

An Efficient Bayesian Approach to Face Recognition based on Wavelet Transform

Esther Annlin Kala James
Professor and HOD/ECE,
Government College of Technology
Coimbatore –641 0 13

S. Annadurai
Additional Director,
Directorate of Technical Education,
Chennai-600 025

ABSTRACT

Face recognition plays a vital role in identification and for authentication purpose, in our day to day lives. In real time, this identification must be liable, perfect and faster. In this paper, face recognition approach using wavelet transform and Bayesian model is proposed, where the images have been recognized with minimal time and improved efficiency.

First the original image is decomposed into Low frequency and High frequency sub-band images by applying wavelet transform and then the Bayesian approach is used to find the intrapersonal difference between the images. The Principal Component Analysis algorithm is used to compute the eigenvector space of the face. The face recognition rate was gained through similarity measure using whitening transform.

In this paper, the performance of the proposed method is verified using the databases of face images, taken at different times, varying the lighting, facial expressions and facial details. In the proposed approach, the recognition rate is highly improved and hence it yields better classification compared to the existing approaches.

Keywords: Face recognition, Wavelet Transform, Whitening Transform, Bayesian Transform, Principal Component Analysis.

1. INTRODUCTION

Face recognition has become an important issue in many applications such as security systems, credit card verification and criminal identification.

The image of a human face has certain features that are common to all persons and some that exhibit the unique characteristics (also discriminatory) for a particular person. The task of face recognition involves the extraction of these unique features from the face image for discriminating it from other persons.

A successful face recognition methodology focus on the particular choice of the features to represent the face images., A face recognition approach combining Bayesian probabilistic model and wavelet transform is proposed[3]. In this paper only the intrapersonal variation is considered and whitening transform is used for the similarity measure.. Hence the computation complexity of the algorithm is considerably reduced.

Major advantage of wavelets is that they are very flexible. Several bases exist, and one can choose the basis which is more suitable for a given application and provide a spatial and frequential decomposition of the image at the same time. Also

computational complexity of wavelets is linear with the number (N) of computed coefficients (O(N)) while other transforms lead to $N \times \log_2(N)$ complexity.

2. WAVELET DECOMPOSITION

The two dimensional wavelet analysis efficiently decompose an image for extracting intrinsic facial features. A single level decomposition of the image is performed. This decomposition generates the coefficient matrices of the one level approximation and horizontal, vertical, and diagonal details, respectively. From the obtained coefficients, the approximation and three detailed images via the high pass and low pass filtering realized with respect to the column vectors and the row vectors of array pixels are constructed. In this manner, two level wavelet decomposition is performed. Thus the 2-D transform uses a family of wavelet functions and its associated scaling function to decompose the original image into different sub bands, namely the low-low (LL), low-high (LH), high-low (HL) and high-high (HH) sub-bands.

3. PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) has been widely adopted as the most promising face recognition algorithm. PCA is used extensively in all forms of analysis from neuroscience to computer graphics because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it.

PCA is an orthogonal transformation of the coordinate system in which the pixels are described. The main idea of the principal component analysis is to find the vectors which best describe the distribution of face images within the entire image space and it aims to extract a subspace where the variance is maximized. PCA is performed by projecting a new image into the subspace called face space spanned by the eigenfaces and then classifying the face by comparing its position in face space with the positions of known individuals.

A face image in 2-dimension with size $N \times N$ can also be considered as one dimensional vector of dimension N^2 . The main idea of the principle component is to find the vectors that best account for the distribution of face images within the entire image space. These vectors define the subspace of face images, which we call "face space". Each of these vectors is of length N^2 , describes an $N \times N$ image, and is a linear combination of the original face images. These vectors are the eigenvectors of the

covariance matrix corresponding to the original face images, and because they are face-like in appearance, they are referred as “eigenfaces”.

