Modified Incremental Linear Discriminant Analysis for Face Recognition

R. K. Agrawal[1] and **Ashish Chaudhary**[2]

Abstract - ***Linear Discriminant analysis is a commonly used and valuable approach for feature extraction in face recognition. In this paper, we have proposed and investigated modified incremental Linear Discriminant Analysis (MILDA). We have compared the performance of proposed MILDA method against Pang et al ILDA in terms of classification accuracy, execution time and memory. It is found on the basis of experimental results with different face datasets that the proposed MILDA scheme is computationally efficient in terms of time and memory in comparison to batch method and Pang et al method. The experimental results also show that the classification accuracy due to MILDA, batch method and Pang et al are in complete agreement with each other.***

Index Terms - Statistical pattern recognition, Feature extraction, Face ecognition, Linear Discriminant Analysis

1. INTRODUCTION

Feature extraction is one of the important steps in pattern recognition which utilizes all the information of the given data to yield feature vector of the lower dimension and thereby eliminates redundant and irrelevant information. Commonly used feature extraction techniques are Principal Component Analysis (PCA) [1], Linear Discriminant Analysis LDA [2-7], Independent Components Analysis ICA [16] etc. Generally while applying these techniques for data classification it is assumed that complete dataset for training is available in advance and, learning is carried out in one batch. However in many real world applications such as pattern recognition and time series prediction we frequently come across situations where complete set of training samples is not available in advance, instead existing dataset keeps on changing with time. For example, in face recognition process human face undergoes facial variation due to different expressions (sad, happy, laughing face etc), lighting conditions, and make up, hairstyles etc. Hence, it is difficult to consider all facial variation when a human face is registered in a face recognition system, first time [8]. Similarly, in intrusion detection system, it is desirable to study the pattern behavior of intruder which can change slightly from its original behavior on account of incremental changes in data set [13]. Hence it is difficult to extract meaningful features only from previously available dataset. A straightforward approach in this situation is that we can collect data whenever new data are presented and then construct a provisional system by batch learning [9] over the collected data so far. However, such system will require large memory and high computational cost because the system would need to maintain a huge memory to store the data either previously learned, or newly presented, possibly without a limit. Moreover, the system will not be able to utilize the knowledge acquired in the past, even if the learning of most of the data is finished, and will repeat the learning from the beginning whenever one additional sample is presented.

[1,2]*School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi – 110067*
E-Mail: [1]rka@mail.jnu.ac.in, [2]ashish01jnu@gmail.com

To address such situation several eigenspace model [10–12] have been proposed. Hall, Marshall and Martin [10] proposed Incremental PCA (IPCA) based on the updating of covariance matrix through a residue estimating procedure. Agrawal and Karmeshu [11] proposed perturbation scheme for online learning of features based on incremental principal component analysis. Recently Pang, Ozawa and Kasabov [9] proposed an incremental linear discriminant analysis (ILDA). In this paper they have updated within class scatter matrix as new samples is added in terms of previously computed scatter matrix. This reduces the cost of computing updated scatter matrix. However the inverse of updated matrix $S_W{}^*$ carried out for the computation of transformation matrix does not employ the previous knowledge of inverse of S_W. In this paper we investigate computationally more efficient scheme to compute the inverse of scatter matrix $S_W{}^*$ which allows computation of dominant eigenvalues and eigenvector much more efficiently thereby increasing the performance of face recognition system.

The paper is organized as follows. Section 2 first provides a brief introduction of sequential ILDA proposed by Pang, Ozawa and Kasabov [9]. Following this, a modified incremental Linear Discriminant Analysis (MILDA) method is proposed in section 3. The performance of the proposed MILDA method in relation to batch method and Pang et al method is examined in terms of discriminability, computational time and memory in section 4. For this we have considered three publicly available face datasets [15]. The last section 5 contains conclusions.

