

Illuminant Invariant Face Recognition by the Fusion of Incremental and Artificial Neural Network Approaches

¹Linda Sara Mathew, ²A.C.Ramesh

¹M.A. College of Engineering, Kothamangalam, Tamilnadu, India

²PSG College of Technology, Coimbatore, Tamilnadu, India

Abstract

Dimension reduction is critical for face recognition applications as the data dimensionality is much higher than the size of the training set, leading to the small sample size problem. The issue of singular scatter matrices has been mitigated by GSVD and Incremental approach overcomes the scalability problem. In this paper, an incremental algorithm ILDA-GSVD has been used for generating the projection matrix and neural network classifier discriminates the face images. This method consists of preprocessing, dimension reduction, feature extraction and classification using neural network. Several methods are adopted to standardize the faces illumination reducing the variations for further features extraction; which are extracted using the image phase spectrum of the histogram equalized image. The incremental approach proved efficient compared to the batch method both in terms of misclassification and time even when a few samples of images are available and neural classifier further reduced the misclassification caused by not-linearly separable classes. The proposed method was tested on both feret and orl face databases. Evaluation results show that the proposed feature extraction scheme, when used together with the neural network classifier, provides a recognition rate of 92% and a verification error lower than 0.04%.

Keywords

Face Recognition, LDA, Scatter matrix, GSVD, ILDA, Multilayer Perceptron, Backpropagation Neural Network

I. Motivation

Pattern recognition is a fast developing field of computer technology which finds much prominence. It has many applications in various fields where individual's identity profiling is critical. Among many recognition subjects, face recognition has drawn considerable interest and attention from many researchers for the last two decades because of its potential applications, such as in the areas of surveillance, secure trading terminals, Closed Circuit Television (CCTV) control, user authentication, HCI, intelligent robot and so on. Recently in many countries face recognition tools were applied to prevent electoral fraud, border crossing etc. Some potential applications of face recognition include its use as a security measure at ATMs. The same concept can be applied to computers so that computer security can be ensured. Since these databases are typically very large, it is impossible for any user to memorize every face. Therefore, it is necessary to have machines aid users in the recognition process [1]. Another motivation for research in the face recognition field is the authentication problem. The demand for reliable personal identification in computerized access control has resulted in an increased interest in biometrics to replace password and Identification (ID) card. The password and ID card can be easily breached since the password can be divulged to an unauthorized user, and the ID card can be stolen by an impostor. Thus, the emergence of biometrics has addressed the problems that plague the traditional verification methods. Biometric which make use of human features such as iris, retina,

face, fingerprint, signature dynamics, and speech can be used to verify a person's identity. The biometrics data have an edge over traditional security methods since they cannot be easily stolen or shared. The face recognition system has the benefit of being a passive, non-intrusive system for verifying personal identity.

II. Introduction

The innate ability of human beings to recognize different patterns is remarkable but from the last few years, it is the machines that have showed tremendous improvement. The computer vision research of last 30 years has focused on building such machines that can recognize human beings. The face gives a unique biometric identity for individual existence of every human being. This biometric feature helps to discriminate between different human beings otherwise everyone looks the same. The distinguishable landmarks (nodal points/facial features) are very efficiently utilized for both detection and recognition tasks.

The literature review shows that face recognition has been done using neural networks, PCA (Principal Component Analysis), LDA, distance matrix approach of ARENA algorithm for neural network training [2] and using 3D geometry. Researchers have moved from 2D recognition to 3D recognition but there are still some issues in 2D recognition that needs further investigation such as dimension reduction approaches, robust classification techniques, consistency in accuracy and time factor issues. Neural networks have shown good results in recognition scenarios but still we believe there remains a room for improvement. The novel scheme of fusing neural network response with incremental algorithm can be applied on various applications to enhance the system accuracy.

Face recognition techniques use face representations found by both supervised and unsupervised statistical methods. Unsupervised methods find a set of basis images and represent faces as a linear combination of those images. To increase the speed of the algorithm we use incremental approach.

III. Proposed Methodology

The proposed face recognition method consists of different parts:

- Pre processing
- Dimensionality reduction
- Feature Extraction using LDA/GSVD and ILDA-GSVD
- Classification using neural network

IV. Preprocessing

Histogram is equalized by stretching and/or redistributing the original histogram over the entire range of discrete levels of the image, in a way that an enhancement of image contrast is achieved [3]. The resultant image histogram is constant for all brightness values. This corresponds to a brightness distribution where all values are equally probable.

