Independent Component Analysis of Edge Information for Face Recognition

Kailash J. Karande kailashkarande@yahoo.co.in
Department of Information Technology,
Sinhgad Institute of Technology, Lonavala,
Dist-Pune, Maharashtra State. (India) 410401.

Sanjay N. Talbar sntalbar@yahoo.com
Department of Electronics & Telecommunication
SGGS Institute of Engineering &Technology, Nanded,
Maharashtra State, India.

Abstract

In this paper we address the problem of face recognition using edge information as independent components. The edge information is obtained by using Laplacian of Gaussian (LoG) and Canny edge detection methods then preprocessing is done by using Principle Component analysis (PCA) before applying the Independent Component Analysis (ICA) algorithm for training of images. The independent components obtained by ICA algorithm are used as feature vectors for classification. The Euclidean distance and Mahalanobis distance classifiers are used for testing of images. The algorithm is tested on two different databases of face images for variation in illumination and facial poses up to 1800rotation angle.

Keywords: Principle Component analysis (PCA), Independent Component Analysis (ICA), Laplacian of Gaussian (LoG) and Canny edge detection Euclidean distance classifier, Mahalanobis distance classifier.

1. INTRODUCTION

Face recognition is a task that humans perform routinely and effortlessly in their daily lives. Wide availability of powerful and low-cost desktop and embedded computing systems has created an enormous interest in automatic processing of digital images and videos in a number of applications, including biometric authentication, surveillance, human-computer interaction, and multimedia management. Research and development in automatic face recognition follows naturally.

Research in face recognition is motivated not only by the fundamental challenges this recognition problem poses but also by numerous practical applications where human identification is needed. Face recognition, as one of the primary biometric technologies, became more and more important owing to rapid advances in technologies such as digital cameras, the Internet and mobile devices, and increased demands on security. Face recognition has several advantages over other biometric technologies: It is natural, non intrusive, and easy to use. Among the six biometric attributes considered by Hietmeyer [1], facial features scored the highest compatibility in a Machine Readable Travel Documents (MRTD) [2] system based on a number of evaluation factors, such as enrollment, renewal, machine requirements, and public perception.

A face recognition system is expected to identify faces present in images and videos automatically. It can operate in either or both of two modes: (1) face verification (or authentication), and (2) face identification (or recognition). Face verification involves a one-to-one match that compares a query face image against a template face image whose identity is being claimed. Face identification involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face.

There are two predominant approaches to the face recognition problem: geometric (feature based) and photometric (view based). As a researcher interest in face recognition continued and many different methods are proposed, like well studied methods using Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Elastic Bunch Graph Matching (EBGM). In order to organize the vast field of face recognition, several approaches are conceivable. For instance, algorithms treating the face and its environment as uncontrolled systems could be distinguished from systems that control the lighting or background of the scene, or the orientation of the face. Or systems that use one or more still images for the recognition task could be distinguished from others that base their efforts on video sequences.

The comparison of face recognition using PCA and ICA on FERET database with different classifiers [3] [4] are discussed and found that the ICA has better recognition rate as compared with PCA with statistically independent basis images and also with statistically independent coefficients. Their findings are based on the frontal face image datasets are encouraging with few face expressions. Marian S Bartlett used version of ICA [5] derived from the principle of optimal information transfer through sigmoidal neurons on face images from FERET database has proved that ICA representation gave the best performance on the frontal face images. Feature selection in the independent component subspace [6] which gives the benefits for face recognition with changes in illumination and facial expressions. Fusion of ICA features like Spatial, Temporal and Localized features [7] [8] for Face Recognition are considered as optimization method. An independent Gabor features (IGFs) method and its application to face recognition using ICA [9] is discussed by Chengjun Liu; Wechsler, H. The novelty of the IGF method comes from 1) the derivation of independent Gabor features in the feature extraction stage and 2) the development of an IGF features-based probabilistic reasoning model (PRM) classification method in the pattern recognition stage. In particular, the IGF method first derives a Gabor feature vector from a set of down sampled Gabor wavelet representations of face images, then reduces the dimensionality of the vector by means of principal component analysis, and finally defines the independent Gabor features based on the independent component analysis (ICA).

The illumination has great influence on how a face image looks. Researchers have proved that for a face image, the difference caused by illumination changes has even exceeded the difference caused by identity changes [10]. The big challenge of face recognition lies in dealing with variations of pose, illumination, and expression. Also there is need to address the problem of identity changes using structural components.