3.1 Steps for the PCA algorithm

Step 1: The normalized training image in the N -dimensional space is stored in a vector of size N . Let the normalized training face image set,

$$T = \{X_1, X_2, \dots, X_N\} \text{ where } X = \{x_1, x_2, \dots, x_m\}^T$$

Step 2: Each of the normalized training face images is mean centered. This is done by subtracting the mean face image from each of the normalized training images. The mean image is represented as a column vector where each scalar is the mean of all corresponding pixels of the training images,

$$\bar{X}_i = X_i - \bar{X}$$

where the average of the training face image set is defined as:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

Step 3: Once the training face images are centered, the next process is to create the Eigenspace which is the reduced vectors of the mean normalized training face images. The training images are combined into a data matrix of size N by P , where P is the number of training images and each column is a single image.

$$\bar{X} = \{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p\}$$

Step 4: The column vectors are combined into a data matrix which is multiplied by its transpose to create a covariance matrix. The covariance is defined as:

$$\Omega = \bar{X}^T \bar{X}$$

Step 5: The eigenvalues and corresponding eigenvectors are computed for the covariance matrix using Jacobian transformation,

$$\Omega V = \Lambda V$$

Where v is the set of eigenvectors associated with the eigenvalues Λ

Step 6: Order the eigenvectors $V_i \in V$ according to their

corresponding eigenvalues $\lambda_i \in \Lambda$ from high to low with non-zero eigenvalues. This matrix of eigenvectors is the eigenspace V , where each column of V is an eigenvector. The principal components are the eigenspace.

$$V_i = \{V_1, V_2, \dots, V_p\}$$

4. BAYESIAN TECHNIQUE

A probabilistic similarity measure based on the Bayesian belief that the image intensity difference, denoted by $\Delta = I_1 - I_2$, is the characteristic of typical variations in appearance of an individual.

The two classes of facial image variations

1. *Intrapersonal* variations Ω_I (corresponding to different facial expressions of the same individual)

2. *Extrapersonal* variations Ω_E (corresponding to variations between different individuals)

The similarity measure is expressed in terms of the probability as,

$$S(I_1, I_2) = P(\Delta \in \Omega) = P(\Omega | \Delta)$$

Where $P(\Omega_I | \Delta)$ is the a posteriori probability given by Bayes rule, using estimates of the likelihood $P(\Delta | \Omega_I)$ and $P(\Delta | \Omega_E)$.

$$S(I_1, I_2) = P(\Omega_I | \Delta)$$

$$= \frac{P(\Delta | \Omega_I) P(\Omega_I)}{P(\Delta | \Omega_I) P(\Omega_I) + P(\Delta | \Omega_E) P(\Omega_E)} \quad (1)$$

This particular Bayesian formulation, casts the standard face recognition task into a binary pattern classification problem with

$$\Omega_I \text{ and } \Omega_E$$

This problem is then solved using the Maximum a Posteriori (MAP) rule – *i.e.*, two images are determined to belong to the same individual if $P(\Omega_I | \Delta) > P(\Omega_E | \Delta)$, or equivalently, if $S(I_1, I_2) > 1/2$.

An alternative probabilistic similarity measure can be defined in simpler form using the intrapersonal likelihood alone,

$$S' = P(\Delta | \Omega_I) \quad (2)$$

thus leading to Maximum Likelihood (ML) recognition as opposed to the Maximum a Posteriori (MAP) recognition. This simplified ML measure can be almost effective as its MAP counterpart in most cases.

The complete likelihood estimate can be written as the product of two independent marginal Gaussian densities

$$\hat{P}(\Delta | \Omega_I) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^M \frac{\Delta_i^2}{\lambda_i}\right)}{(2\pi)^{M/2} \prod_{i=1}^M \lambda_i^{1/2}} \left[\frac{\exp\left(-\frac{\epsilon^2(\Delta)}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}} \right]$$

As large number of images in the database is used, the computation of this method is tedious and hence whitening transformation is applied.

4.1 Use of PCA in Bayesian Method

The difference Δ between the two images is modeled by the three components

- Intrinsic difference (I) that discriminates different individuals.
- Transformation difference (T) caused by such transformations as lighting or expression changes, and
- Random noise (N).