For image $I(x,y)$ with discrete k gray values histogram is defined by:

$$p(i) = \frac{n_i}{N} \quad (1)$$

where $i, \in 0, 1, \dots, k-1$ grey level and N is total number of pixels in the image.

Transformation to a new intensity value is defined by:

$$i_{out} = \sum_{i=0}^{k-1} p(i) \quad (2)$$

Output values are from domain of $[0,1]$. To obtain pixel values into original domain, it must be rescaled by the $k-1$ value.

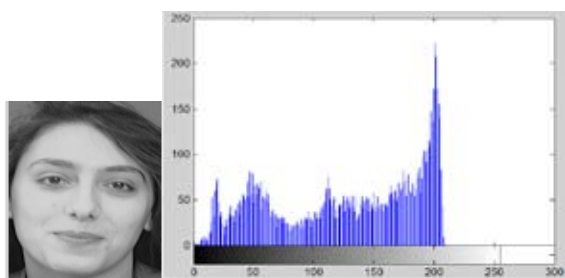


Fig. 1: Input Image and Its Histogram

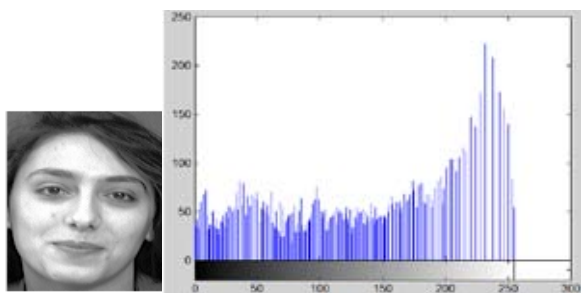


Fig. 2: Histogram Equalized Image and Its Histogram

Filtering in images is done for noise reduction, image enhancement and feature extraction. The pixels in an image is modified based on some function of a local neighborhood of the pixels. Convolution kernel gives a prescription for the linear combination [4]. Linear filtering has its roots in Fourier transform in the frequency domain. Each pixel in the neighbourhood is multiplied by a corresponding coefficient and the results are summed to obtain the response at each point. The process consists of moving the center of the filter mask from point to point in an image. At each point the response of the filter is the sum of products of the filter coefficients and the corresponding neighborhood pixels in the area spanned by the filter mask. The principal focus is on masks of odd sizes because of the unambiguous center point.

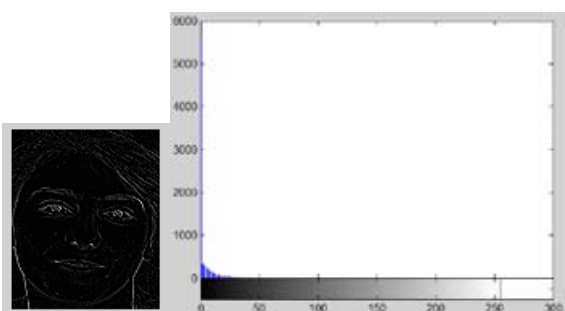


Fig. 3: Laplacian Filtered Image

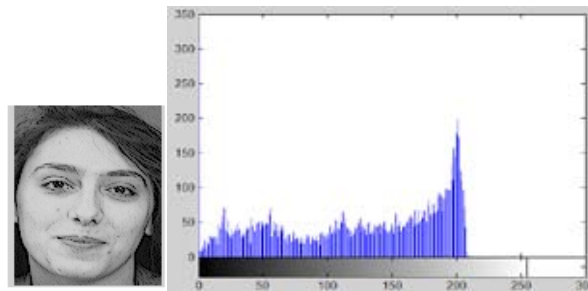


Fig. 4: Enhanced Image

Digital images are prone to various types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. There are several ways that noise can be introduced into an image, depending on how the image is created. If the image is scanned from a photograph made on film, the film grain is a source of noise. Noise can also be the result of damage to the film, or be introduced by the scanner itself. If the image is acquired directly in a digital format, the mechanism for gathering the data can introduce noise. Electronic transmission of image data can also introduce noise.

