International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 42-45
ISSN: 2278-2419

Comparison on PCA ICA and LDA in Face Recognition

C.Christry[1], J.Pavithra[2], F.Mahimai Sathya[3]
St. Joseph's College of Arts & Science (Autonomous), Cuddalore
Email:[1]canschris@yahoo.com[2] pavithrasam27@gmail.com[3] mahimaisathya@gmail.com

Abstract-Face recognition is used in wide range of application. In recent years, face recognition has become one of the most successful applications in image analysis and understanding. Different statistical method and research groups reported a contradictory result when comparing principal component analysis (PCA) algorithm, independent component analysis (ICA) algorithm, and linear discriminant analysis (LDA) algorithm that has been proposed in recent years. The goal of this paper is to compare and analyze the three algorithms and conclude which is best. Feret Dataset is used for consistency.

Keywords**:** Face Recognition, PCA, ICA, LDA.

I. INTRODUCTION

Face recognition is defined as identification of a person from an image of their face. It is one of the most successful application of image analysis and understanding. A lot of face recognition algorithm along with their modification have been developed during their past decades [1].In recent years, face recognition has gained much attention. Identifying a face whether it is a known face or unknown face is compared by a person from a database. The interest of researchers in face recognition has grown rapidly in recent years, since; there is a wide range of commercial and law enforcement application on face recognition.Applications of face recognition are credit card, passport checks, criminalinvestigation,etc.The solutions for face recognition problem are of three parts.Facerecognition,Feature extraction done from face region,Decisiontaken.Decision result is based on recognition, verification or categorization of unknown face by comparing the face with the database. Solving a problem is not easy.There are some technical problem for face recognition such as lack of robustness, to illuminate and pose variation.In this paper two techniques is used appearance based techniques, feature based techniques.

A. Appearance based technique***:*** It uses holistic features that are applied to whole face or a specified region in face.
B. Feature based technique: It uses geometric facial features and relations between them.

The goal of this paper is to compare and analyses the threealgorithms and tell them which is best. In previous work it proposes a result of equal working condition when comparing.

The three algorithms are PCA, ICA, and LDA.

a. *PCA*: PCA find a set of projection vector such that the projected sample retains the information about original sample [2].
b. *ICA*: ICA capture second and higher-order statistic and project the I/P data on basic vector that are statistically independent [3][4].
c. *LDA*: LDA is also known as fisherman discriminate analysis. It uses class information and find a set of vector that maximize between class scatter matrixand minimize within class scatter matrix [5][6].

The comparison in done by feret dataset[7]theferet dataset is used for consistency.

II. LITERATURE REVIEW

Bartlett et al. [3] and Liu [8] claim that ICA outperforms PCA, while Baek et al. [9] claim thatPCA is better. Moghaddam [10] states that there is no significant statistical difference.Beveridge et al. [11] claim that in their tests LDA performed uniformly worse than PCA, Martinez [12] states that LDA is better for some tasks, and Belhumeur et al. [5] and Navarreteet al. [13] claim that LDA outperforms PCA on all tasks in their tests (for more than twosamples per class in training phase).

A. PCA (Principal Component Analysis)

Each face in a training set of M images is represented as S-dimension vector. Here PCA finds t-dimension subspace whose basis vector corresponds to maximum variance direction in original image space. This is normally low dimensiont<<s). Basic vector define subspace on image called "face space". To identify the known face the images are projected onto face space and it finds the weight by contributing each vector. To identify the unknown face the images are projected onto face space to obtain set of weight. By comparing the weight of known and unknown face, the face can be identified. If the images are considered as random variables, PCA basics vector is identified as Eigen vector of scatter matrix denoted as S_T, and it is defined as

$$S_{T=\sum}^{M}(X_i-\mu).(X_i-\mu)^T$$

μ- mean of all image in training set.

X_i- i^{th} image with column concatenated in a vector

W_{PCA-} projection matrix is composed of t- Eigen vector correspond to t- largest Eigen value thus PCA finds t dimension face space[2].

The steps in finding the principal components can be summarized as follows:

• Collect xi of an n dimensional data set x, i=1,2, …, m
• Mean correct (center) all the points: Calculate mean mx and subtract it from each data point, xi – mx
• Calculate the covariance matrix: C = (xi - mx)(xi - mx)T
• Determine eigenvalues and eigenvectors of the matrix C.

