ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Feature Extraction and Dimension Redution Using LDA and MFA

R Dorathy[1]*, K Amareswar[2], P Pavan Kumar[2].

[1]Department of CSE, Saveetha school of engineering, Saveetha university, Chennai, Tamil Nadu, India.
[2]Department of CSE/IT, Saveetha school of engineering, Saveetha university, Chennai, Tamil Nadu, India.

ABSTRACT

Numerous learning applications are portrayed by high measurements. Typically, not these measurements are significant and some are repetitive. There are two primary methodologies; highlight determination and highlight change are utilized to lessen dimensionality. The same structure permits us to propose less difficult criteria, which measure two parts of the inserting, in particular its general quality and its inclination to support either interruptions or expulsions. In the end, a basic examination delineates the soundness of the methodology. Face recognizable proof includes one-to-numerous matches that look at an inquiry face picture against all the layout pictures in the database to decide the personality of the question face. Another face acknowledgment situation includes a watch-list check, where an inquiry face is coordinated to a rundown of suspects (one-to-few matches).

Keywords: extraction, LDA, MFA.

Corresponding author

ISSN: 0975-8585

I.INTRODUCTION

Low dimensional representations from high dimensional information, it endures algorithmic execution is delicate to the measure of neighbors, the calculation experiences the notable little specimen size (SSS) issue and the calculation de-stresses little separation sets. Keeping in mind the end goal to accomplish this we propose exponential implanting utilizing network exponential and give a general structure to dimensionality decrease.

II.DIGITAL IMAGE PROCESSING

A picture might be characterized as a two-dimensional capacity, f(x, y), in which x and y are the spatial (Plane) co-ordinates and the

Abundancy off at any pair of directions (x, y) is known as the power or dark level of the picture by then. Whenever x, y, and the abundancy estimations of f are all limited, discrete amounts, we call the picture a computerized picture. Pixel is the term most broadly used to signify the components of an advanced picture. Advanced picture handling is a subset of the electronic space wherein the picture is changed over to a variety of little whole numbers, called pixels, speaking to a physical amount, for example, scene brilliance, put away in a computerized memory, and prepared by PC or other computerized equipment.

III.FUNDAMENTAL STEPS IN DIGITAL IMAGE PROCESSING

Image processing is important for the digital image. The fundamental steps are used to access the image, the steps as follows.

III.1 IMAGE ACQUISITION

Image acquisition is the first process acquisition could be as simple as being given an image that is already in digital form.

III.2 IMAGE ENHANCEMENT

Picture upgrade is among the easiest and most engaging zones of computerized picture handling. Fundamentally, the thought behind improvement systems is to bring out point of interest that is darkened, or basically to highlight certain components of enthusiasm for a picture.

III.3 IMAGE RESTORATION

Picture reclamation is a zone that likewise manages enhancing the presence of a picture. In any case, not at all like improvement, which is subjective, picture rebuilding is objective, as in reclamation systems have a tendency to be founded on scientific or probabilistic models of picture debasement

Color image processing is an area that has been gaining in importance because of the significant increase in the use of digital images over the Internet, which covers fundamental concepts in color models and basic color processing in a digital domain.

Wavelets are the foundation for representing images in various degrees of resolution.

Compression deals with techniques for reducing the storage required to save an image, or the bandwidth required to transmit it.

III.4 IMAGE COMPRESSION

Picture pressure is recognizable (maybe unintentionally) to most clients of PCs as picture record expansions, for example, the jpg document augmentation utilized as a part of the JPEG(Joint Photographic

Experts Group) picture pressure standard. Morphological handling manages devices for extricating picture parts that are helpful in the representation and depiction of shape
Segmentation procedures partition an image into its constituent parts or objects.

IV.COMPONENTS OF AN IMAGE PROCESSING SYSTEM

Albeit huge scale picture handling frameworks still are being sold for enormous imaging applications, for example, preparing of satellite pictures, the pattern proceeds toward scaling down and mixing of universally useful little PCs with specific picture handling equipment. The capacity of every part is talked about in the accompanying sections, beginning with picture detecting. With reference to detecting, two components are required to obtain computerized pictures.