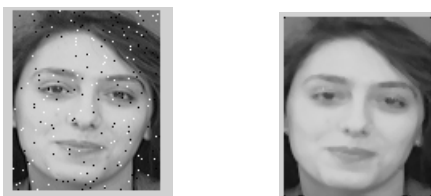


Fig. 5: Salt & Pepper Noise Added and Removed

V. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classical statistical approach for feature Extraction and dimension reduction [5]. LDA computes the optimal transformation (projection), which minimizes the within-class distance (of the data set) and maximizes the between-class distance simultaneously, thus achieving maximum discrimination [6]. In face recognition systems class represents person, but the class consists of many images per face. The optimal transformation can be readily computed by applying an eigen-decomposition on the scatter matrices of the given training data set. However classical LDA requires the total scatter matrix to be nonsingular. In many applications such as information retrieval, face recognition, and microarray data analysis, scatter matrices can be singular since the data points are from a very high-dimensional space and in general the sample size does not exceed this dimension. This is known as the singularity or under sampled problems [7].

In recent years, many approaches have been brought to bear on such high-dimensional, under sampled problems, including PCA+LDA, Regularized LDA, Pseudo inverse LDA and LDA/GSVD [7]. To deal with the singularity problem encountered by classical LDA there are three major extensions: regularized LDA, PCA+LDA, and LDA/GSVD. The common point of these algorithms is the use of Singular Value Decomposition (SVD) or Generalized Singular Value Decomposition (GSVD) Regularized.

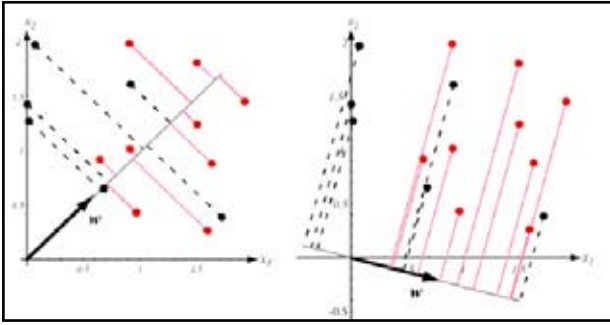


Fig. 6: Projection of the Same set of Data onto Two Different Lines

Let the between-class scatter matrix be

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (3)$$

and the within-class scatter matrix be

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (4)$$

Where, μ_i is the mean image of class X_i , and N_i is the number of samples in class X_i . If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples [8].

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} \quad (5)$$

VI. LDA/GSVD

A common limitation of trace optimization in discriminant analysis is that one of the scatter matrices must be nonsingular, which restricts its application to document sets in which the number of terms does not exceed the number of documents. By using the generalized singular value decomposition (GSVD), this problem can be mitigated [9].

Algorithm:

- 1) Compute H_W and H_B
- 2) Let $t = \text{rank}(H)$
- 3) Compute SVD on $H = [H_B^T; H_W^T]^T \in \mathbb{R}^{(c+p) \times n}$ as
 $H = P(R \ 0) Q^T$
 Where $R = \text{diag}(\xi_1, \dots, \xi_t)$, $\xi_i > 0 (i=1, \dots, t)$
- 4) Compute SVD on $P(1:c, 1:t) = UV^T$
- 5) Let $T = \text{rank}(H_B)$
- 6) Compute the first T columns of
 $Y = Q \cdot (R^{-1} \cdot W)$

VII. Incremental Approaches

In batch computation all observations are used simultaneously to compute the model. In an incremental computation, an existing model is updated using new observations. Most previous work on PCA and LDA requires that all the training data be available before the dimension reduction step which is known as the batch method [10].

Linear Discriminant Analysis (LDA), also known as Fisher Discriminant Analysis-FDA, seeks directions for efficient discrimination, while principal component analysis (PCA), seeks directions efficient for representation. The typical implementation

of these two techniques assumes that a complete dataset for training is given in advance, and learning is carried out in one batch [11]. However, when we conduct LDA/PCA learning over datasets in real-world application, we often confront difficult situation where a complete set of training samples is not given in advance. A straightforward approach is that, we can collect data whenever new data is presented, and then construct a provisional system by batch learning over the collected data so far [12]. However, it is obvious that such system only works under a condition of a large memory and high computation expenses, because the system would need to maintain a huge memory to store the data either previously learnt or newly presented, possibly without a limit. Moreover, the system has to discard the knowledge acquired in the past, even if the learning of 99.9% data is finished, and repeat the learning from the beginning whenever one additional sample is presented. Obviously, one-pass incremental learning gives a solution to the above problem. In this learning scheme, a system must acquire knowledge with a single presentation of the training data, and retaining the knowledge acquired in the past without keeping a large number of training samples. Incremental PCA (IPCA) is based on the updating of covariance matrix through a residue estimating procedure [13]. The discriminant eigenspace is updated for classification, when data that contains new or existing classes is being added to an initial discriminant eigenspace.