T and N are two components deteriorating the recognition performance. Normally N is of small energy. Under a large transformation, T could potentially be greater than I.

A successful subspace method should be able to reduce the effect of T and N as much as possible without sacrificing much of I. By analyzing the distribution of I, T and N in the PCA and Bayesian analysis, T and N can be effectively removed.

4.2 Whitening Transformation

Another useful preprocessing strategy in Bayesian method is to first whiten the observed variables. This means that before the application of the Bayesian algorithm (and after centering), transform is applied to the observed vector x linearly so that a new vector \bar{x} is obtained which is white, i.e. its components are uncorrelated and their variances equal to unity. In other words, the covariance matrix of \bar{x} equals the identity matrix:

$$E\{\bar{x}\bar{x}^T\} = I$$

These computations can be greatly simplified by offline transformations. To compute the likelihoods $P(\Delta|\Omega_j)$ and $P(\Delta|\Omega_e)$ pre-process the I_k images with the whitening transformation and consequently every image is stored as two vectors of whitened subspace coefficients; i for intrapersonal and e for extrapersonal

$$i_j = \Lambda^{-1/2} V_I^T I_j \quad e_j = \Lambda^{-1/2} V_E^T I_j \quad (3)$$

where, Λ and V are matrices of the largest eigenvalues and eigenvectors of Σ_E or Σ_I , with subspace dimensionalities of M_I and M_E , respectively. After this pre-processing and with the normalizing denominator pre-computed, evaluating the likelihood is reduced by computing simple Euclidean distances for the exponents

$$P(\Delta | \Omega_E) = \frac{e^{-\frac{1}{2} \|e_j - e_k\|^2}}{(2\pi)^{\frac{D}{2}} |\Sigma_E|^{\frac{1}{2}}}$$

$$P(\Delta | \Omega_I) = \frac{e^{-\frac{1}{2} \|i_j - i_k\|^2}}{(2\pi)^{\frac{D}{2}} |\Sigma_I|^{\frac{1}{2}}}$$

These likelihoods are then used to compute the Maximum a Posteriori (MAP) similarity S in Equation 1. Since the Euclidean distances in the exponents of Equation 3 are of dimensions M_I and M_E for the i and e vectors, respectively, only $2 \times (M_I + M_E)$ arithmetic operations are required for each similarity computation. Thus, one avoids unnecessary and repeated image

differencing and online projections. The Maximum Likelihood (ML) similarity matching equation is simpler than Maximum a Posteriori (MAP) to implement. In this framework, since only the intrapersonal class is evaluated, it leads to the simplified similarity measure, by computing just the I vectors alone.

$$S' = P(\Delta | \Omega_I) = \frac{e^{-\frac{1}{2} \|i_j - i_k\|^2}}{(2\pi)^{\frac{D}{2}} |\Sigma_I|^{\frac{1}{2}}} \quad (4)$$

5. PROPOSED METHOD

In the proposed method the face is recognized using wavelet transform with Bayesian classifier; which improves the efficiency by providing high recognition rate with minimal elapsed time. The block diagram for proposed method is shown in Figure 1.

1. Two level of discrete wavelet decomposition is applied for training images using Daubechies 4 DWT. In the matrix form of db4 algorithm, there is a overlap between iterations, hence it picks up details missed by haar transform.
2. For the decomposed LL image, intrapersonal differences are calculated using Bayesian classifier.
3. The difference matrix is given to PCA to find the eigenvalue and eigenvector.
4. In Whitening transform, the whitening coefficient for top non-zero Eigen values are computed. The top most fifty five eigen values are taken for the recognition purpose, thus discarding the noise in data while preserving meaningful information. Thus the dimensionality is greatly reduced and in this transformation there is no loss of information since reduction is achieved through elimination of the redundant data .
5. Similarly the above steps are repeated for the test images.
6. Finally similarity measure is carried out between train and test images using Bayesian technique.