Volume: 03 Issue: 01 June 2014, Page No. 42-45
ISSN: 2278-2419

- Sort the eigenvalues (and corresponding eigenvectors) in decreasing order.
- Select the first $d \leq n$ eigenvectors and generate the data set in the new representation.
- The projected test image is compared to every projected training image by using a similarity measure. The result is the training image which is the closest to test image.

B. ICA (Independent Component Analysis)

With Gaussian distribution, PCA considered the image element as random variable and it minimizes the second order statistic. If it is a non Gaussian distribution large variation does not correspond to PCA basics vector. Here ICA [3][4] minimizes both second-order and a higher-order dependency in input data and try to find the basic along which data is statistically independent. ICA is a statistical method of transforming observed multidimensional random vector into its component that is statistically independent. ICA is a special case of redundancy technique and it represents the data in terms of statistically independent variables. Two types architecture is provided by the author bartlet[3] for face recognition. Architecture 1: Statistically independent basic image. Architecture 2: Factorial code representation. Also INFOMAX algorithm was implemented and used in ICA by the author bell and sejnouski [3]. As a result to perform ICA, PCA is used reduce its dimensionalitypriorly before performing ICA. Compared with other statistical method ICA provides more powerful data than PCA.

The basic steps to derive the independent components are as follows:

- Collect xi of an n dimensional data set x, i=1,2, …., m
- Mean correct all the points: Calculate mean mx and subtract it from each data point,

Xi – mx

- Calculate the covariance matrix: C = (xi -mx)(xi - mx)T
- The ICA of x factorizes the covariance matrix C into the following form:

C = F Δ FT where Δ is a diagonal real positive matrix.

- F transforms the original data x into Z such that the components of the new data Z are independent: X = F Z.

C. LDA (Linear Discriminant Analysis)

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique which is used for classification problem. It finds the vectors in the underlying space that best discriminate among classes [5] [6]. The goal of LDA is to maximize the between-class scatter matrix measure while minimizing the within-class scatter matrix measure. The between-class scatter matrix *SB* and the within-class scatter matrix are denoted as SW.

The basic steps in LDA are as follows [14]:

- Samples for class1 and class2
- Calculate the mean of class1 and class2 i.e. Mu1 and Mu2.
- Covariance Matrix of the first class and second class i.e. S1 and S2.
- Calculate within-class scatter matrix by using given equation Sw = S1 + S2.
- Calculate between-class scatter matrix by using given equation SB = (Mu1-Mu2)*(Mu1-Mu2).
- Calculate the mean of all classes.
- Compute the LDA projection invSw =inv(Sw) invSw _ by _SB = invSw * SB .
- The LDA projection is then obtained as the solution of the generalized eigen value problem Sw-1SBW = λW W = eig(Sw-1 sb) Where W is projection vector
- Compare the test image's projection matrix with the projection matrix of each training image by using a similarity measure. The result is the training image which is the closest to the test image.

III. SAMPLE MODEL

The gallery contains 1,196 face images, the training images are a randomly selected subset of 500 gallery images. Most importantly, there are four different sets of probeimages such as fa,fb,duplicateI, duplicateII. The fafbprobe set contains 1,195 images of subjects taken at the same time as the gallery images. The only difference is that the subjects were told to assume a different facial expression then in the gallery image. The duplicate I probe set contains 722 images of subjects taken between one minute and 1,031 days after the gallery image was taken. The duplicate II probe setis a subset of the duplicate I probe set, where the probe image is taken at least 18 months after the gallery image. The duplicate II set has 234 images. All images in the data set are of size 384×256 pixels and grayscale.

IV. TRAINING

To train the PCA algorithm we used a subset of classes for which there wereexactlythree images per class. We found 225 such classes (different persons), so our trainingset consisted of 3×225 = 675 images (M = 675, c = 225). One important question worthanswering at this stage is: in what extent does the training set and gallery and probe setsoverlap? Out of 675 images in the training set, 224 were taken from the gallery (33%),another 224 (33%) were taken from the fbset and were of the same subject as the ones takenfrom the gallery, while 3 are in dup1 set. The remaining 224 were not in any set used inrecognition stage. We can therefore conclude that algorithms were trained roughly on 33% ofsubjects later used in the recognition stage. The effect that this percentage of overlap has onalgorithm performance needs further exploration and will be part of our future work. PCAderived, in accordance with theory, M - 1 = 674 meaningful eigenvectors. Weadopted theFERET recommendation and kept the top 40% of those, resulting in 270-dimensional PCAsubspace (40% of 674 ≈270). It was calculated that 97.85% of energy was retained in those270 eigenvectors. This subspace was used for recognition as PCA face space and as input toICA and LDA (PCA was the preprocessing dimensionality reduction step). ICA yielded tworepresentations (ICA1 & ICA2) using the input from PCA (as in [3]). Dimensionality of bothICA representations was also 270. However, LDA yielded only 224-dimensional space sinceit can, by theory, produce a maximum of c - 1 basis vectors. All of those were kept to stayclose to the dimensionality of PCA and ICA spaces and thus make comparisons as fair aspossible. After all the subspaces have been derived, all images from data sets were projectedonto each subspace and recognition using nearest