Since it is difficult to design an incremental solution for the eigenvalues problem on the product of scatter matrices, little effort has been made to design incremental LDA algorithms. Based on the generalized singular value decomposition LDA (LDA/GSVD), a new ILDA algorithm is called GSVD-ILDA. Different from the existing techniques in which the new projection matrix is found in a restricted subspace, GSVD-ILDA determines the projection matrix in full space [14]. The GSVD-ILDA algorithm has two advantages: (1). The algorithm can process the samples in chunks or in sequence, which is desirable for large image databases, and (2). With the dynamically adding samples, the algorithm can constrain the computational cost.

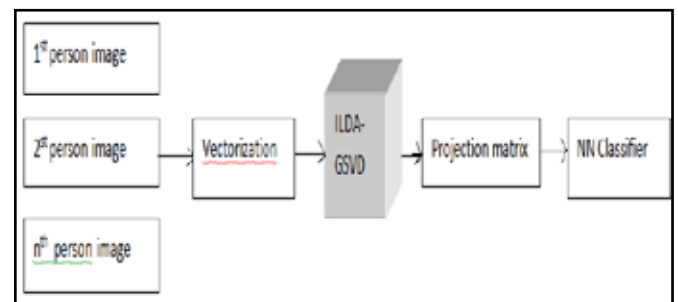


Fig. 7: GSVD-ILDA Block Diagram

GSVD-ILDA algorithm:

Input: the data matrix X , the number of classes c , the number of samples in the j^{th} class $p_j (j = 1, \dots, c)$, the total number of samples $p = \sum p_j$ and the newly added data matrix X^s

Output: the updated mean of the i^{th} class, the updated global mean, the updated number of samples in the j^{th} class j , and the updated projection matrix Φ .

- 1) For the data matrix X , let c be the number of classes, p_j be the number of samples in the j^{th} class ($j = 1, \dots, c$), and p be the total number of samples.
- 2) Compute the mean m_j of the j^{th} class ($j = 1, \dots, c$) and the global mean m .
- 3) Compute the matrix $H_i = X - m e^T$

- 4) Compute the SVD of the matrix H_t as $H_t = Q\Xi Z^T$.
- 5) Let Q_k be the first k columns of Q and Ξ_k be the leading principal submatrix of Ξ .
- 6) Let $s = 1$.
- 7) For the newly added data matrix $X_s = [x_1, \dots, x_{q_s}]$, let c be the class number of the data $[X, X_s]$.
- 8) Compute the number of samples p_i' and the sample mean m_i' of the newly added samples of class i ($i=1, \dots, c$), and then, compute the mean m' of the data X_s .
- 9) Compute $p_i = p_i + p_i'$ and $m_i = (p_i m_i + p_i' m_i')/p_i$, where p_i' and m_i' are the number of samples and the sample mean of the newly added samples of class i ($i=1, \dots, c'$), respectively.
- 10) Compute the matrix $X = [X' - m\epsilon^T (pq/p+q) (m - m')]$.
- 11) Compute the QR decomposition $(I - Q_k Q_k^T)X = JK$.
- 12) Compute the SVD of the smaller matrix
- 13) Obtain Q as $Q = ([Q_k, J]Q)$
- 14) Let Q_k and Ξ_k be the rank- k approximation of the matrix Q and Ξ .
- 15) Compute H_b
- 16) Let $t = \text{rank}(\Xi)$ and $\tau = \text{rank}(H_b)$.
- 17) Compute $P(1:c, 1:t) = H_b^T Q(:, 1:t) (\Xi(1:t, 1:t))^{-1}$.
- 18) Compute the SVD on $P(1:c, 1:t)$, which is $P(1:c, 1:t) = UYW^T$.
- 19) Compute the first τ columns of $Y = Q[\Xi(1:t, 1:t) - I]W$ and assign them to Φ .
- 20) Let $s = s + 1$. If $s \leq S$
- 21) Let $X = [X, X_{s-1}]$ go to Step 7. else
- 23) End.

