Comparison of PCA, LDA, ICA, SVM & HGPP

Ajeet Singh
Lecturer, ECE
VITS
Ghaziabad, India
ajeetranaut@gmail.com

Bhupesh Bhatia
Asst. Prof. ECE
GPMCE
Delhi India
bhupeshbhatia30@gmail.com

VijayRaj Shokeen
Lecturer ECE
GPMCE Delhi
India
vrshokeen@gmail.com

Dr. B.K. Singh
Prof. ECE
BTKIT Almorha
India
bksapkec@yahoo.com

***Abstract*— In the field of face recognition, this paper explores a comparison of five most popular algorithms. These algorithms are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) , Support Vector Machine (SVM) and Histogram of Gabor Phase Patterns(HGPP). The performance of the algorithms have been measured in terms of the accuracy, training time, testing time, total execution time and memory usage for train and test the databases. The algorithms have been tested on the AT&T and IFD face database. The investigation shows that SVM outperforms the rest of the algorithms.**

Keywords— PCA, ICA, LDA, SVM, HGPP and face recognition

I. INTRODUCTION

Face of human being is the prime attraction and it is very easy to convey identity and emotion. Still, it is not possible to infer intelligence and character from the facial appearance but ability of faces recognition of human beings are remarkable. Various changes in the visual stimulus of a faces of human beings due to changes in view conditions, facial expression, aging and distractions such as change in hair style, facial hairs, glasses on eyes do not restrict this skill of human beings. This robust skill inspired the researchers to develop the computational model for face recognition in the prospect of various practical applications such as criminal identification, security systems, image and film processing and human computer identification.

After analyzing the performance of all the algorithms discussed in literature survey, it is interesting to note that there is often 'contradictory' and confusing claims that have been made in the literature. For example, Bartlett et al. [1] and Liu et al. [2] claim that ICA outperforms PCA, while Baek et al. [3] claim that PCA is better. Moghaddam [4] states that there is no significant difference. Beveridge et al. [5] claim that in their tests LDA performed uniformly worse than PCA, Martinez [6] states that LDA is better for some tasks, Belhumeur et al. [7] and Navarrete et al. [8] claim that LDA outperforms PCA and Brain C. Becker et al. [9] claim that SVM performs well as compare to above said algorithms but Zhang et al found HGPP as a novel object representation approach for face recognition.

While all these claims may in fact hold a good degree of truth, one should bear in mind that there were differing control factors surrounding each conclusion i.e. the actual task statement, the subspace distance metrics, dimensionality retention and the non-standardized database choices etc [10, 11]. All these conclusions have contributed too much debate and confusion over the years, particularly for an individual who is new in the field of face recognition and who seeks a good comparative understanding of the available techniques.

Presently, large numbers of approaches or algorithms are available for the face recognition. Among of them, well known five algorithms i.e. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component analysis (ICA), support Vector Machine (SVM) and Histogram Gabor Phase Pattern (HGPP) have been considered for the comparison. A brief description of these algorithms is given below:

II. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) also known as Karhunen-Loeve expansion. It is a classical feature extraction and data representation algorithm, widely used in the areas of pattern recognition and computer vision. Sirovich and Kirby [12], [13] first used PCA to efficiently represent pictures of human faces. They argued that any face image could be reconstructed approximately as a weighted sum of a small

SEEK DIGITAL LIBRARY

collection of images that define a facial basis (eigenimages), and a mean image of the face. Within this context, Turk and Pentland [14] presented the well-known Eigenfaces method for face recognition in 1991.

This approach of face recognition involves the following initialization operations [14]:

- Acquiring an initial set of N face images (training images).
- Calculation of the Eigenface from the training set keeping only the M images that correspond to the highest eigenvalues. These M images define the “face space”. As new faces are encountered, the “eigenfaces” can be updated or recalculated accordingly.
- Calculation of the corresponding distribution in M-dimensional weight space for each known individual by projecting their face images onto the “face space”.
- Calculation of a set of weights projecting the input image to the M “eigenfaces”.
- Determine whether the image is a face or not by checking the closeness of the image to the “face space”.
- If it is close enough, classify the weight pattern as either a known person or as an unknown based on the Euclidean distance measured.
- If it is close enough then cite the recognition successful and provide relevant information about the recognized face form the database which contains information about the faces.