International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 42-45
ISSN: 2278-2419

neighbour classification with various Distance measures were performed.

SamplesTable 1: Algorithm performance across four metrics. Left part contains the results for rank 1 and the best algorithm-metric combinations are bolded.

V. STATISTICAL SIGNIFICANCE

The simplest method for determining significance is to model each probe image as a binomial test that either succeeds or fails. Under this model, PCA is significantly better than ICA on every probe set. For the fafb probe sets, the differences are significant to a probability of 99.99% for both. For the duplicate I and duplicate II probe sets, the differences are again significant to 99.96% and 99.87%, respectively. When L2 norm is used, ICA performs significantly better on the fafb and duplicate I probe sets, but not on duplicate II probe sets.

Table 1: Showing best Algorithm – Metric combination

Results at rank 1	CMS Result					
Metrics	**L1**	**L2**	**MAH**	**COS**	**Highest Curve**	**Same as Rank 1**
Algorithm	**Fa**					
PCA	**82,26%**	82,18%	64,94%	81,00%	PCA+COS	N
ICA1	81,00%	**81,51%**	64,94%	80,92%	ICA1+L2	Y
ICA2	64,94%	74,31%	64,94%	**83,85%**	ICA2+COS	Y
LDA	78,08%	**82,76%**	70,88%	81,51%	LDA+COS	N
	Fb					
PCA	**5567%**	25,26%	32,99%	18,56%	PCA+L1	Y
ICA1	18,04%	17,53%	**32,99%**	12,89%	ICA1+L1	N
ICA2	15,98%	44,85%	32,99%	**64,95%**	ICA2+COS	Y
LDA	26,80%	26,80%	**41,24%**	20,62%	LDA+L2	N
	Dup1					
PCA	**36,29%**	33,52%	25,62%	33,52%	PCA+L1	Y
ICA1	**32,55%**	31,86%	25,62%	32,27%	ICA1+L1	Y
ICA2	28,81%	31,99%	25,62%	**42,66%**	ICA2+COS	Y
LDA	**34,76%**	32,96%	27,70%	33,38%	LDA+L1	Y
	Dup2					
PCA	**17,09%**	10,68%	14,53%	11,11%	PCA+L1	Y
ICA1	8,97%	7,69%	**14,53%**	8,97%	ICA1+MAH	Y
ICA2	16,24%	19,66%	14,53%	**28,21%**	ICA2+COS	Y
LDA	16,24%	10,26%	**16,67%**	10,68%	LDA+L1	N