In this paper we propose the face recognition using Edge information and ICA with large rotation angles with poses and variation in illumination conditions. Here we used the edge information of face images using different standard edge detection function operators like canny, Laplacian of Gaussian (log). The edge information is being used to calculate independent components for face recognition. We used the database which has the large rotation angles up to 180^0 change between the images of person while looking right and or left. We considered the face images having various orientations of the face i.e.: looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, looking down. In this study we considered the samples of individual person which consist of sufficient number of images having expressions, changes in illumination and large rotation angles. For illumination variation the effects of light on face image from left, right, top and bottom sides are considered. The paper is organized as follows. In Section 2 we introduce the ICA and information about edge detection in section 3. The section 4 specifies the need of the preprocessing before applying the ICA algorithm. The modified Fast ICA algorithm is presented in the Section 5 and classifiers are discussed in section 6. Experimental results are discussed in Section 7 and accordingly the conclusions are drawn in Section 8.

2. Introduction to ICA

ICA will be rigorously defined as a statistical 'latent variable' model. Assume that we observe *n* linear mixtures x_1, ..., x_n of *n* independent components

$$x_j = a_{j1}s_1 + a_{j2}s_2 + ... + a_{jn}s_n \quad \text{For all j.} \tag{1}$$

we have now dropped the time index *t*; in the ICA model, we assume that each mixture x_j as well as each independent component s_k is a random variable, instead of a proper time signal. The observed values $x_j(t)$, e.g. the microphone signals in the cocktail party problem, are then a sample of this random variable. Without loss of generality, we can assume that both the mixture variables and the independent components have zero mean. If this is not true, then the observable variables x_i can always be centered by subtracting the sample mean, which makes the model zero-mean.

It is convenient to use vector-matrix notation instead of the sums like in the previous equation. Let us denote by **x** the random vector whose elements are the mixtures x_1, ..., x_n. And likewise by **s** the random vector with elements s_1, ..., s_n. Let us denote by **A** the matrix with elements a_{ij}. Generally, bold lower case letters indicate vectors and bold upper-case letters denote matrices. All vectors are understood as column vectors; thus $\mathbf{x}^T$ or the transpose of **x**, is a row vector. Using this vector-matrix notation, the above mixing model is written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2}$$

Sometimes we need the columns of matrix **A**; denoting them by $\mathbf{a}_j$ the model can also be written as,

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i \tag{3}$$

The statistical model in equation.(2) is called independent component analysis or ICA model[11]. The ICA model is a generative model which means that it describes how the observed data are generated by a process of mixing the components s_i. The independent components are latent variables, meaning that they cannot be directly observed. Also the mixing matrix is assumed to be unknown. All we observe is the random vector **x**, and we must estimate both **A** and **s** using it. This must be done under as general assumptions as possible.

The starting point for ICA is the very simple assumption that the components s_i are statistically independent. It will be seen that we also assume that the independent component must have nongaussian distributions. However, in the basic model we do not assume these distributions known (if they are known, the problem is considerably simplified). For simplicity, we are also assuming that the unknown mixing matrix is square, but this assumption can be sometimes relaxed. Then, after estimating the matrix **A**, we can compute its inverse, say **W**, and obtain the independent component simply by:

$$\mathbf{s} = \mathbf{W}\mathbf{x}. \tag{4}$$

ICA is very closely related to the method called blind source separation (BSS) or blind signal separation. A “source” means here an original signal, i.e. independent component, like the speaker in a cocktail party problem. “Blind” means that we know very little, if anything, on the mixing matrix, and make little assumptions on the source signals. ICA is one method perhaps the most widely used, for performing blind source separation.

3. Edge Detection

Edges are boundaries between different textures. Edge also can be defined as discontinuities in image intensity from one pixel to another. The edges for an image are always the important characteristics that offer an indication for a higher frequency. Detection of edges for an image may help for image segmentation, data compression, and also help for well matching, such as image reconstruction and so on.

Laplacian of a Gaussian (LoG) Detector: Consider the Gaussian function

$$h(r) = -e^{-\frac{r^2}{2\sigma^2}} \qquad (5)$$

where $r^2=x^2+y^2$ and σ is the standard deviation. This is smoothing function which, if convolved with an image, will blur it. The degree of blurring is determined by the value of σ. The Laplacian function [13] is

$$\nabla^2 h(r) = -\left[\frac{r^2-\sigma^2}{\sigma^4}\right] e^{-\frac{r^2}{2\sigma^2}} \qquad (6)$$

For obvious reason, this function is called the Laplacian of a Gaussian (LoG). Because of the second derivative is a linear operation, convolving an image with $\nabla^2 h(r)$ is the same as convolving the image with the smoothing function first and then computing the Laplacian of the result. This is the key concept underlying the LoG detector.