The first is a physical gadget that is touchy to the vitality emanated by the article we wish to picture. The second, called a digitizer, is a gadget for changing over the yield of the physical detecting gadget into computerized structure. For example, in an advanced camcorder, the sensors deliver an electrical yield corresponding to light power. The digitizer changes over these yields to computerized information.

Particular picture handling equipment for the most part comprises of the digitizer just said, in addition to equipment that performs other primitive operations, for example, a number-crunching rationale unit (ALU), which per shapes number-crunching and sensible operations in parallel on whole pictures. One case of how an ALU is utilized is as a part of averaging pictures as fast as they are digitized, with the end goal of commotion lessening. This sort of equipment in some cases is known as a front-end subsystem, and its most recognizing trademark is velocity.

As it were, this unit performs capacities that require quick information throughputs (e.g., digitizing and averaging video pictures at 30 outlines) that the ordinary primary PC can't deal with. The PC in a picture handling framework is a universally useful PC and can run from a PC to a super PC. In committed applications, now and again extraordinarily composed PCs are utilized to accomplish a required level of execution.

Be that as it may, our enthusiasm here is on broadly useful picture preparing frameworks. In these frameworks, any very much prepared PC-sort machine is appropriate for disconnected picture handling errands. Programming for picture preparing comprises of particular modules that perform particular assignments. An all around planned bundle likewise incorporates the capacity for the client to compose code that, as a base, uses the particular modules. More complex programming bundles permit the combination of those modules and universally useful programming orders from no less than one script.

Mass stockpiling ability is an absolute necessity in picture preparing applications. A picture of size 1024*1024 pixels, in which the power of every pixel is a 8-bit amount, requires one megabyte of storage room if the picture is not packed. At the point when managing thousands, or even millions, of pictures, giving satisfactory stockpiling in a picture preparing framework can be a test.

Advanced capacity for picture preparing applications falls into three foremost classifications: (1) fleeting capacity for use amid handling, (2) on-line stockpiling for generally quick review, and (3) documented capacity, portrayed by rare access. Capacity is measured in bytes (eight bits), Kbytes (one thousand bytes), Mbytes (one million bytes), G bytes (which means giga, or one billion, bytes), and T bytes (which means tera, or one trillion, bytes).

One strategy for giving transient stockpiling is PC memory. Another is by specific sheets, called outline cushions, that store one or more pictures and can be gotten to quickly, for the most part at video rates (e.g., at 30 complete pictures for every second). The last strategy permits for all intents and purposes immediate picture zoom, and in addition scroll (vertical moves) and container (level movements). Outline cradles as a rule are housed in the particular picture handling equipment unit.

Online capacity by and large appears as attractive circles or optical-media stockpiling. The key element portraying on-line stockpiling is continuous access to the put away information. At long last, chronicled capacity is portrayed by monstrous stockpiling prerequisites yet occasional requirement for access. Attractive tapes and optical circles housed in "jukeboxes" are the standard media for authentic applications.

Picture shows being used today are for the most part shading (ideally level screen) TV screens. Screens are driven by the yields of picture and representation show cards that are a vital part of the PC framework. At times are there prerequisites for picture show applications that can't be met by presentation cards accessible industrially as a major aspect of the PC framework. Sometimes, it is important to have stereo showcases, and these are executed as headgear containing two little shows inserted in goggles worn by the client.

Printed version gadgets for recording pictures incorporate laser printers, film cameras, heat-delicate gadgets, inkjet units, and advanced units, for example, optical and CD-ROM circles. Film gives the most elevated conceivable determination, yet paper is the undeniable medium of decision for composed material. For presentations, pictures are shown in video form transparencies or in an advanced medium if picture projection hardware is utilized. The last approach is picking up acknowledgment as the standard for picture presentations.

Systems administration is right around a default capacity in any PC framework being used today. In light of the huge measure of information characteristic in picture handling applications, the key thought in picture transmission is bandwidth.In devoted systems, this normally is not an issue, but rather correspondences with remote locales by means of the Internet are not generally as effective. Luckily, this circumstance is enhancing rapidly as a consequence of optical fiber and other broadband innovations.