2. SEQUENTIAL INCREMENTAL LDA

Suppose that initially we have a set of $\mathbf{N}$ $\mathbf{d}-$dimensional samples $\mathbf{x_1}, \mathbf{x_2}, \ldots\ldots\ldots \mathbf{x}_N$ belonging to $\mathbf{c}$ different classes with $\mathbf{N_i}$ samples in the $\mathbf{i^{th}}$ class. Then, the objective of LDA is to seek the direction $\mathbf{w}$ which maximizes the between-class scatter and minimizes the within class scatter of the projected images, such that the following criterion function [1]:

$$\mathbf{J(w)} = \frac{\mathbf{w^T S_B w}}{\mathbf{w^T S_W w}} \qquad (1)$$

is maximized, where S_B and S_W are between-class scatter and within class scatter matrices and are defined as

$$\mathbf{S_B} = \sum_{1}^{c} \mathbf{N_i}(\mathbf{\bar{x}_i} - \mathbf{\bar{x}})(\mathbf{\bar{x}_i} - \mathbf{\bar{x}})^{\mathbf{T}} \quad (2)$$

Where $\mathbf{\bar{x}}$ is d- dimensional sample mean of all the images and defined by

$$\mathbf{\bar{x}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{1}^{n} \mathbf{x_k} \quad (3)$$

and $\mathbf{\bar{x}_i}$ is i^{th} class mean given by

$$\mathbf{\bar{x}_i} = \frac{\mathbf{1}}{\mathbf{N_i}} \sum_{\mathbf{x_i} \in \mathbf{C_i}} \mathbf{x_i} \quad (4)$$

Within-class scatter matrix $\mathbf{S_W}$ is defined as

$$\mathbf{S_W} = \sum_{1}^{c} \mathbf{S_i} \quad (5)$$

Where $\mathbf{S_i}$ is i^{th} -class scatter matrix

$$\mathbf{S_i} = \sum_{\mathbf{x} \in \mathbf{c_i}} (\mathbf{x} - \mathbf{\bar{x}_i})(\mathbf{x} - \mathbf{\bar{x}_i})^{\mathbf{T}} \quad (6)$$

To find the transformation matrix W, a generalized eigenvalue problem needs to be solved which is given by

$$\mathbf{S_B w} = \lambda \mathbf{S_w w} \quad (7)$$

If S_w is non-singular then we have

$$\mathbf{S_W}^{-1}\mathbf{S_B w} = \lambda \mathbf{w} \quad (8)$$

Hence in case of LDA the transformation matrix $\mathbf{W}$ is represented in terms of eigenvectors of matrix $\mathbf{U} = \mathbf{S_w}^{-1}\mathbf{S_B}$. It is obvious that the parameter needed for classification at any point of time are $\mathbf{S_W}, \mathbf{S_B}, \mathbf{\bar{x}}$ and $\mathbf{N}$. So discriminant eigenspace can be represented as $\Omega = (\mathbf{S_W}, \mathbf{S_B}, \mathbf{\bar{x}}, \mathbf{N})$.

The traditional LDA works in a batch mode assuming that the whole dataset is given in advance and is trained in one batch only [9]. However, in a streaming environment, addition of any new samples will result in changes in original mean vector $\mathbf{\bar{x}}$, within class scatter matrix $\mathbf{S_W}$, as well as between-class distance matrix $\mathbf{S_B}$. Hence, the discriminant eigenspace model Ω needed to be updated. Pang, Ozawa and Kasabov [9] suggested incremental linear discriminant analysis (ILDA) for updating discriminant eigenspace Ω.

Pang et al [9] proposed that as a new sample $\mathbf{y}$ belonging to $\mathbf{k^{th}}$ class is added to existing samples with mean $\mathbf{\bar{x}}$, within class scatter matrix S_W and the between class scatter matrix $\mathbf{S_B}$ then the new mean vector, the new between scatter matrix and the new within class scatter matrix are respectively given by $\mathbf{\bar{x}}^*$, S_W^* and $\mathbf{S_B^*}$ i.e.

$$\mathbf{\bar{x}}^* = \frac{\mathbf{N\bar{x}} + \mathbf{y}}{\mathbf{N} + \mathbf{1}} \quad (9)$$

If $\mathbf{k} = \mathbf{c} + \mathbf{1}$ i.e. the incoming sample belongs to a new class, then updated between-class scatter will be

$$\mathbf{S_B^*} = \sum_{\mathbf{i=1}}^{\mathbf{c+1}} \mathbf{N_i}^*(\mathbf{\bar{x}_i}^* - \mathbf{\bar{x}}^*)(\mathbf{\bar{x}_i}^* - \mathbf{\bar{x}}^*)^{\mathbf{T}} \quad (10)$$