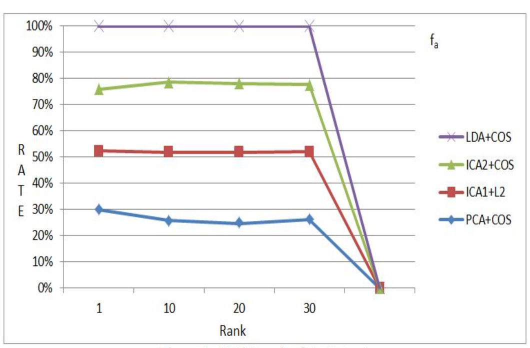

Figure 1: CMS Result of Fa Dataset

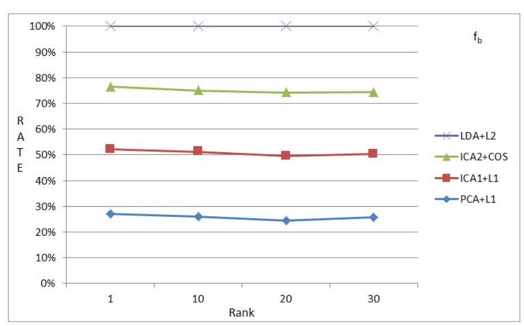

Figure 2: CMS Result of Fb Dataset

Volume: 03 Issue: 01 June 2014, Page No. 42-45
ISSN: 2278-2419

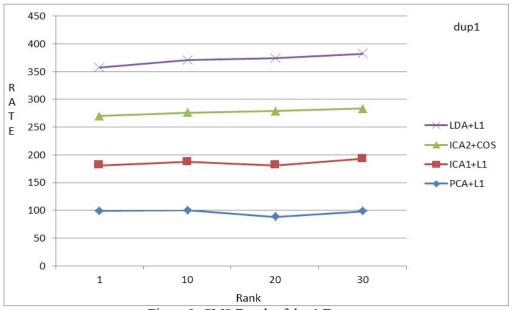

Figure 3: CMS Result of dup1 Dataset

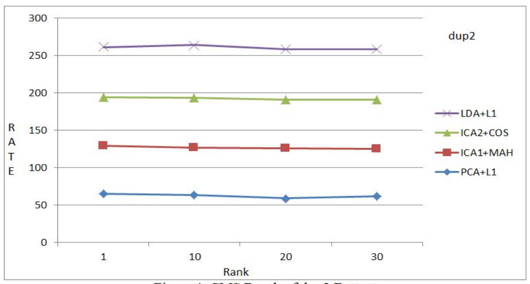

Figure 4: CMS Result of dup2 Dataset

VI. CONCLUSION

In this based on rank we conclude that PCA, ICA and LDA are in equal working conditions on all four probe sets. In future work the difference in performance between PCA, ICA, and LDA can be statistically significant.

REFERENCES

[1] W. Zhao, R. Chellappa, J. Phillips, A. Rosenfeld, "Face Recognition in Still and Video Images: A Literature Survey", *ACM Computing Surveys*, Vol. 35, Dec. 2003, pp. 399-458
[2] M. Turk, A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86
[3] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, "Face Recognition by Independent Component Analysis", *IEEE Trans. on Neural Networks*, Vol. 13, No. 6, November 2002, pp. 1450-1464
[4] B. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, "Recognizing Faces with PCA and ICA", *Computer Vision and Image Understanding (Special Issue on Face Recognition)*, Vol. 91, Issues 1-2, July-August 2003, pp. 115-137
[5] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant Analysis of PrincipalComponents for Face Recognition", *Proc. of the 3rd IEEE International Conference onAutomatic Face and Gesture Recognition*, 14-16 April 1998, Nara, Japan, pp. 336-341
[6] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, "The FERET Evaluation Methodology forFace Recognition Algorithms", *IEEE Trans. on Pattern Analysis and MachineIntelligence*, Vol. 22, No. 10, October 2000, pp. 1090-1104
[7] C. Liu, H. Wechsler, "Comparative Assessment of Independent Component Analysis(ICA) for Face Recognition", *Second International Conference on Audio- and VideobasedBiometric Person Authentication*, Washington D.C., USA, 22-23 March 1999
[8] K. Baek, B. Draper, J.R. Beveridge, K. She, "PCA vs. ICA: A Comparison on the FERETData Set", *Proc. of the Fourth International Conference on Computer Vision, PatternRecognition and Image Processing*, Durham, NC, USA, 8-14 March 2002, pp. 824-827
[9] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for VisualRecognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6,October 2002, pp. 780-788
[10] J.R. Beveridge, K. She, B. Draper, G.H. Givens, "A Nonparametric StatisticalComparison of Principal Component and Linear Discriminant Subspaces for FaceRecognition", Proc. of the IEEE Conference on Computer Vision and PatternRecognition, December 2001, Kaui, HI, USA, pp. 535- 542
[11] A. Martinez, A. Kak, "PCA versus LDA", IEEE Trans. on Pattern Analysis and MachineIntelligence, Vol. 23, No. 2, February 2001, pp. 228-233
[12] P. Navarrete, J. Ruiz-del-Solar, "Analysis and Comparison of Eigenspace-Based FaceRecognition Approaches", *International Journal of Pattern Recognition and ArtificialIntelligence*, Vol. 16, No. 7, November 2002, pp. 817-830
[13] Aly A. Farag, ShireenY.Elhabian, “A Tutorial on Data Reduction”, October 2 2008.