Canny Edge Detection:

Finds edge by looking for local maxima of the gradient of f(x, y). The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds to detect strong and weak edges and includes the weak edges at the output only if they are connected to strong edges. Therefore this method is more likely to detect true weak edges. The Canny edge detector[13] is the most powerful edge detector provided by function edge. In this method the image is smoothed using a Gaussian filter with a specified standard deviation, σ, to reduce noise. The local gradient,

$$g(x,y) = \left[G^2 x + G^2 y\right]^{1/2} \qquad (7)$$

and edge direction,

$$\alpha(x,y) = \tan^{-1}(Gy/Gx) \qquad (8)$$

are computed at each point. Gx and Gy computed using sobel, prewitt or Roberts method of edge detection. An edge point is defined to be a point whose strength is locally maximum in the direction of the gradient.

4. Preprocessing by PCA

There are several approaches for the estimation of the independent component analysis (ICA) model. In particular, several algorithms were proposed for the estimation of the basic version of the ICA model, which has a square mixing matrix and no noise. Practically when applying the ICA algorithms to real data, some practical considerations arise and need to be taken into account. To overcome these practical considerations we implemented a preprocessing technique in this algorithm that is dimension reduction by principal component analysis. That may be useful and even necessary before the application of the ICA algorithms in practice. Overall face recognition benefits from feature selection of PCA and ICA combination [6].

A common preprocessing technique for multidimensional data is to reduce its dimension by principal component analysis (PCA) [3]. Basically, the data is projected linearly onto a subspace

$$\tilde{X} = E_n x \qquad (9)$$

so that the maximum amount of information (in the least-squares sense) is preserved. Reducing dimension in this way has several benefits. First, let us consider the case where the number of independent components (ICs) *n* is smaller than the number of mixtures; say m. Performing ICA on the mixtures directly can cause big problems in such a case, since the basic ICA model does not hold anymore. Using PCA we can reduce the dimension of the data to *n.* After such a reduction, the number of mixtures and ICs are equal, the mixing matrix is square, and the basic ICA model holds.

The question is whether PCA is able to find the subspace correctly, so that the *n ICs* can be estimated from the reduced mixtures. This is not true in general, but in a special case it turns out to be the case. If the data consists of n ICs only, with *no* noise added, the whole data is contained in an n-dimensional subspace. Using PCA for dimension reduction clearly finds this n-dimensional subspace, since the eigenvalues corresponding to that subspace, and only those eigenvalues, are nonzero. Thus reducing dimension with PCA works correctly. In practice, the data is usually not exactly contained in the subspace, due to noise and other factors, but if the noise level is low, PCA still finds approximately the right subspace. In the general case, some weak ICs may be lost in the dimension reduction process, but PCA may still be a good idea for optimal estimation of the strong ICs.

5. ICA Algorithm

Here we used the modified FastICA algorithm [11] for computing of independent components. Before computing the independent components we use to calculate the principle components and then whitened those components to reduce the size of matrix. In the previous section we have seen the need to use PCA before applying ICA. After finding principle components we whiten the eigenvector matrix to reduce the size of matrix and make it square. To estimate several independent components, we need to run the ICA algorithm several times with weight vectors **w**1, ...,**w***n*. To prevent different vectors from converging to the same maxima we must decorrelates the outputs $\mathbf{w}^T_1\mathbf{x}$, ...,$\mathbf{w}^T_n\,\mathbf{x}$ after every iteration.