Where N_i^* is the number of samples in class i after addition of y. If $\mathbf{1} \le \mathbf{i} \le \mathbf{c}$ then the updated matrix $\mathbf{S_B^*}$ is given by

$$\mathbf{S_B^*} = \sum_{\mathbf{i=1}}^{\mathbf{c}} \mathbf{N_i}^*(\mathbf{\bar{x}_i}^* - \mathbf{\bar{x}}^*)(\mathbf{\bar{x}_i}^* - \mathbf{\bar{x}}^*)^{\mathbf{T}} \quad (11)$$

Where $\mathbf{\bar{x}_i}^* = (\mathbf{1}/(\mathbf{N_i} + \mathbf{1}))(\mathbf{N_i}\mathbf{\bar{x}_i} + \mathbf{y})$ and $\mathbf{N_i^*} = \mathbf{N_i} + \mathbf{1}$, if $\mathbf{y}$ belongs to class i otherwise $\mathbf{\bar{x}_i}^* = \mathbf{\bar{x}_i}$ and $\mathbf{N_i}^* = \mathbf{N_i}$.

If y is a new class sample, which means $\mathbf{k}$ is the $(\mathbf{c}+\mathbf{1})^{\mathbf{th}}$ class, then the updated within class scatter matrix does not change:

$$S_W^* = \sum_{\mathbf{i=1}}^{\mathbf{c}} \mathbf{S_i} + \mathbf{S_k} = \sum_{\mathbf{i=1}}^{\mathbf{c+1}} \mathbf{S_i} = \sum_{\mathbf{i=1}}^{\mathbf{c}} \mathbf{S_i} = \mathbf{S_W} \quad (12)$$

However if $1 \le i \le c$, then the updated S_W matrix is given by [9]

$$S_W^* = \sum_{\mathbf{i=1, i \ne k}}^{\mathbf{c}} \mathbf{S_i} + S_k^*$$

$$S_k^* = \mathbf{S_k} + \frac{\mathbf{N_k}}{\mathbf{N_k} + \mathbf{1}}(\mathbf{y} - \mathbf{\bar{x}_k})(\mathbf{y} - \mathbf{\bar{x}_k})^{\mathbf{T}} \quad (13)$$

To determine transformation matrix $\mathbf{W}$, dominant eigenvectors of $\mathbf{U} = S_W^{*\,-1}\mathbf{S_B}^*$ is computed. However, this requires evaluation of inverse of matrix S_W^* i.e. $(S_W^*)^{-1}$ which is highly computational intensive operation. It would be useful to determine an alternative approach to compute inverse of $(S_W^*)^{-1}$ which reduces the cost of computation without decreasing its accuracy.

3. ODIFIED INCREMENTAL LINEAR DISCRIMINANT ANALYSIS (MILDA)

It will be noteworthy if we are able to calculate $(S_W^*)^{-1}$ in terms of the previously calculated S_w^{-1}, thereby decreasing the cost of computation. Equation (13) can also be rewritten as

$$S_W^* = \mathbf{S_W} + \frac{\mathbf{N_k}}{\mathbf{N_k} + \mathbf{1}}(\mathbf{y} - \mathbf{\bar{x}_k})(\mathbf{y} - \mathbf{\bar{x}_k})^{\mathbf{T}} \quad (14)$$

According to Woodbury Formula [14]: If A is a matrix of dimension $n \times n$ and U and V are vectors of size n then

$$\left(\mathbf{A}+\mathbf{U}\mathbf{V}^{\mathbf{T}}\right)^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{U}\mathbf{V}^{\mathbf{T}}\mathbf{A}^{-1}}{1+\mathbf{V}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{U}} \qquad (15)$$

Equation (15) allows computing inverse of a perturbative matrix in terms of a given matrix A and change to the given matrix A.

Using (14) and (16), we get

$$(S_W^*)^{-1}=\mathbf{S_W}^{-1}-\frac{\mathbf{S_W}^{-1}\mathbf{U}\mathbf{V}^{\mathbf{T}}\mathbf{S_W}^{-1}}{1+\mathbf{V}^{\mathbf{T}}\mathbf{S_w}^{-1}\mathbf{U}} \qquad (16)$$

Where $\mathbf{U}=\frac{\mathbf{N_k}}{\mathbf{N_k}+\mathbf{1}}\left(\mathbf{y}-\overline{\mathbf{x}}_{\mathbf{k}}\right)$ and $\mathbf{V}=\left(\mathbf{y}-\overline{\mathbf{x}}_{\mathbf{k}}\right)$.