VIII. Feedforward Neural Network

The architecture of neural network consists of three layers – input, hidden and output layers. The linear inseparability can be resolved using FFNN. The layered structure gets input from the previous layers.

Backpropagation is a kind of the gradient descent method, which search an acceptable local minimum in the NN weight space in order to achieve minimal error. Error is defined as a root mean square of differences between real and desired outputs of NN [15].

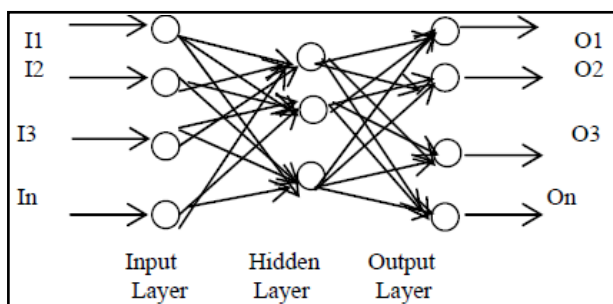


Fig. 8: FFNN Architecture

Pre-processed images in the vector format are feed into the input layer which contains 10304 neurons, hidden layer contains 100 neurons, the number of neurons in hidden layer comes from experience and guess work considering optimal performance. Finally, output layer contains the neurons equal to the number of subjects under consideration. The gradient of the error is minimized through adjusting weights and biases continuously with momentum. Momentum acts like a low pass filter and ignores small features in error surface so that network do not stuck into a local minimum. Neural Network is trained upon some set of images, and tested upon unseen images. In proposed technique neural

network uses back propagation algorithm for error computation and new weights calculation for each link. The network undergoes the process of training, continuously in an iterative manner, it calculates output from each layer, extracting the mean square error and propagating it backwards and check if it is approaching targets. Error signal for each neuron is calculated which in fact is used for neuron weight updation. If it is approaching targets then training is considered done. It has been observed that as the number of subjects increases, the training time also increases (as the complexity of the input increases with the increase in the number of classes). In this way network approaches the known correct outputs (targets) in order to be trained [16].

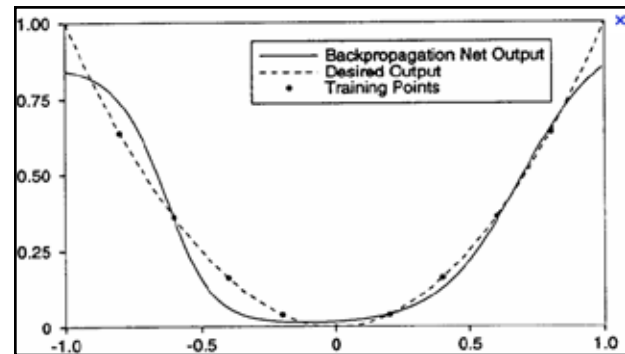


Fig. 9: Backpropagation Net Output

The change in weights between hidden layer and input layer, is given by

$$\Delta w_{ij} = \alpha \delta_i h_j \quad (7)$$

where α , is the learning rate coefficient and

$$\delta_i = (t_i - o_i) o_i (1 - o_i) \quad (8)$$

The change in weights between hidden layer and input layer, is given by

$$\Delta w_{ij} = \beta \delta_{Hi} x_j \quad (9)$$

where β , is the learning rate coefficient and

$$\delta_{Hi} = x_i (1 - x_i) \sum_{j=1}^k \delta_j w_{ij} \quad (10)$$

IX. Face Database

The publicly available ORL database consists of ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). These images were taken over a period of a few years, so one person can either have only images taken on the same day or images taken with longer time in between. The images can be manually centered and cropped to different scales. The database is distributed free to facilitate research in face recognition.



Fig. 10: Sample Faces from ORL Database

X. Results

The present study indicates that the dimensionality reduction efficiency varies with the algorithms. LDA-GSVD and ILDA-GSVD algorithms are implemented in ORL and FERET databases. The results indicate that the incremental approach GSVD-ILDA perform faster than its batch method. Both algorithms perform better in FERET database than in ORL database.