Mathematically it can be shown as follows:

- Assume (X_1, X_2, X_3, …, X_m) is a set of M train set from N face images arranged as column vector. $[X_i] = [X_1, X_2, X_3, \dots, X_m]^T$ ---(1)
- Average face of set can be defined as: $\Psi = \left(\frac{1}{M}\right) \sum_{n=1}^{M} (X_n)$ --(2)
- Each face differs from the average by vector $\Phi_i = X_i - \Psi$ - -(3)
- When applied to PCA, this large set of vectors seeks a set of M orthogonal vectors U_n, which describes the distribution of data.
- The K^{th} vector $\mathbf{U}_k$ is chosen such that $\lambda_k = \left(\frac{1}{M}\right) \sum_{n=1}^{M} [[(\mathbf{U}_k)^T * \Phi_n]^2$ --(4)

is maximum, applied to $(\mathbf{U}_l)^T \mathbf{U}_k = \delta_{lk} =$

$$f(x) = \begin{cases} 1, if\ l = k \\ 0, otherwise \end{cases} \quad \text{--(5)}$$

- The vector U_k and scalar λ_k are the eigenvectors and eigenvalues respectively of the covariance matrix (Scatter matrix)

$$C = \left(\frac{1}{M}\right) \sum_{n=1}^{M} (\Phi_n)(\Phi_n)^T \quad \text{--(6)}$$

$= AA^T$

Where the matrix $A = [\Phi_1, \Phi_2 \dots\dots \Phi_M]$.

- Consider the eigenvector v_i of AA^T such that $AA^T v_i = \mu_i v_i$
- Multiplying both side by A, $AA^T A v_i = \mu_i A v_i$

Av_i is the eigenvectors of $C = AA^T$

- By following this analysis, we construct the $L = A^T A$, where $L_{mn} = \Phi_m{}^T \Phi_n$ and find the eigenvectors v_i of L.
- These vectors form the eigenvalues u_l. $u_l = \sum_{k=1}^{M} v_{lk} \Phi_k$, $l = 1, \dots\dots$M

The last step is to classify an image of face.

Now, there is need to transform the image of face into its eigenfaces components to calculate the vector of weights $W^T = [w_1, w_2, \dots, w_m]$, where $w_k = u_k^T (X - \Psi)$ for $k = 1, 2, \dots, m$; here m does not represent the total eigenfaces but it is the one of the greater value. Now by comparing all the test images with the weights of the training images, the best possible match is found out. The comparison is done using the “Euclidean distance” measurement. Minimum distance shows the maximum match.

Minimum Euclidean distance can be expressed as:

$$\varepsilon_k = \|\Omega - \Omega_k\|^2 \quad \text{--(7)}$$

Where, Ω_k represent the vector of face image of number k.The input face belongs to a class if ε_k which is below the established threshold θ_ε, then the face image is considered to be a ‘known’ face. If the ε_k is above the given threshold, but below a second threshold, the image can be considered as an ‘unknown’ face. If the input image is above these two thresholds, the image is considered not to be a face.

III. LINEAR DISCRIMINANT ANALYSIS (LDA)

Although it is true that the Eigenface approach is good in selecting a subspace which retains most of

SEEK DIGITAL LIBRARY

the data variation but it does not employ any optimal properties which may facilitate class discriminality [15, 16]. In Eigenface approach or PCA, the training data is taken as a whole, wherein projections are made very close together resulting in information defining class membership and identity being mixed or lost all together i.e. PCA makes no use of between-class scatter [17].

Here, a class based scheme has been proposed in which the faces of the same subjects are grouped into separate classes and the variations the images between the images of different classes are discriminated using the eignvectors. At the same time covariance is minimized within the same class [18]. By defining all the face images of the same person in one class and faces of the other people in the different classes we can establish a model for performing cluster separation analysis. It is done by achieving two terms named "between class scatter matrix" and "within class scatter matrix". This approach is named as Linear Discriminant Analysis or Fisher's Discriminant Analysis (FLD) [19].

Mathematically, it can be achieved by calculating an optimal projection matrix(W_{LDA}) such that the Fisher Discriminant Criterion is maximizing in the following form:

$$W_{LDA} = arg\left(max_W \frac{|W^T S_B W|}{|W^T S_W W|}\right)$$

$$= [w_1, \ldots\ldots, w_D] \qquad \text{--(8)}$$

Now, the aim is to maximizing the between – class measure and Within-class measure, given below:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \Psi)(\mu_i - \Psi)^T \qquad \text{--(9)}$$

and

$$S_W = \sum_{i=1}^{c} \sum_{x_m \epsilon x_m} (x_m - \mu_i)(x_m - \mu_i)^T \qquad \text{--(10)}$$

where N_i represents the number of training samples per class i, μ_i represents the mean vector of samples belonging to class i and X_i is the sets of samples belonging to a particular class. S_B represents the scatter of features around the overall mean for all face classes and S_W represents the scatter of features around the mean of each class [18].

After it, it is not difficult to show that the solution to the maximization problem in (6) lies in the solution of a generalized eigensystem given by:

$$S_B.W = \Lambda.S_W.W \qquad \text{--(11)}$$

Where $W = [w_1, \ldots\ldots, w_D]$represents the eigenvector (fisher face) matrix and $\Lambda = [\lambda_1, \ldots.., \lambda_D]$ represent the corresponding eigenvalues of the between and within-class matrices [18]. In directly solving this eigensystem, one can run into problems that may include the fact that:

i. The eigensystem does not have orthogonal eigenvector because $S_W^{-1}.S_B$ is, is general, not symmetric [18].

ii. The matrices of S_W andS_B are usually too large, given the pixel dimensionality.

iii. There are at most *c* – 1 nonzero generalized eigenvectors meaning that *D* is limited to an upper bound of *c* – 1 [6].

iv. Because the rank of S_W is at most *X* – *c*, and often the number of images in the learning set *X* is much smaller than the image dimensionality, S_W easily becomes singular and hence noninvertible [6, 18].