The computation of inverse of a matrix of size $n\times n$ requires $O(n^3)$ time. However, inverse of a matrix S_W^* in terms of inverse of matrix S_w can be computed in $o(n^2)$ time thereby decreasing the cost of computation. Hence, it will be more appropriate to represent discriminant eigenspace by $\mathbf{\Phi}=\left(\mathbf{S_W}^{-1},\mathbf{S_B},\overline{\mathbf{x}},N\right)$ which allows updating eigenspace when a new sample is considered in addition to existing samples.

The outline of the procedure based on MILDA is given below:

MILDA Algorithm

Input: [X_1, X_2… X_n]

1. Compute $\overline{\mathbf{x}}$, $\mathbf{S}_B$, $\mathbf{S}_W$
2. Compute $\mathbf{S}_W{}^{-1}$
3. For each tuple **y** do the following:
   If **y** belongs to new class then
   Compute $\mathbf{S}_B^*$ using equation (10)
   $(S_W^*)^{-1}=\mathbf{S_W}^{-1}$
   Else
   Compute $\mathbf{S}_B^*$ using equation (11)
   $(S_W^*)^{-1}=\mathbf{S_W}^{-1}-\frac{\mathbf{S_W}^{-1}\mathbf{U}\mathbf{V}^{\mathbf{T}}\mathbf{S_W}^{-1}}{1+\mathbf{V}^{\mathbf{T}}\mathbf{S_w}^{-1}\mathbf{U}}$
   End
4. Compute (c-1) dominant eigenvectors of $(S_W^*)^{-1}\mathbf{S}_B^*$ *i.e.* **[e_1, e_2,…,$e_{(c-1)}$]**

Output: W =[e_1, e_2, …, $e_{(c-1)}$]

4. EXPERIMENTAL SETUP AND RESULTS

In this section, we have examined the efficiency and accuracy of our modified approach for updating discriminant eigenspace i.e. $\mathbf{\Omega}=(\mathbf{S_W^{-1}},\mathbf{S_B},\overline{\mathbf{x}},\mathbf{N})$ against Sequential ILDA $\Omega=(S_W,\mathbf{S_B},\overline{\mathbf{x}},\mathbf{N})$ proposed by Pang et al. The modified approach is evaluated in terms of discriminability, execution time (time taken to update eigenspace) and memory usage against Sequential ILDA. For all experiments matlab code running on a PC with Intel Pentium 4 2.8 GHz CPU and 256-Mb RAM is used.

Extensive experiments are carried out on three publicly available databases [15]: Yale, ORL, and JAFFE to check the efficacy of the proposed MILDA. The ORL database [15] consists of 40 different individuals with 10 images for each individual. All the images are taken against a dark homogeneous background and the subjects are in up-right, frontal position (with tolerance for some side movement). The images from ORL database were cropped from 112 × 92 to 75 × 50 in our experiment. The Yale database [15] consists of 165 images, which are made up of 16 different individuals with 11 images for each individual. The size of images is changed from 320 × 243 to 100 × 102 in our experiment. The JAFFE database [15] comprises 10 Japanese females. Each person has seven facial expressions: "happy," "sad," "surprise," "angry," "disgust," "fearful," and "neutral." There are three or four images for each facial expression of each person. The images from JAFFE database were cropped from 256× 256 to 128 × 128 in our experiment.

For every test, first we constructed an initial feature space using 20% of the total samples in which at least one image from each class is ensured to be included. For carrying out incremental learning one sample is chosen randomly from the remaining training samples. For incremental learning, we first encode features by projecting data presented in terms of updated eigenspace. We have used K-nearest neighbor classifier (K-1) [17] in our experiments. The "leave-one-out" strategy is adopted for testing and training. We found that proposed MILDA method can classify data with same accuracy as batch method and Pang et al method for all the three face datasets.