A simple way of achieving decorrelation is a deflation scheme based on a Gram-Schmidt-like decorrelation[13]. This means that we estimate the independent components one by one. When we have estimated *p* independent components, or *p* vectors $\mathbf{w}_1$, ...,$\mathbf{w}_p$, we run the one-unit fixed-point algorithm for $\mathbf{w}_{p+1}$, and after every iteration step subtract from $\mathbf{w}_{p+1}$ the projections $\mathbf{w}^T{}_{p+1}\,\mathbf{w}_j\,\mathbf{w}_j$, j = 1, ..., p of the previously estimated p vectors, and then renormalize $\mathbf{w}_{p+1}$:

$$W_{p+1} = W_{p+1} - \sum_{j=1}^{p} W^T{}_{p+1} W_j W_j \qquad (10)$$

$$W_{p+1} = W_{p+1} / \sqrt{W^T{}_{p+1} W_{p+1}} \qquad (11)$$

In certain applications, however, it may be desired to use a symmetric decorrelation, in which no vectors are privileged over others. This can be accomplished, e.g., by the classical method involving matrix square roots,

$$W = (WW^T)^{-1/2} W \qquad (12)$$

where **W** is the matrix $(\mathbf{w}_1, ..., \mathbf{w}_n)^T$ of the vectors, and the inverse square root $(\mathbf{WW}^T)^{-1/2}$ is obtained from the eigenvalues decomposition of $\mathbf{WW}^T = \mathbf{FDF}^T$ as $(\mathbf{WW}^T)^{-1/2} = \mathbf{FD}^{-1/2}\mathbf{F}^T$. A simpler alternative is the following iterative algorithm,

$$W = W / \sqrt{\| WW^T \|} \qquad (13)$$

As we are using the ICA algorithm the training time required is more as compared to other methods. For the algorithm developed in this paper required around few seconds time as training time.

6. Similarity Metrics using Classifiers

As we are using the ICA algorithm the feature extraction time required is more as compared to other methods. For the algorithm developed in this paper required around few seconds time as extraction time. The query image will be more similar to the database images if the distance is smaller. For similarity measurement we used the Euclidean distance classifier and Mahalanobis distance classifier [12][14], for calculating the minimum distance between the query image and images to be matched from the database. If x and y are the two d-dimensional feature vectors of the database image and query image respectively, then these distance metrics are defined as follows. Euclidean distance to determine closeness reduces the problem to computing the distance measures:

$$d_E(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \qquad (14)$$

If the distance is small, we say the images are similar and we can decide which the most similar images in the database are. Another distance metrics for comparison of the retrieval of images used is Mahalanobis metric:

$$d_{Mah}(x, y) = \sqrt{(x-y)^2 Cov(x)^{-1} (x-y)} \qquad (15)$$

The results of these classifiers are very much close to each other. In Mahalanobis metrics the time required for similarity measure is more due to involvement of the covariance matrix.

7. Experimental Results

The experimental results presented in this paper are divided in to two parts. Both parts evaluate the face representation using Laplacian of Gaussian (LoG) and Canny edge detection for feature extraction methods. In first part the face images having variation in illumination conditions are used. In second part face images are used with large rotation angles. There are four stages of the algorithm developed in MATLAB environment. In the first stage we calculate the edge information using LoG or Canny edge detection method. The second stage is used for preprocessing of image matrix using PCA for dimension reduction. Third stage is used to find out independent components as feature vectors using ICA algorithm. In last stage we used different two classifiers for testing of input image to be recognized.

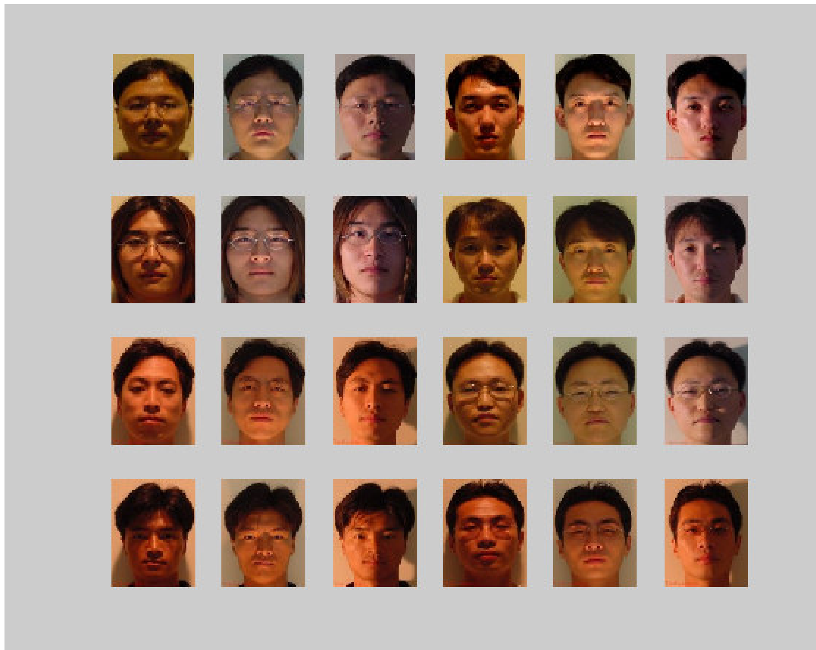

Figure 1: Various view of face images with different illumination conditions in Asian face database

