. Fig. 7 shows that the mixed images produced by the proposed method (using β=8, σ=3, and k=2 for image Lena) can resist attacks such as brightness adjustment (±45%), JPEG2000 coding with a compression ratio (CR) of 1.58, JPEG coding (with CR=1.36), and inversion. Although the bit correct ratio (BCR) for the watermarks in Fig. 7(b) and 7(c) are a bit low, the extracted watermarks are recognizable. Although the BCR of Fig. 7(e) is only 18.65%, the extracted watermark is still recognizable. It is interesting that the BCR of Fig. 7(f) is 100%, which means that the mixed images generated by our method are immune to an inversion attack. BCR is defined as:

$$BCR = \left(\sum_{i=0}^{MN-1} \overline{w_i \oplus w_i'} \Big/ M \times N \right) \times 100\% \qquad (3)$$

where w_i and w_i' represent the values of the original watermark and the extracted watermark, respectively. The watermark has a size of $M \times N$.

Images	Lin et al.'s tech. [12]	Lin et al.'s appr. [13]	Hsiao et al.'s alg. [14]	Proposed method
Lena	30.0/ 1.18	30.19/ 1.322	30.00/ 1.159	34.99/ 0.491
Jet	30.3/ 1.40	30.19/ 1.384	30.00/ 1.093	35.29/ 0.527
Peppers	30.2/ 1.36	30.19/ 1.305	30.00/ 1.159	35.00/ 0.487
Goldhill	30.1/ 1.16	-	30.00/ 0.936	34.12/ 0.410

TABLE 3: PSNR and embedding rate for the proposed method and other schemes.



(a) Attack-free (BCR=100%)
(b) Brightness (+45%) (BCR=60.801%)
(c) Brightness (-45%) (BCR=67.880%)
(d) JPEG2000 (CR=1.58) (BCR=74.936%)
(e) JPEG (CR=1.36) (BCR=18.65%)
(f) Inversion (BCR=100%)

FIGURE 7: Examples of extracted watermarks (size of 117×117 with 8 bits/pixel, 2 colors) after various attacks. (a) Attack-free, (b) Brightness (+45%), (c) Brightness (-45%), (d) JPEG2000, (e) JPEG, and (f) Inversion.

4. CONCLUSION

An effective lossless data hiding scheme that embeds data bits in the spatial and frequency domains was proposed. The proposed method consists of two approaches, namely, the min-max algorithm and coefficient-bias approach. The min-max algorithm is used to hide a secret message in a host media in the spatial domain. In the frequency domain, a watermark is embedded in the LH- and HL-subbands of IWT using the coefficient-bias approach. Experiments indicate that not only a hidden data is successfully extracted but also a host image is losslessly restored. Moreover, the resultant perceptual quality generated by the proposed method is good. The mixed images can survive various manipulations, such as JPEG2000 and JPEG brightness adjustment, and inversion.

5. REFERENCES

1. F. Y. Shih. "*Digital watermarking and steganography: fundamentals and techniques*". CRC Press., FL (2008).

2. I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich and T. Kalker. "*Digital watermarking and steganography, 2nd Ed.*". Morgan Kaufmann., MA (2008).

3. G. Xuan, J. Zhu, J. Chen, Y. Q. Shi, Z. Ni and W. Su, "*Distortionless data hiding based on integer wavelet transform*". Electronics Letters, 38(25): 1646-1648, 2002.

4. H. C. Wu, N. I. Wu, C. S. Tsai and M. S. Hwang. "*Image steganographic scheme based on pixel-value differencing and LSB replacement methods*". IEE Proc. Vision Image Signal Processing, 152:611-615, 2005.

5. R. Z. Wang and Y. S. Chen. "*High-payload image steganography using two-way block matching*". IEEE T. Signal Processing Letter, 13(3):161-164, 2006.

6. H. M. Al-Otum and N. A. Samara. "*Adaptive blind wavelet-based watermarking technique using tree mutual difference*". Journal of Electronic Imaging, 15(4):043011-1~12, 2006.

7. X. Zhu, A. T. S. Ho and P. Marziliano. "*A new semi-fragile image watermarking with robust tampering restoration using irregular sampling*". Signal Processing: Image Communications, 22: 515-528, 2007.

8. Y. Govindarajan and S. Dakshinamurthi, "*Quality-security uncompromised and plausible watermarking for patent infringement*". International Journal of Image Processing, 1(2):11-20, 2007.

9. J. Tian. "*Reversible data embedding using a difference expansion*". IEEE T. Circuits and Systems for Video Technology, 13(8):890-896, 2003.

10. A. M. Alattar. "*Reversible watermark using the difference expansion of a generalized integer transform*". IEEE T. Image Processing, 13(8):1147-1156, 2004.

11. Z. Ni, Y. Q. Shi, N. Ansary and W. Su, "*Reversible data hiding*," IEEE T. Circuit and System for Video Technology, 16:354-362, 2006.

12. C. C. Lin and N. L. Hsueh. "*A lossless data hiding scheme based on three-pixel block differences*". Pattern Recognition, 41:1415-1425, 2008.

13. C. C. Lin, W. L. Tai and C. C. Chang. "*Multilevel reversible data hiding based on histogram modification of difference images*". Pattern Recognition, 41:3582-3591, 2008.

14. J. Y. Hsiao, K. F. Chan and J. M. Chang. "*Block-based reversible data embedding*". Signal Processing, 89:556-569, 2009.

15. H. W. Tseng and C. C. Chang. "An extended difference expansion algorithm for reversible watermarking". Image and Vision Computing, 26:1148-1153, 2009.

16. P. Tsai, Y. C. Hu and H. L. Yeh. "*Reversible image hiding scheme using predictive coding and histogram shifting*". Signal Processing, 89:1129-1143, 2009.

17. C. Saravanan and R. Ponalagusamy, "*Lossless grey-scale image compression using source symbols*". International Journal of Image Processing, 3(5):246-251, 2009.

CALL FOR PAPERS

Journal: International Journal of Image Processing (IJIP)
Volume: 3 **Issue**: 6
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
- Autonomous vehicles
- Character and handwritten text recognition
- Chemical and spectral sensitization
- Chemistry of photosensitive materials
- Coating technologies
- Coding and transmission
- Cognitive aspects of image understanding
- Color imaging
- Communication of visual data
- Data fusion from multiple sensor inputs
- Display and printing
- Document image understanding
- Generation and display
- Holography
- Image analysis and interpretation
- Image capturing, databases
- Image generation, manipulation, permanence
- Image processing applications
- 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: 1
Paper Submission: January 31 2010
Author Notification: February 28 2010
Issue Publication: March 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 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
CSC Press: cscpress@cscjournals.org
Info: info@cscjournals.org

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