In addressing the latter problems, particularly the issue of singularity it was proposed, by Belhumer [7] and Martinez [6], that an intermediate subspace be used before projecting the image into the final LDA subspace. This intermediate space was achieved by employing the PCA algorithm prior to LDA projection. This method known as the Fisher face approach [7], found much success in improving recognition results and maintaining the non-singularity of S_W [16, 20].

After solving for the eigenvectors and hence the optimal transformation matrix W_{LDA} is found, projection follows very similar to their PCA algorithm whereby the basis vectors are again formulated by calculating the dot product of the images with each of the eigenvectors such that:

$$\boldsymbol{Y} = W_{LDA}{}^{T}(\boldsymbol{X} - \boldsymbol{\Psi}) \qquad \text{--(12)}$$

Recognition again follows the PCA methodology, whereby the probe images are projected into the same face space as the training centroid and classification is achieved by virtue of nearest-neighbor distance metrics.

IV. INDEPENDENT COMPONENT ANALYSIS (ICA)

Face recognition is a task in which most of the important information may be contained in the higher order relationships among the image pixels. Therefore, it is more advantages to investigate the generalization of PCA which are sensitive to high order relationships, rather not just to second order relationships. Independent Component Analysis (ICA) is that type of generalization which was expected [21]. It minimizes both second and higher-order dependencies in an effort to provide a more powerful data representation.

There are number algorithms for performing ICA. Among of them we chose the following one.

SEEK DIGITAL LIBRARY

A. INFOMAX ALGORITHM

Here, we have chosen the infomax algorithm proposed by Bell and Sejnowshi [23, 24, 21, 22]. It was derived from the principle of optimal information transfer in neurons with sigmoidal transfer function [93]. This algorithm is motivates as: let **X** be an n-dimensional (n-D) random vector. It is representing a distribution of inputs in the environment. Assume W is a n $\times$ n invertible matrix, **U** = W**X** and **Y** = f (**U**) an n-D random variable which is representing the outputs of n-neurons. Each component of $f = (f_1, \dots\dots, f_n)$ is an invertible squashing function and it is mapping real numbers into [0,1] interval.

Typically, the logistic function is used

$$f_i(u) = \frac{1}{1+e^{-u}} \quad \text{--(13)}$$

The $U_1, \dots\dots, U_n$ are random variables. These random variables are linear combinations of inputs and it can interpret as presynaptic activations of n – neurons.

The $Y_1, \dots\dots, Y_n$ are random variables. These random variables are represented as postsynaptic activation rates. These are bounded by the interval [0, 1].

The goal of this algorithm is to maximize the mutual information between the **X** (inputs in the environment) and **Y** (output of the neural network). This can be achieved by performing gradient ascent on the entropy of the output with respect to the weight matrix W. The gradient update rule for the weight matrix, W is as follows:

$$\Delta W \propto \nabla_w H(\mathbf{Y}) = (W^T)^{-1} + E(Y'X^T) \quad \text{--(14)}$$

Here, $Y_i' = \frac{f''(U_i)}{f'(U_i)}$ is the ratio between the second and the first partial derivatives of the activation function, T represents the transpose, E is the expected value, H(**Y**) represents the entropy of the random vector **Y** and $\nabla_w H(\mathbf{Y})$ represents the gradient of the matrix form i.e. the cell in row I, column j of this matrix represent the derivatives of $H(\mathbf{Y})$ with respect to W_{ij}.

Here, we can avoid the computation of the matrix inverse by employing the natural gradient [25], which is amount to multiplying the absolute gradient by $W^T W$. it results the following learning [26].

$$\Delta W \propto \nabla_w H(\mathbf{Y}) W^T W = (I + Y'U^T)W \quad \text{--(15)}$$

Here, I is the identity matrix. The logistic transfer function gives $Y_i' = (1 - 2Y_i)$. --(16)

When, we have multiple inputs and outputs. It maximizes the joint of the output **Y**. further it encourages the individual outputs to move towards statistical independence.

This algorithm can be speeded up by including a "sphering" step prior to learning [26]. The row means of **X** are subtracted and then **X** is passed through the whitening matrix W_z, which is twice the inverse square root of the covariance matrix

$$W_z = 2 * (Cov(X))^{(-1/2)} \quad \text{--(17)}$$

It removes the first and the second order statistics of the data. Due to this, both the mean and covariance are set to zero and the variances are equalized. When the inputs to ICA are "sphered" data, the full transform matrix W_I is the product of the sphering matrix learned by ICA

$$W_I = W W_z \quad \text{--(18)}$$

Mackey [27] and Pearlmutter [28] have shown that the ICA algorithm converges to the maximum likelihood estimate of W^{-1} for the following generative model of the data:

$$X = W^{-1} S \quad \text{--(19)}$$

Where, $S = (S_1, \dots\dots, S_n)'$ is a vector of independent random variable with cumulative distribution equal to f_i. **S** is called as source.