In order to have an efficient similarity measure the zero mean intrapersonal difference image is used and the principal eigenvectors are used to find the similarity score.

Computation of this score, involves calculating the intrapersonal differences between each image of a subject and projecting onto the principal eigenvectors. The exponentials are then evaluated, normalized and combined as likelihoods. This operation is iterated over all members of database. And the one which has the maximum score is chosen as the best match.

In this offline transformation, the whitened subspace is obtained using the eigenvectors and eigenvalues. Thus the Maximum Likelihood similarity measure is obtained using the whitening coefficients.

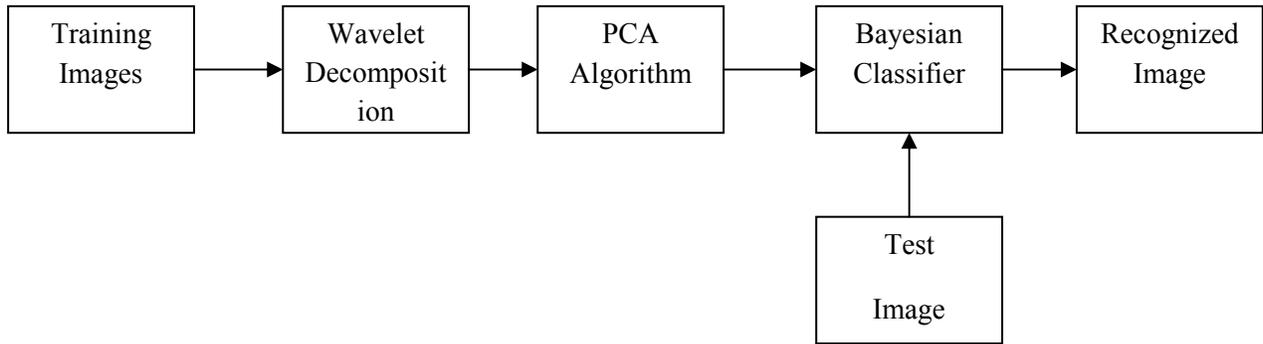


Figure 1: Block diagram of proposed method

6. EXPERIMENTAL RESULT

The above proposed method is tested on ORL, Yale and Yale B face databases. The images for the training set and testing set are selected randomly with a size of 112 x 112.

This method is quite robust, despite large changes in the visual stimulus due to viewing conditions, expression, aging and distractions such as glasses, beards or changes in hairstyle.

To evaluate a face matching, a system was built that uses both PCA matching and Bayesian method. Compared to the standard euclidean eigenfaces, the new probabilistic similarity measure using the intrapersonal eigenfaces shows better experimental results on Table 1 than the performance of conventional algorithms.

Table 1 : Percentage of recognition rate

Database	% of recognition rate		
	PCA	LDA	Proposed Method
ORL	91.23	93.60	97.82
Yale	90.18	92.56	95.76
YaleB	89.05	91.34	94.66



Figure 2: Some images of ORL database



Figure 3: Some images of Yale database

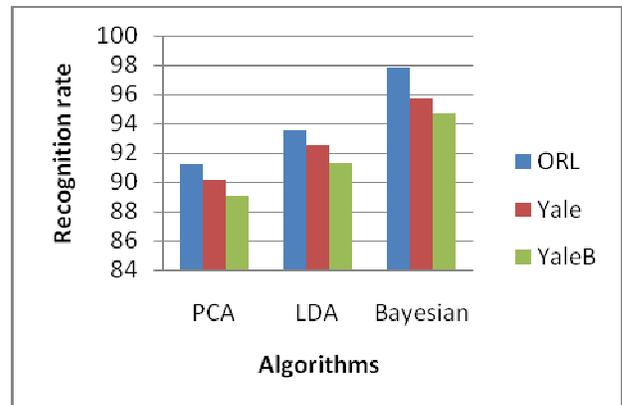


Figure 4: Comparison of recognition rates

7. CONCLUSION

In this paper, an efficient bayesian approach to face recognition based on wavelet transform is proposed. The experimental result shows that the face recognition performance is relatively unaffected even though there is a transformation of the face including translation, small rotation and illumination.