The computational gain of proposed MILDA scheme for calculating discriminant eigenspace $((S_W^*)^{-1},S_B^*)$ as compared to Pang et al method and batch method is shown in Figure 1 for the ORL database. It can be observed that the difference between the execution time in batch method and both incremental methods (Pang et al method and MILDA) is quite significant. It can also be observed that the difference between the execution time in proposed MILDA scheme and Pang et al method is not significant for small set of samples but becomes pronounced as sample size increases.

Figures 2-3 shows the variations in execution time to compute discriminant eigenspace $((S_W^*)^{-1},S_B^*)$ with number of samples for MILDA and Pang et al method for ORL and JAFFE face datasets. Figures 2-3 show that the proposed MILDA scheme requires less computation time in comparison to Pang et al method. It can also be observed that the difference in execution time to compute eigenspace $((S_W^*)^{-1},S_B^*)$ is more significant when sample size is large. Experimental results on Yale dataset also show that the proposed MILDA scheme outperforms Pang et al method in terms of computation time.

We have also estimated the amount of memory required by Pang et al method and MILDA method for incremental linear Discriminant analysis for face datasets. The results are shown in Figures 5-7. It can be observed that memory requirement is more in Pang et al method in comparison to MILDA method when sample size is large for all the three face datasets.

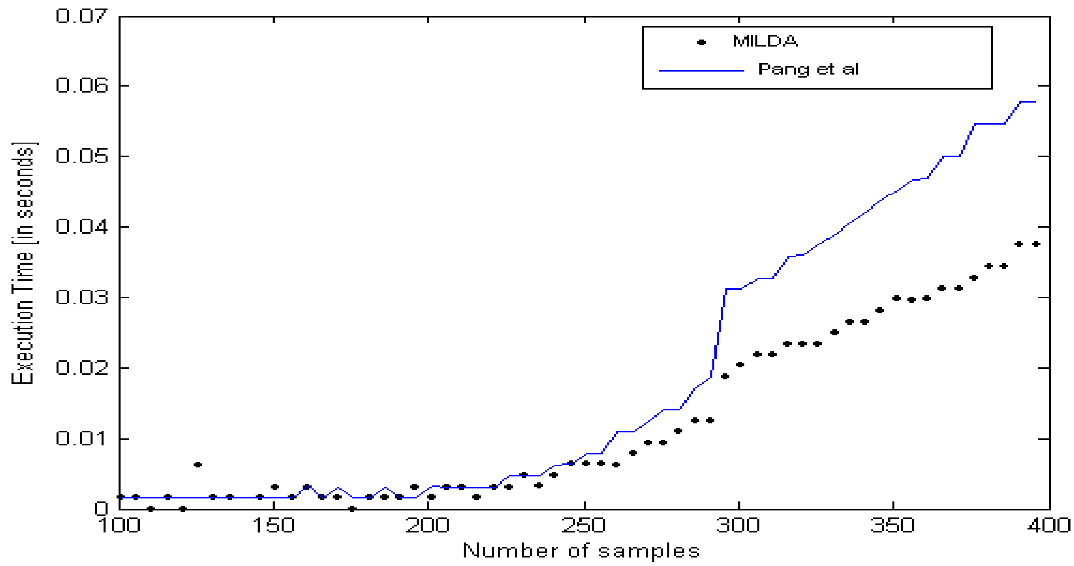

Figure 1: Variation in Execution time for Batch method, Pang et al method and MILDA method when new features are added for ORL dataset