From here, we can conclude that by using logistic activation functions corresponds to assuming logistic random sources and using the standard cumulative Gaussian distributions as activation functions corresponds to assuming Gausssian random sources. Thus, W^{-1}, the inverse of the weight matrix in Bell and Sejnowski's algorithm, can interpreted as the source mixing and $U = WX$ variables can be interpreted as the maximum-likelihood (ML) estimates of the sources that generated the data.

At last we can conclude, Independent component analysis (ICA) extracts statistically independent variables from a set of measured variables, where each measured variable is affected by a number of underlying physical causes. Extracting such variables is desirable because independent variables are usually generated by different physical processes. Thus, by extracting independent variables, ICA can effectively extract the underlying physical causes for a given set of measured variables.

Here in this investigation, we have used Arch I.

V. SUPPORT VECTOR MACHINES (SVMs)

The Support Vector Machine is based on VC theory of statistical learning. It is implement

SEEK DIGITAL LIBRARY

structural risk minimization [29]. Initially, it was proposed as per a binary classifier. It computes the support vectors through determining a hyperplane. Support Vectors maximize the distance or margin between the hyperplane and the closest points.

Assume a set of N points and $X_i \in \boldsymbol{R}^n$, i=1, 2, 3,……,N. Each point belongs to one of the two classes i.e. $\boldsymbol{Y}_i \in \{-1,1\}$. Here optimal separating hyperplane (OHS) can be defined as

$$f(x) = \sum_{i=1}^{l} \alpha_i Y_i \boldsymbol{X}_i . \boldsymbol{X} + b \qquad \text{… (20)}$$

The coefficients α_i and b are the solution of a quadratic equation [30]. Sign of f(x) decides the 'Classification' of a new point data in the above equation.

In the case of multi-class classification the distance between hyperplane and a data set can be defined as:

$$d(x) = \frac{\sum_{i=1}^{l} \alpha_i Y_i X_i . X + b}{\|\sum_{i=1}^{l} \alpha_i Y_i X_i\|} \qquad \text{… (21)}$$

Larger |d| shows the more reliable classification.

VI. HISTOGRAM OF GABOR PHASE PATTERNS (HGPP)

HGPP is the combination of spatial histogram and Gabor phase information [33]. Gabor phase information is of two types. These are known as Global Gabor phase pattern (GGPP) and Local Gabor phase pattern (LGPP). Both of the Gabor phase patterns are based on quadrant-bit codes of Gabor real and imaginary parts ($P_{u,v}^{Re}(Z), P_{u,v}^{Im}(Z)$). Quadrant–bit codes have been proposed by Daugman for iris recognition [31]. Here GGPP encodes orientation information at each scale whereas LGPP encodes the local neighborhood variations at each orientation and scale. Finally, both of the GPP's are combined with spatial histograms to model the original object image.

Gabor wavelet is well known algorithm for the face recognition. Conventionally, the magnitude of the Gabor coefficients are considered as valuable for face recognition and phase of the Gabor coefficients are considered useless and always discarded. But use of the spatial histograms, encodes the Gabor phases through Local binary Pattern (LBP) and provides the better recognition rate comparable with that of magnitude based methods. It shows that combination of Gabor phase and magnitudes provides the higher classification accuracy. These observation paid more attention towards the Gabor phases for face recognition.

So, Gabor Wavelet can be defined as [32]:

$$\Psi_{u,v}(Z) = \frac{\|k_{u,v}\|^2}{\sigma} e^{\left(\frac{-\|k_{u,v}\|^2 \|Z\|^2}{2\sigma^2}\right)} \left[e^{ik_{u,v}Z} - e^{\frac{-\sigma^2}{2}}\right] \qquad \text{--(22)}$$

Where $\overline{k_{u,v}} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \emptyset_u \\ k_v \sin \emptyset_u \end{pmatrix}$, $k_v = \frac{f_{max}}{2^{v/2}}$,

$\emptyset_u = u\left(\frac{\pi}{8}\right)$, v = 0, ...,v_{max} - 1,

$v = 0, \dots, u_{max} - 1, v\ is\ the\ frequency$ and

$u\ is\ the\ orientation\ with\ v_{max} = 5,$ and

$u_{max} = 8, \sigma = 2\pi.$

Here, in the R.H.S the term in the square bracket determines the oscillatory part of the kernel and the second term compensates for the magnitude of the DC value. σ determines the ratio of the Gaussian window width to the wavelength [9].