V.DIMENSIONALITY REDUCTION

Dimensionality lessening is the change of high-dimensional information into a lower dimensional information space. At present, the most broadly utilized dimensionality diminishment strategies are subspace change. Primary Component Analysis (PCA) [2] is a widely utilized direct subspace change technique augmenting the difference of the changed elements in the anticipated subspace. Straight Discriminant Analysis (LDA) [3] encodes discriminant data by expanding the between-class covariance, and in the mean time minimizing the inside class covariance in the anticipated subspace. Another critical subspace technique is the Bayesian calculation utilizing probabilistic subspace [4]. Wang et al. [5] demonstrated face contrast with three parts and utilized them to bind together PCA, LDA and Bayesian into a general structure. LDA experiences the Small Sample Size (SSS) issue.

PCA and LDA neglect to find the hidden complex structure, in which the high dimensional picture data. Territory Preserving Projections (LPP) [11] to safeguard the region of picture tests. LPP is notable as a Laplacian inserting calculation. LPP and its varieties just describe the area of tests, so they don't promise a decent projection for arrangement purposes. To address this, Unsupervised Discriminant Projections (UDP) [12] presents the idea of no territory and describes the no area of tests by utilizing the nonlocal diffuse. A succinct model for highlight extraction can be acquired by boosting the proportion of nonlocal scramble to neighborhood dissipate. A succinct paradigm for highlight extraction can be gotten by boosting the proportion of nonlocal disseminate to neighborhood diffuse. The vast majority of the above existing calculations were brought together into a general diagram inserting structure proposed by Yan et al. [13]. Also, another administered dimensionality lessening calculation Marginal Fisher Analysis (MFA) was proposed by them under this structure too.

VI.FACE DETECTION

Face recognition is a PC innovation that decides the areas and sizes of human countenances in discretionary (advanced) pictures. It distinguishes facial components and overlooks whatever else, for example, structures, trees and bodies. Early face-location calculations concentrated on the discovery of frontal human appearances, while fresher calculations endeavor to illuminate the more broad and troublesome issue of multi-perspective face recognition.

The location of countenances that are either pivoted along the hub from the face to the eyewitness (in-plane revolution), or turned along the vertical or left-right. Face location is utilized as a part of biometrics, regularly as a part of (or together with) a facial acknowledgment framework. It is additionally utilized as a part of video observation, human PC interface and picture database administration. Some late advanced cameras use face recognition for self-adjust Also, confront discovery is valuable for selecting locales of enthusiasm for photograph slideshows that utilization a container and-scale Ken Burns impact.

VII.FEATURE DETECTION

In PC vision and picture preparing the idea of highlight recognition alludes to techniques that go for processing reflections of picture data and settling on nearby choices at each picture point whether there is a picture highlight of a given sort by then or not. The subsequent components will be subsets of the picture area, frequently as confined focuses, ceaseless bends or associated locales.

Highlight recognition is characterized as an "intriguing" part of a picture, and elements are utilized as a beginning stage for some PC vision calculations. Since components are utilized as the beginning stage and primary primitives for ensuing calculations, the general calculation will regularly just be on a par with its element locator. Therefore, the attractive property for an element finder is repeatability: regardless of whether the same component will be identified in two or more diverse pictures of the same scene**.**

VII.1 Marginal Fisher Analysis:

Differing from LPP and UDP, MFA uses the class label information to construct two graphs based on *k* nearest neighbors: an intrinsic graph that characterizes the intraclass compactness and a penalty graph which characterizes the interclass separability. The intrinsic graph illustrates the intraclass point adjacency relationship, where each sample is connected to its *k*1-nearest neighbors of the same class. The corresponding similarity matrix is denoted as **H***c*.

$$H_{ij}^{c} = \begin{cases} 1 & \text{if } i \in N_{k_1}^{+}(j) \text{ or } j \in N_{k_1}^{+}(i); \\ 0 & \text{otherwise;} \end{cases}$$

where*N*+*k*1

*(i)*indicates the index set of the *k*1 nearest neighbors of the sample **x***i* in the same class. The penalty graph illustrates the interclass marginal point adjacency relationship and the marginal point pairs of different classes are connected. The corresponding similarity matrix is denoted as **H***p*.