The PCA algorithm is used because it is the basic and straight forward method for feature extraction. It provides efficient results and requires less storage.

The ML classifier outperforms the Bayesian MAP which is complex and expensive than the Bayesian ML classifier.

The future efforts will be on the recognition of face images in dynamic video sequences and real time tasks using fuzzy logic.

8. REFERENCES

- [1] W. Zhao, R. Chellappa, A. Rosenfeld and P.J. Phillips, "Face Recognition: A Literature Survey", *ACM Computer Survey.*, vol 35, no. 4, pp 299-458, 2003
- [2] Tsai, G.S.-Y.; Tang, A.C.-W.; , "Two-view Face Recognition using Bayesian Fusion," *Systems, Man and Cybernetics*, 2009. SMC 2009. IEEE International Conference on , vol., no., pp.157-162, 11-14 Oct. 2009.
- [3] Niu Liping, Zheng Yanbin et.al. "Combined Face Recognition Using Wavelet Transform and Bayesian" *IEEE Second International Conference on Information and Computing Science*, 2009.
- [4] Moghaddam. B, T. Jebara, and A. Pentland, "Bayesian Face Recognition", *Pattern Recognition*, Vol. 33, pp, 1771-1782, 2000
- [5] Ming-Yuan Shieh, Choung-Ming Hsieh, Jian-Yuan Chen, Juing - Shian Chiou, "PCA and LDA based Fuzzy Face Recognition System", *SICE Annual Conference 2010*, August 18-21, 2010.
- [6] Jian Yang, Zhang. Dand Jing-yu Yang, "Two dimensional PCA: A new approach to Appearance based Face Representation and Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence* .Vol26.(2004)131-137.
- [7] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi and Hong-jiang Zhang, "Face Recognition using Laplacianfaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol.27, No.3, March 2005.
- [8] Kresimir Delac, Mislav Grgic, Sonja Grgic, "Independent comparative study of PCA, ICA and LDA on the FERET Data set," *Wiley Periodicals*, vol.15, 252-260(2005).
- [9] Mogaddam B.,Nastar C.,Pentland A., "A Bayesian Similarity Measure for Direct Image Matching", *Massachusetts Institute of Technology, Cambridge, MIT Media lab PCS TR39*, 1996.
- [10] Byung-joo oh , "Face Recognition by using Neural Networks Classifiers based on PCA and LDA," *Systems, Man and Cybernetics*, 2005 *IEEE International Conference on*, vol.2, no.pp.1699-1703 vol.2,10-12 Oct. 2005
- [11] Hadi Sadoghi Yazdi, Abedin Vhedian., "Fuzzy Bayesian classification of LR Fuzzy numbers", *IACSIT International Journal of Computer Theory and Engineering*, Vol.1., no 5, December 2009.
- [12] Keun-Chang Kwak, Witold Pedrycz., "Face Recognition using Fuzzy integral and wavelet decomposition method", *IEEE Transactions systems, Man and cybernetics*, Vol 34., No.4, August 2004.
- [13] Minghua Wan, Zhihui Lai, Jie shao, Chuanbo Huang, Zong Jin, "Fuzzy Local Discriminant Embedding (FLDE) for Face Recognition", *The 1st International conference on Information Science and Engineering (ICISE 2009)*.
- [14] P.N. Belhumeur, J.P. Hespanha and D.J.Kriegman, "Eigen faces versus. Fisher Faces: Recognition using class specific linear projection", *IEEE Trans. Pattern Analysis Machine Intelligence*, July 1997, Vol.19, pp.711-720.
- [15] S.Yan, M. Li, H.J. Zhang and Q. Cheng, "Ranking Prior Likelihood Distributions for Bayesian Shape Localization Framework", *Proceedings IEEE International Conference Computer Vision 2003*.