Table 1: Recognition Rate as Number of Classes Varied

No: of classes	Recognition rate
15	92%
30	90%
40	89%

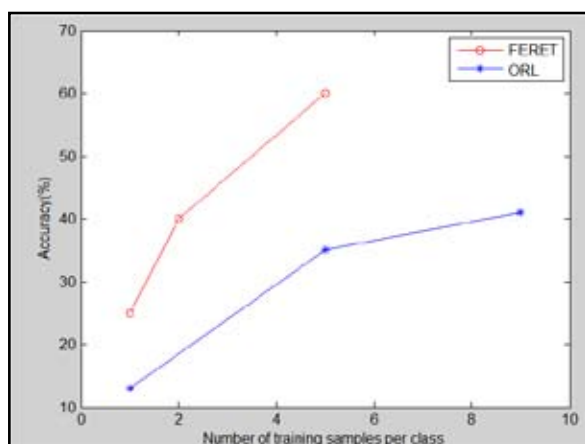


Fig. 11: Comparison of Accuracy

Many metrics have been adopted to evaluate algorithms, such as learning time, execution time, the number of samples required in training, and the ratio between detection rates and false alarms. In general, detectors can make two types of errors: false negatives in which faces are missed resulting in low detection rates and false

positives in which an image is declared to be face.

$$\text{False negative} = \frac{\text{Number of Missed Faces}}{\text{Total Number of Actual Faces}}$$

$$\text{False positive} = \frac{\text{Number of Incorrectly detected Faces}}{\text{Total Number of Actual Faces}}$$

The following table and figures show the accuracy and output error at each epoch for the data set.

Table 2: Accuracy in Different Epochs

Epoch	Accuracy	Output Error
10	18	0.40
20	20	0.39
30	38	0.28
40	53	0.098
50	60	0.096
60	70	0.075
70	78	0.067
80	88	0.059
90	90	0.046
100	92	0.035

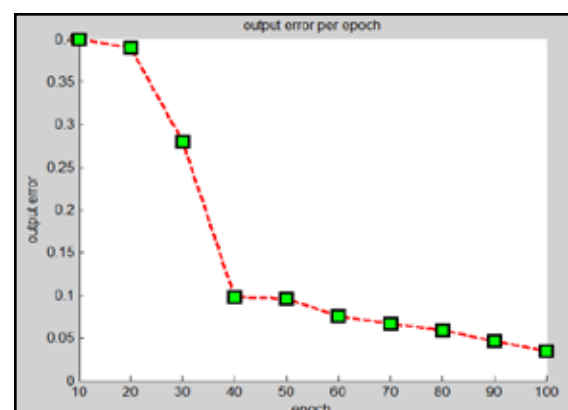


Fig. 12: Output Error Per Epoch

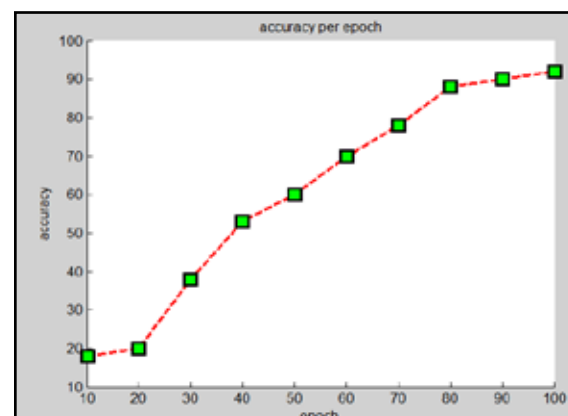


Fig. 13: Accuracy Per Epoch

XI. Conclusion and Future Work

Face recognition still remains as a challenging problem and a lot of work needs to be done in this area. The proposed algorithm demonstrated how a face recognition system can be designed by the combination of an incremental approach and neural network. The very high dimensionality of the data is a big issue but the use

of an incremental approach, GSVD-ILDA generated a projection matrix with the discriminatory information of the whole classes. ILDA effectively synthesizes three useful improvements on the current linear discrimination technique: it improves the selection of discrimination vectors, adds a measure so that the discrimination vectors satisfy the statistical correlation using less computing time, and provides a strategy to select the principal components. Future work shall be focused on investigating the best rank – k approximation of incremental LDA. The Back Propagation neural network used as a classifier in the final stage improves the accuracy of the method. Experiments on different standard benchmark databases have been done.