Now, the Gabor transformation of a given image can be defined as:

$$G_{u,v}(Z) = I(Z) * \Psi_{u,v}(Z) \qquad \text{… (23)}$$

$G_{u,v}(Z)$ is the convolution of corresponding to the Gabor kernel at scale v and orientation u. Again, the Gabor wavelet coefficient $G_{u,v}(Z)$ can be rewritten as a complex number.

$$G_{u,v}(z) = A_{u,v}(Z) . \exp\left(i\theta_{u,v}(Z)\right) \qquad \text{-- (24)}$$

Here,$A_{u,v}(Z)$ is the magnitude and $\theta_{u,v}(Z)$ is the phase of the Gabor wavelets. Magnitude varies slowly whereas phase varies with some rate with respect to spatial position. The rotation of the phases takes different values of the image but it represents

SEEK DIGITAL LIBRARY

almost the same value features. This causes severe problem in the face matching, that is the reason people used to make use of only the magnitude for face classification.

But Daugman's approach demodulated the Gabor phase with phase – quadrant demodulation coding. He used this coding for Iris recognition [31]. This coding assigns the each pixel into two bits$\left(P_{u,v}^{Re}(Z), P_{u,v}^{Im}(Z)\right)$. It is also known as quadrant bit coding (QBC). QBC is relatively stable. It actually quantifies the Gabor features.

$$P_{u,v}^{Re}(Z) = \begin{cases} 0, if\ Re\left(G_{u,v}(Z)\right) > 0 \\ 1, if\ Re\left(G_{u,v}(Z)\right) \leq 0 \end{cases} \quad \ldots (25)$$

$$P_{u,v}^{Im}(Z) = \begin{cases} 0, if\ Im\left(G_{u,v}(Z)\right) > 0 \\ 1, if\ Im\left(G_{u,v}(Z)\right) \leq 0 \end{cases} \quad \ldots (26)$$

Above these equations encoded by Daugman and named as Daugman's encoding method, are followed as:

$$P_{u,v}^{Re}(Z) = \begin{cases} 0, if\ Re\ \theta_{u,v}(Z) \in \{I, IV\} \\ 1, if\ Re\ \theta_{u,v}(Z) \in \{II, III\} \end{cases} \quad \ldots (27)$$

$$P_{u,v}^{Im}(Z) = \begin{cases} 0, if\ Im\ \theta_{u,v}(Z) \in \{I, II\} \\ 1, if\ Im\ \theta_{u,v}(Z) \in \{III, IV\} \end{cases} \quad \ldots (28)$$

$\theta_{u,v}(Z)$ defines the Gabor phase angle for the pixel at the spatial position Z. It transforms the same feature ("00") for the phase angle in $(0°, 90°)$ and so on.

From here, the GGPP algorithm computes one binary string for each pixel by concatenating the real or imaginary bit codes for different orientations for a given frequency at a given position. Now $GGPP_v(Z_0)$ formulates the values of GGPP at the frequency v and at the position (Z_0), which is shown as follows:

$$GGPP_v^{Re}(Z_0) = P_{0,v}^{Re}(Z_0), P_{1,v}^{Re}(Z_0), \ldots, P_{k,v}^{Re}(Z_0)] \quad \text{--(29)}$$

$$GGPP_v^{Im}(Z_0) = [P_{0,v}^{Im}(Z_0), P_{1,v}^{RIm}(Z_0), \ldots, P_{k,v}^{Im}(Z_0)] \quad \text{– (30)}$$

There are total eight orientations which can represent 0-255 different orientation modes.

Further, we can encode the local variations for each pixel, denoted as LGPP. This scheme encodes the sign difference of the central pixel from its neighbors. This shows the spots and flat area in the any given images. It can be computed using local XOR pattern or LXP operator. It can formulate as given below:

$$LGPP_{u,v}^{Re}(Z_0) = P_{u,v}^{Re}(Z_0)\ XOR\ P_{u,v}^{Re}(Z_1), P_{u,v}^{Re}(Z_0)\ XOR\ P_{u,v}^{Re}(Z_2) \ldots\ldots, P_{u,v}^{Re}(Z_0) XOR P_{u,v}^{Re}(Z_8)] \quad \ldots (31)$$

$$LGPP_{u,v}^{Im}(Z_0) = P_{u,v}^{Im}(Z_0)\ XOR\ P_{u,v}^{Im}(Z_1), P_{u,v}^{Im}(Z_0)\ XOR\ P_{u,v}^{Im}(Z_2) \ldots, P_{u,v}^{Im}(Z_0)\ XOR\ P_{u,v}^{Im}(Z_8)] \quad \text{--(32)}$$

Here $Z_1, Z_2, Z_3, \ldots, Z_8$ are the eight neighbors around Z_0 and XOR denotes the bit exclusive or operator.