Two standard databases are used for evaluation of these results. The first database is published by IIT Kanpur [16] widely used for research purpose known as *Indian face database*. In this database images of 60 persons with 10 sample images with different orientations and views are available. The second database known as *Asian face image database* [15] is from Intelligent Multimedia research laboratories having face images of 56 male persons with 10 samples each; which consist of variation in illumination conditions and different views. The resolution of all images we used in the algorithm is 128 x128 for computational purpose. Few face images are shown in Figure 1 and Figure 4 from both the databases with various views.

In this study we have explored feature selection techniques using Edge information and ICA for face recognition. Feature selection techniques are warranted especially for ICA features, since these are devoid of any importance ranking based on energy content. The study is carried out on the face database that contains both facial expression with pose variation and illumination variations. We implemented the algorithms for face recognition using Edge information and ICA on the different conditions of face images with different set of images. These set of images we used are by selecting the first 50, 100, 150 or 200 independent components.

Figure 2: Edge information of few face images

The edge detected images are shown in Figure 2 and are then preprocessed by PCA. The independent components are obtained by ICA algorithm and few are shown in Figure 3. These are the independent components obtained for face images with variation in illumination. The light condition on the face images are from left and right sides as well as from top and down directions. Under different illumination conditions the results are encouraging as shown in table 1.The images used in this part are from Asian face database with different illumination conditions as shown above in Figure 1. For training we used different sets of independent components varying from 50 to 200. The edge information for the images with large rotation angles are shown in Figure 5 and corresponding few independent components are shown in Figure 6. For this part we used images from Indian face database with large rotation angle up to 180^0.

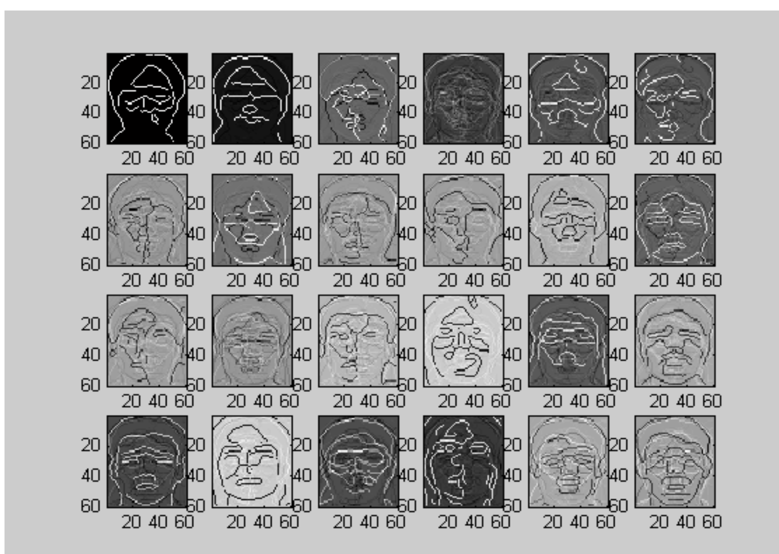

Figure 3: Independent Components of few face images

Figure 4: Various view of face images with different face orientations in Indian face database