References

- [1] Zhao, W., Chellappa, R., Rosenfeld, A.; Phillips, P., "Face recognition: A literature survey", ACM Comput. Surv., Vol. 35, No. 4, 2003, pp. 399–458.
- [2] Fukunaga, K., "Introduction to Statistical Pattern Recognition", New York: Academic, USA, 1990.
- [3] Ye, J., Li, Q., "LDA/QR: An efficient and effective dimension reduction algorithm and its theoretical foundation", Pattern Recognit., Vol. 37, No. 4, 2004, pp. 851–854.
- [4] Ye, J., Janardan, R., Park, C. H., Park, H., "An optimization criterion for generalized discriminant analysis on undersampled problems", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 26, No. 8, 2004, pp. 982–994.
- [5] Martinez, A. M.; Kak, A. C., "PCA versus LDA", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 23, No. 2, 2001, pp. 228–233.
- [6] Ye, J., "Characterization of a family of algorithms for generalized discriminant analysis on under sampled problems", J. Mach. Learn. Res., Vol. 6, 2005, pp. 483–502.
- [7] Ruiz-del-Solar, J., Navarrete, P., "Eigenspace-based face recognition: A comparative study of different approaches", IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., Vol. 35, No. 3, 2005, pp. 315–325, 2005.
- [8] Haitao Zhao, Pong Chi Yuen., "Incremental Linear Discriminant Analysis for Face Recognition", IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, Vol. 38, No. 1, 2008.
- [9] Jing, X., Zhang, D., Tang Y., "Improved LDA approach", IEEE Trans. Syst., Man, Cybern. B, Cybern., Vol. 34, No. 5, 2004, pp. 1942–195.
- [10] Ye, J., Li, Q., Xiong, H., Park, H., Janardan, R.; Kumar, V., "IDR/QR: An incremental dimension reduction algorithm via QR decomposition", in Proc. 10th ACM SIGKDD Int. Conf. KDD, 2004, pp. 364–373.
- [11] Pang, S., Ozawa, S., Kasabov, N., "Incremental linear discriminant analysis for classification of data streams", IEEE Trans. Syst., Man, Cybern. B, Cybern., Vol. 35, No. 5, 2005, pp. 905–914.
- [12] Chatterjee, C., Roychowdhury, V. P., "On self-organizing algorithms and networks for class-separability features", IEEE Trans. Neural Netw., Vol. 8, No. 3, 1997, pp. 663–678.
- [13] Chen, L., Liao, H., Ko, M., Lin, J., Yu, G., "A new LDA-based face recognition system which can solve the small sample size problem", Pattern Recognition., Vol. 33, No. 10, 2000, pp. 1713–1726.

- [14] J. Ye, R. Janardan, C.H. Park, H. Park., "An optimization criterion for generalized discriminant analysis on undersampled problems", IEEE Trans. Pattern Analysis and Machine Intelligence, 26(8), pp. 982-994, 2004.
- [15] J. Mao, K. Jain., "Artificial neural networks for feature extraction and multivariate data projection", IEEE Trans. Neural Networks, 6(2), pp. 296-317, 1995.
- [16] Eleyan, A., Demirel, H., "Face Recognition System based on PCA and Feedforward Neural Networks", Proceedings of Computational Intelligence and Bioinspired Systems, pp. 935-942, Spain, June 2005, Springer-Verlag, Barcelona.



Linda Sara Mathew received her B.Tech degree in Computer Science and Engineering from Mar Athanasius College of Engineering, Kothamangalam, Kerala in 2002 and ME degree in Computer Science and Engineering from PSG College of Technology, Coimbatore in 2011. She is currently, working as Assistant Professor, with Department of Computer Science and Engineering in Mar Athanasius College of Engineering, Kothamangalam and has a teaching experience of 8 years. Her area of interests include digital signal processing, Image Processing and Soft Computing.



A.C Ramesh received his BE degree from Noorul Islam College of Engineering, Kumaracoil, Tamil Nadu in the year 2000, ME degree from National Engineering College, Kovilpatti, Tamil Nadu in the year 2002 and is currently pursuing Ph.D. He is currently working as Assistant Professor, with Department of Computer Science and Engineering in PSG College of Technology, Coimbatore and has a teaching experience of 8 years. His area of interests include Digital Image Processing, Artificial Neural Networks and Genetic Algorithms and Biometrics.