Above process to encode the both GPP's provide 90 images (five real GGPP's, five imaginary GGPP's, 40 real LGPP's and 40 imaginary LGPP's) with the same size as the original face images. These images are in the form of micro – pattern and look like the images with rich structural textures. Histogram serves as a good description tool for above said micro – pattern and structural textures. In order to preserve the spatial information in the histogram features, both the GPP's are spatially subdivided into the non-overlapping rectangular region. Further spatial histogram can extract easily from non – overlapping rectangular regions. Then all of these histograms are concatenated into a single extended histogram features. It is also named as Joint local – histogram features (JLHF). It works on all frequencies and orientations.

The HGPP can be defined as:

SEEK DIGITAL LIBRARY

HGPP = $(H^{Re}_{GGPP}, H^{Im}_{GGPP}, H^{Re}_{LGPP}, H^{Im}_{LGPP})$ … (33)

Where H^{Re}_{GGPP} and H^{Im}_{GGPP} are the sub-region histograms of the real and imaginary part of GGPP whereas H^{Re}_{LGPP} and H^{Im}_{LGPP} are the sub region histograms of the real and imaginary part of LGPP. Both can formulate as given below:

H^{Re}_{GGPP}=

$(H^{Re}_{GGPP}(v,l): v = 0, \ldots, 4; l = 1, \ldots, L)$ --(34)

H^{Im}_{GGPP}=

$(H^{Im}_{GGPP}(v,l): v = 0, \ldots, 4; l = 1, \ldots, L)$ --(35)

H^{Re}_{LGPP}=

$(H^{Re}_{LGPP}(v,l): u = 0, \ldots 7;\ v = 0, \ldots, 4; l = 1, \ldots, L)$ …(36)

H^{Im}_{LGPP}=

$(H^{Im}_{LGPP}(v,l): u = 0, \ldots 7;\ v = 0, \ldots, 4; l = 1, \ldots, L)$ -- (37)

Where L is the number of sub-regions divided for the histogram computation.

VII. RESEARCH METHODOLOGY

We used ATT and IFD database for comparison of different face recognition algorithms such as PCA, LDA, ICA, SVM and HGPP. Based on algorithm, we extract different features from a training set. Using these feature we trained the classifier. We extract features from testing set and find the accuracy, training time, testing time, total execution time and total memory usage of the algorithm.

VIII. DATA ANALYSIS

We used ATT [36] and IFD [37] databases for training and testing different algorithms. We took 40 persons images from ATT and IFD database. 5 images of each person are used for training and 5 images of each person are used for testing algorithms. From Fig. 3 it is observed that all algorithms give better result on ATT database then IFD database. HGPP give best result on ATT database and LDA give best result on IFD database. A few images of both databases are shown below:

Fig. 1: Images of a subject From the ATT database

Fig. 2: Images of a subject from the IFD database

IX. EXPERIMENTAL RESULTS

Here, two face databases have been employed for comparison of performance. These are - 1. ATT face database [36] and 2. Indian face database (IFD) [37]. These two databases have been chosen because the ATT contains images with very small changes in orientation of images for each subject involved, whereas the IFD contains a set of 10 images for each subject where each image is oriented in a different angle compared to another.

CSU Face Identification Evaluation system [35] is used to provide the pre-processed databases which are converted to JPEG format and resizes them to smaller size to speed up computation.

The evaluation is carried out using the Face Recognition Evaluator. It is an open source MATLAB interface. Comparison is done on the basis of rate of recognition accuracy, testing time, training time, total execution time and total memory usage. Comparative results obtained by testing the five i.e. PCA, LDA, ICA, SVM and HGPP algorithms on both the IFD and the ATT databases.

B. COMPARATIVE RESULTS

The Comparative results have been organized by testing the above algorithms under the following criteria:

- Following result shows the comparisons of accuracy when the five algorithms applied on both datasets.

Accuracy (%)		
DATASET	ATT	IFD

SEEK DIGITAL LIBRARY

PCA	91.3	74.2
LDA	94.4	86.3
ICA	91.3	71.7
SVM	95.6	85.4
HGPP	76.25	43.25

Table 9.1 (a): Comparison of recognition accuracy.

- Following result shows the comparisons of training time when the five algorithms applied on both datasets.

Training Time		
DATSSET	ATT	IFD
PCA	0.2 (ms/image)	0.2(ms/image)
LDA	0.4(ms/image)	0.5(ms/image)
ICA	9.5(ms/image)	9(ms/image)
SVM	0.6(ms/image)	0.8(ms/image)
HGPP	9.32(s/image)	2.8(s/image)

Table 9.1 (b): Comparison of Training Time

- Following result shows the comparisons of testing time when the five algorithms applied on both datasets.

Testing Time		
DATASET	ATT	IFD
PCA	0.3(ms/image)	0.1(ms/image)
LDA	0.2(ms/image)	0.1(ms/image)
ICA	0.1(ms/image)	0.1(ms/image)
SVM	0.3(ms/image)	0.7(ms/image)
HGPP	19.39 (s/image)	10.0(s/image)

Table 9.1 (c): Comparison of Testing Time

- Following result shows the comparisons of total execution time when the five algorithms applied on both datasets.