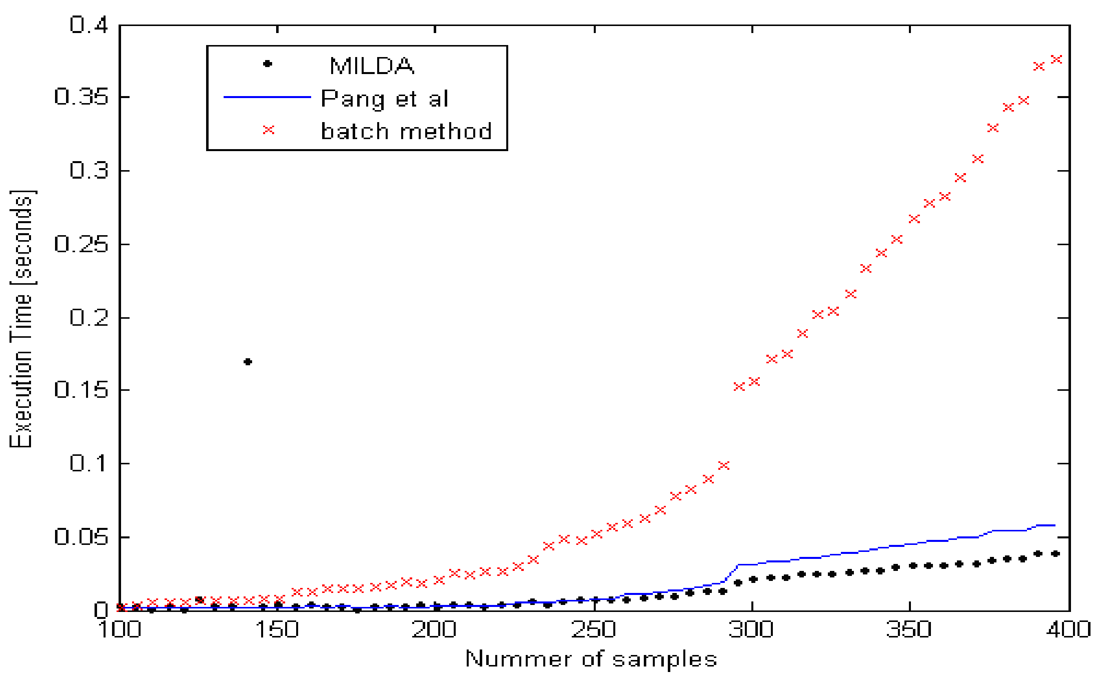

Figure 2: Variation in Execution time for Pang etal method and MILDA method when new features are added for ORL datas

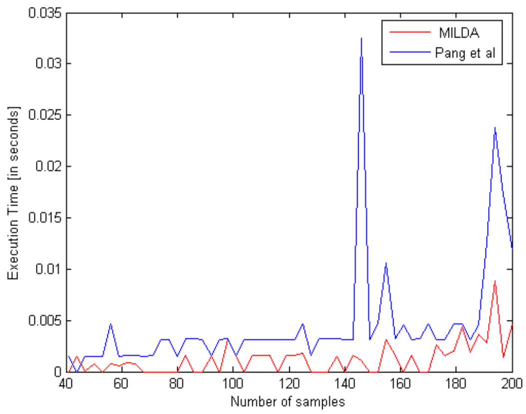

Figure 3: Variation in Execution time for Pang et al method and MILDA method when new features are added for JAFFE dataset

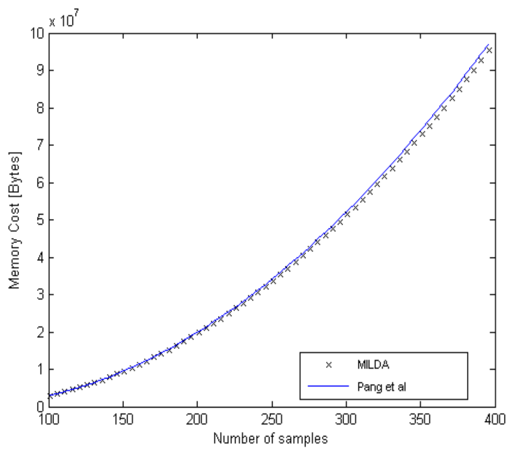

Figure 5: Variation in Memory usage for Pang et al method and MILDA method when new features are added for ORL dataset

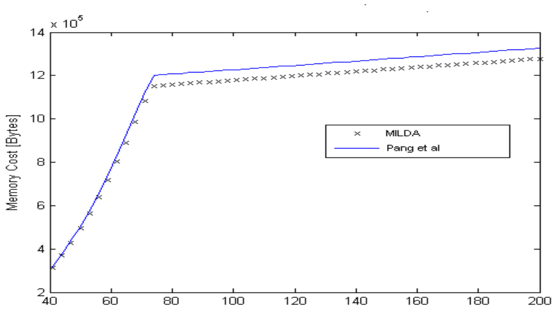

Figure 6: Variation in Memory usage for Pang et al method and MILDA method when new features are added for JAFFE dataset