Figure 5: Edge information of few face images

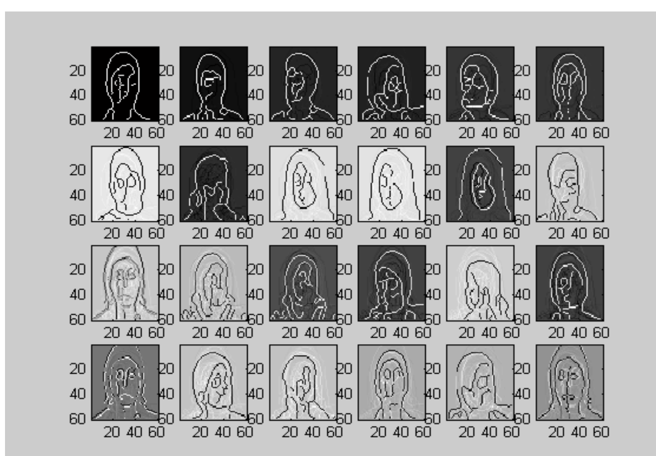

Figure 6: Independent Components of few face images

Conditions of images used for algorithm	Database used	No of Components used (PCs & ICs)	Recognition rate using LoG Edge detector and ICA (%)		Recognition rate using Canny Edge detector and ICA (%)	
Part I- Pose variation with large rotation angles	Indian face database.	Classifier used	Euclidean distance	Mahalanobis metric	Euclidean distance	Mahalanobis metric
		200	74	76.5	80.5	83.5
		150	79.33	82	84.67	86.66
		100	79	83	85	87
		50	82	86	88	91.5
Part II- Variation of illumination conditions.	Asian face database.	200	74	78	86	92.33
		150	81	83	90	93.33
		100	82	88.67	90	94
		50	84.5	87.5	91.5	95

Table 1: Results of the face recognition algorithm

8. Conclusion

In this paper an independent component analysis of Edge information of face images has been discussed and used for face recognition. Two standard databases are used, which contains the face images with different orientations, expressions and change in illumination. The performance of the algorithm suggested produces very good recognition rate varying from 74% to 95%. Applying ICA on face images for recognition; selection of ICs are required in order to give better performance. We used two different approaches for edge detection where Canny edge detector has given better results as compared to LoG edge detector. We adopted modified FastICA algorithm for computing the independent components. If we observe the results of face recognition using ICA under variation of illumination conditions the results are very good and encouraging. The results of the first part with ICA method are also close to the results of second part. This implies that ICA for edge information of face images is less sensitive to illumination change as compared to face orientations as shown in results.

9. REFERENCES

1. R. Hietmeyer. Biometric identification promises fast and secure processing of airline passengers. *The International Civil Aviation Organization Journal*, 55(9):10–11, 2000.
2. Machine Readable Travel Documents (MRTD). http://www.icao.int/mrtd/overview/overview.cfm.
3. Bruce A. Draper, Kyungim Baek, Marian Stewart Bartlett, “Recognizing faces with PCA and ICA”, Computer Vision and Image Understanding 91 (2003) 115-137.
4. Jian Yang, David Zhang, Jing-yu Yang, “Is ICA Significantly Better than PCA for Face Recognition?” Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05) 1550- 5499/05.
5. Marian Stewart Bartlett, Javier R. Movellan, Terrence J. Sejonowski, “Face Recognition by Independent Component Analysis”, IEEE Transactions on Neural Networks, vol-13, No-6, November 2002, PP 1450-1464.
6. H.K.Ekenel, B.Sankur, “Feature Selection in the Independent Component Subspace for Face Recognition”, Pattern Recognition Letters 25 (2004) 1377-1388.

7. Jiajin Lei, Chao Lu, “Face recognition by Spatiotemporal ICA using Facial Database Collected by AcSys FRS Discover System”, Proceedings of the Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD’06).
8. Jiajin Lei, Chao Lu, “Fusion of ICA Spatial, Temporal and Localized Features for Face Recognition”, Proceedings of the Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD’06).
9. Chengjun Liu; Wechsler, H.” Independent component analysis of Gabor features for face recognition”, Neural Networks, IEEE Transactions , Volume 14, Issue 4, July 2003 Page(s): 919 - 928
10. R. M. Bolle, J. H. Connell, and N. K. Ratha, “Biometric perils and patches,” Pattern Recognition vol. 35, pp. 2727 – 2738, 2002.
11. Aapo Hyvärinen and Erkki Oja “Independent Component Analysis: Algorithms and Applications” Neural Networks Research Centre Helsinki University of Technology P.O. Box 5400, FIN-02015 HUT, Finland, Neural Networks, 13(4-5):411-430, 2000.
12. Rafael C. Gonzalez and Richard E. Woods. “Digital image processing”, Second Edition, published by Pearson Education, 2003.
13. Aapo Hyvarinen, Juha Karhunen, Erkki Oja “Independent Component Analysis” Book by A Wiley Interscience Publication, John Wiley & sons, inc, New York.
14. Manesh Kokare, B.N.Chatterji and P K Biswas “Comparison of similarity metrics for texture image retrieval” International conference TENCON 2003, 571-574.
15. Asian face image database from Intelligent MultimediaLaboratory www.nova.postech.ac.kr / special / imdb /paper_pdf.pdf.
16. Indian face database www.cs.umass.edu / ~vidit / face database.