Total Execution Time (ms/img)		
DATASET	ATT	IFD
PCA	0.5(ms/image)	0.4(ms/image)
LDA	0.6(ms/image)	0.5(ms/image)
ICA	9.7(ms/image)	9.1(ms/image)
SVM	0.9(ms/image)	1.5(ms/image)
HGPP	28.71(s/image)	10.18(s/image)

Table 9.1 (d): Comparison of Total Excution time

- Following result shows the comparisons of memory taken testing when the five algorithms applied on both datasets

Training Memory (MB)		
DATASET	ATT	IFD
PCA	1	1
LDA	1	1
ICA	1	1
SVM	1	1
HGPP	8.08	2.12

Table 9.1 (e): Comparison of Memory Usage for Testing

- Following result shows the comparisons of total memory taken to test, when the five algorithms applied on both datasets.

Testing Memory (MB)		
DATASET	ATT	IFD
PCA	1	1
LDA	1	1
ICA	1	1
SVM	1	1
HGPP	127.64	49.49

Figure 9.1 (f): Comparison of Memory Usage for test

- Following result shows the comparisons of total memory taken for training & Testing when the five algorithms applied on both datasets

Total Memory (MB)		
DATASET	ATT	IFD
PCA	2	2
LDA	2	2
ICA	2	2
SVM	2	2
HGPP	135.72	51.58

Figure 9.1 (f): Comparison of Total Memory Usage

X. PERFORMANCE ANALYSIS

Above analysis shows the performance of the five algorithms on the database of the ATT and IFD. Following points we have observed in this experiment.

- It is observed that recognition rate of the ATT database is higher as compare to IFD database.

SEEK DIGITAL LIBRARY

- It is observed that when algorithms employed on AT&T database, SVM has 95.6% rate of accuracy of recognition. LDA has 94.4% rate of accuracy of recognition, which outperforms the PCA, ICA and HGPP.
- It is observed that when five algorithms employed on IFD database then LDA outperform all remaining four algorithms. LDA has highest rate of accuracy of recognition i.e. 86.3%. Although LDA has the highest rate but it is marginally higher than SVM i.e. 85.4%. PCA and ICA the moderate rate of accuracy of recognition i.e. 74.2% and 71.7% respectively.
- It is observed that when five algorithms employed on ATT database and IFD then HGPP take the longest time to train the system with database that is 9.32 and 2.79 s / image respectively.
- It is observed that when five algorithms employed on ATT database and IFD then HGPP take the longest time to test the system with database that is 19.39 and 10.0 s / image respectively
- It is observed that despite HGPP remaining four algorithms takes very less time to train & test the data when it is employed both databases.
- It is observed that the total execution time taken by HGPP is the highest than other four methods i.e., 28.71 s / image (for ATT) and 10.18 s / image (for IFD).
- It is observed HGPP uses total memory 135.72 & 51.58 MB for the train & test the images of the ATT and IFD database respectively.

XI. CONCLUSION

In this paper, the comparisons of performance of five algorithms of the face recognition i.e. PCA, LDA, ICA, SVM and HGPP have been made. These algorithms have been employed on the two database AT&T and IFD.

It is concluded recognition rate of the ATT database is higher as compare to IFD database. This observation is due to the nature of images contain in the IFD. In this database, each subject is portrayed with highly varying orientation angles. It also shows that each image has rich background region than the ATT database. SVM has 95.6% rate of accuracy of recognition. LDA has 94.4% rate of accuracy of recognition, which outperforms the PCA, ICA and HGPP. HGPP is the lowest rate of accuracy of recognition i.e.43.25%. It shows that HGPP is effective but suffers from the local variations. The complex mathematically steps take more time to compute. Due to the size of images and complex mathematical steps in HGPP, uses total memory 135.72 & 51.58 MB for the train & test the images of the database. The image of ATT database has the dimensions of 92 x 112, while the image of the IFD database has the size of 64 x 48.

XII. FUTURE SCOPE

Lot of work can be done in field of face recognition such as most of the algorithms give good result on Frontal Face recognition but at different angles they do not give good result. To recognize a face at an angle we have to give some 3D face recognition algorithm. We can club other modality with face recognition algorithm for best results example face- iris, face-fingerprint, face-iris-fingerprint. Face recognition algorithm rate can be improved by first detecting the face from image and then crop the detected face and process it for recognition.

REFERENCE

[1] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, “Face Recognition by Independent Component Analysis”, IEEE Trans. on Neural Networks, Vol. 13, pp. 1450-1464, 2002.

[2] C. Liu and H. Wechsler, “Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition”, Second International Conference on Audio and Video-based Biometric Person Authentication, AVBPA’99,Washington D. C., USA, March 22-24, 1999.

[3] K. Baek, B. Draper, J.R. Beveridge, and K. She, “PCA vs. ICA: A Comparison on the FERET Data Set”, Proc. of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing, pp 824-827, 2002.

[4] B. Moghaddam, “Principal Manifolds and Probabilistic Subspaces for Visual Recognition”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, pp. 780-788, 2002.