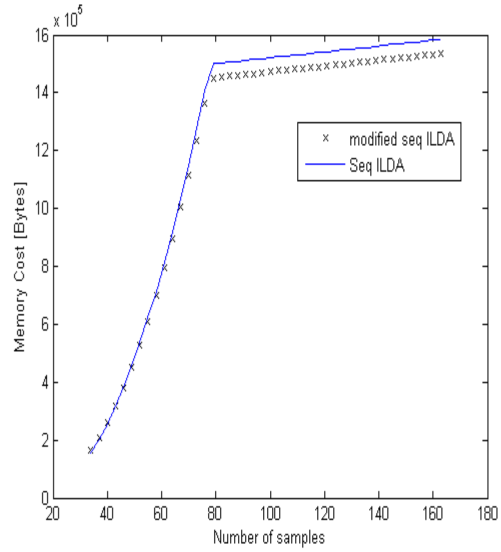

Figure 7: Variation in Memory usage for Pang et al method and MILDA method when new features are added for Yale dataset

5. CONCLUSION

In this paper, we have proposed and investigated modified incremental Linear Discriminant Analysis. Results from this study suggest that the modified incremental Linear Discriminant Analysis is computationally more efficient in comparison to batch method and Pang et al method. This is due to the fact that the batch method and Pang et al method involve intensive matrix operations. The time complexity of inverse of a matrix of size $n \times n$ requires $O(n^3)$ whereas the proposed scheme requires $O(n^2)$. The performance is evaluated in terms of (a) Discriminability, (b) Time required to carry out inverse of matrix (b) Memory requirement. Experimental results with different face datasets show that the proposed scheme is computationally more efficient in terms of time and memory in comparison to batch method and Pang et al method. These investigations suggest that the proposed scheme may be useful for online face detection systems where both memory and computation time is of utmost importance.

REFERENCES

[1]. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs.Fisherfaces: Recognition using class specific linear projection, ”*IEEE Trans. Pattern Anal Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[2]. D. L. Swets and J.Weng, “Using discriminant eigenfeatures for image retrieval,” *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[3]. L. Chen, H. M. Liao, M. Ko, J. Lin, and G. Yu, “A new LDA-based face recognition system which can solve the small sample size problem,” *Pattern Recognit.*, vol. 33, pp. 1713–1726, Oct. 2000.

[4]. A. M. Martlnez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[5]. H.Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognit.*, vol. 34, pp. 2067–2070, 2001.

[6]. C. Liu and H. Wechsler,"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.

[7]. X. Wang and X. Tang,"A unified framework for subspace face recognition," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 26, no. 9, pp. 1222–1228, Sep. 2004.

[8]. S. Ozawa, S. Pang, N. Kasabov, A modified incremental principal component analysis for online learning of feature space and classifier, in C. Zhang, H.W. Guesgen, W.K. Yeap (Eds.), PRICAI : Trends in Artificial Intelligence LNAI, Springer-Verlag (2004) 231-240.

[9]. S. Pang, S. Ozawa, and N. Kasabov, ", Incremental Linear Discriminant Analysis for Classification of Data Streams" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 35, NO. 5, OCTOBER 2005

[10]. P. M. Hall, A. D. Marshall, R. R. Martin, "Incremental eigenanalysis for classification," in *Proc. Brit. Machine Vision Conf.*, vol. 1, pp. 286–295.

[11]. R. K Agrawal and Karmeshu,'' Perturbation scheme for online learning of features ,Incremental Principal Component Analysis". Pattern Recognition, vol 41, no,. 5, pp. 1452-1460, (2008).

[12]. S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkler, and H. Zhang, "An eigenspace update algorithm for image analysis," *Graphical Models Image Process.*, vol. 59, no. 5, pp. 321–332, Sep. 1997.

[13]. S. Mukkamala, A. H. Sung, Artificial intelligent techniques for intrusion detection, IEEE International Conference on Systems, Man and Cybernatics **2** (2003) 1266-1271.

[14]. Gene H. Golub, Charles F. Van Loan, Matrix Computations, 3rd Edition, The John Hopkins University Press Baltimore, 1996

[15]. http:// www.face-rec.com (Face Recognition Homepage)

[16]. M.SBartlett and T.JSejowski," Independent Component of face images: A representation for Face Recognition" in Proc.4th Ann.Jount Symp. Neural Comput. Pasadena,CA,may 17,1997.

[17]. R. O. Duda, P. E. Hart and D. G. Stork, Pattern classification, Second Edition,Wiley, 2001