[5] J.R. Beveridge, K. She, B. Draper, and G.H. Givens, “A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition”, Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535- 542, 2001.

[6] A.M. Martinez and A.C Kak, “PCA versus LDA”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, pp 228-233, 2001.

[7] V. Belhumeur, J. Hespanha, and D. Krieg man. “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19 pp 711-720, 1997.

[8] P. Navarrete and J. Ruiz-del-Solar, “Analysis and Comparison of Eigenspace-Based Face Recognition Approaches”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 16, pp. 817-830, 2002.

[9] Brain C. Becker and Enrique G. Ortig, “Evaluation of Face Recognition Techniques for Application to Facebook”, In Proceedings of the 8th IEEE International Automatice Face and Gesture Recognition Conference, 2009.

SEEK DIGITAL LIBRARY

[10] K. Delac, M. Grgic, S. Grgic, “Independent Comparative study of PCA, ICA and LDA on the FERET Data set”, Proc. of the 4th International Symposium on Image and Signal Processing and Analysis, pp 289-294, 2005.
[11] B.A. Draper, K.Baek, M.S. Bartlett and J.R Beveridge, “Recognizing Faces with PCA and ICA”, Computer Vision and Image Understanding, Vol. 91, pp 115-137, 2003
[12] L. Sirovich and M. Kirby, “Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A - Optics, Image Science and Vision”, 4(3):519–524, March 1987.

[13] M. Kirby and L. Sirovich, “Application of the karhunen-loeve procedure for the characterization of human faces”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1):103–108, 1990.
[14] M. Turk and A. Pentland, “Eigenfaces for recognition. Journal of Cognitive Neuroscience”, 3(1):71–86, 1991.
[15] G.Shakhnarovich and B.Moghaddam, “Face recognition in Subspaces”, Handbook of Face Recognition, Springer-Verlag, 2004.
[16] A. Selinger and D.A. Socolinsky, “Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A comparative Study”, Equinox Corporation, 2002.
[17] A. Nes, “Hybrid Systems for Face Recognition”, Master of Science Graduate Thesis, Faculty of Information Technology, Norwegian University of Science and Technology, 2003.
[18] K. Delac, M. Grgic and P. Liatis, “Appearance-based Statistical Methods for Face Recognition”, 47th International Symposium ELMAR, Zadar, Croatia, 8-10 June 2005.
[19] R. A. Fisher, “The Use of Multiple Measures in Taxonomic Problems”, Ann. Eugenics, Vol. 7, pp. 179-188, 1936.
[20] R. Brunelli and D. Falavigna, “Person identification using multiple cues”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, pp 955-966, 1995.
[21] P.Common, “Independent components analysis – A new concept?”, Signal Processing, Vol. 36, pp. 287- 314,1994.
[22] C. Jutten and J. Herault, “Blind separation of sources in an adaptive algorithm based on neuromimetic architecture”, Signal Processing, Vol. 24, pp. 1-10, 1991.
[23] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution”, Neural Comput., vol. 7, no. 6, pp. 1129–1159, 1995.
[24] A. Cichocki, R. Unbehauen, and E. Rummert, “Robust learning algorithm for blind separation of signals”, Electron. Letter, Vol., 30, no. 7, pp.1386–1387, 1994.
[25] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” in Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 1996, vol.8.
[26] ----, “The independent components of natural scenes are edge filters”, Vision Res., vol. 37, no. 23, pp. 3327–3338, 1997.
[27] D. J. C. MacKay, “Maximum Likelihood and Covariant Algorithms for Independent Component Analysis”, 1996.
[28] B. A. Pearlmutter and L. C. Parra, “A context-sensitive generalization of ICA,” in Advances in Neural Information Processing Systems”, Mozer, Jordan, and Petsche, Eds. Cambridge, MA: MIT Press, 1996, vol. 9.
[29] Vapnik N., “The Nature of Statistical Learning Theory”, Springer, 1995.

[30] Guo G., Li S. Z, and Chan K, “Face Recognition by Support Vector Machines”, Image and Vision Computing, vol. 19, pp. 631-638, 2001.
[31] Daugman J. G.. “High confidence visual recognition of persons by a test of Statistical Independence”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 11, pp. 1148– 1161, Nov. 1993.

[32] Liu C. and Wechsler H., “Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition,” IEEE Trans. Image Process., vol. 11, no. 4, pp. 467–476, Apr. 2002.
[33] Baochang Zhang, Shiguang Shan, Xilin Chen and Wen Gao, “Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation Approach for Face recognition” IEEE Trans. On Image Processing,Vol. 16, NO. 1, January 2007.
[34] C.C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," 2001.
[35] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A.Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure," in *ICVS* Graz, Austria, 2003.
[36] www.cl.cam.ac.uk/research/dtg/attarchive/facebook.html
[37] www.cs.umass.edu/~vidit/IndianFaceDatabase

SEEK DIGITAL LIBRARY