$$H_{ij}^{p} = \begin{cases} 1 & \text{if } (i, j) \in P_{k_2}(c_i) \text{ or } (i, j) \in P_{k_2}(c_j); \\ 0 & \text{otherwise;} \end{cases}$$

where*Pk*2 *(c)* is a set of data pairs that represent the *k*2 nearest pairs among the set {*(i, j), i ∈πc, j /∈πc*}. *πc*denotes a set of the elements belonging to *c*th class. Intraclass compactness is characterized as follows:

$$\mathbf{S}_c = \frac{1}{2}\sum_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T H_{ij}^{c},$$

where= **D***c*−**H***c*is the Laplacian matrix from the intrinsic graph. Similarly, the interclass separability is characterized by

$$\mathbf{S}_p = \frac{1}{2}\sum_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T H_{ij}^{p},$$

where= **D***p*−**H***p*is the Laplacian matrix from the penalty graph. MFA tries to find a transformation matrix which will make interclass more compact while simultaneously making interclass more separable. The criterion function of MFA is as follows:

$$\min_{\mathbf{W}} tr\left(\left(\mathbf{W}^T \mathbf{S}_p \mathbf{W}\right)^{-1}\left(\mathbf{W}^T \mathbf{S}_c \mathbf{W}\right)\right)$$

VII.2 Small Sample Size Problem

Generally, the number of training samples is always less than their dimensionality. This results in the consequence that LPP, UDP and MFA suffer from the SSS problem. We first investigate the SSS problem for UDP. Due to the symmetry of **H**, Eq. (8) can be rewritten:

$$\begin{aligned}\mathbf{S}_L &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(H_{ij}\mathbf{x}_i\mathbf{x}_i^T + H_{ij}\mathbf{x}_j\mathbf{x}_j^T - 2H_{ij}\mathbf{x}_i\mathbf{x}_j^T)\\ &= \sum_{i=1}^{N} D_{ii}\mathbf{x}_i\mathbf{x}_i^T - \sum_{i=1}^{N}\sum_{j=1}^{N} H_{ij}\mathbf{x}_i\mathbf{x}_j^T\\ &= \mathbf{XDX}^T - \mathbf{XHX}^T\\ &= \mathbf{XLX}^T\end{aligned}$$

Theorem 1: Let *D* and *N* be the dimensionality of the sample and the number of the samples, respectively. If *D* > *N*, then the rank of **S***L* is at most *N* − 1.

Proof: According to the definition of the Palladian matrix and the fact that the similarity matrix is symmetrical,

$$|\mathbf{L}| = \begin{vmatrix} \sum_j H_{1j} - H_{11} & -H_{12} & \cdots & -H_{1N} \\ -H_{12} & \sum_j H_{2j} - H_{22} & \cdots & -H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -H_{1N} & -H_{2N} & \cdots & \sum_j H_{Nj} - H_{NN} \end{vmatrix}$$

we add the 2nd, 3rd,... Nth rows into the 1st row, and then obtain |**L**| = 0. So, the rank of **L** is at most *N* −1. It is known that the maximum possible rank of the product of two matrices is smaller than or equal to the smaller of the ranks of the two matrices. Hence, *rank(***S***L) = rank(***XLX***T)≤N*−1. From Theorem 1, when SSS problem occurs, **S***L* is singular. cannot be solved. So, UDP suffers from SSS problem. Using similar proof of Theorem 1, we can obtain *rank(***S***D) ≤N* with SSS problem. LPP also suffers from SSS problem. So does MFA.

VII.3 Linear discriminant analysis:

VII.4 Linear discriminant analysis (LDA) and the related **Fisher's linear discriminant** are methods used in statistics and machine learning to find the linear combination of features which best separate two or more classes of object or event. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification.

LDA is firmly identified with ANOVA (examination of fluctuation) and relapse investigation, which likewise endeavor to express one ward variable as a straight blend of different components or estimations. In the other two techniques be that as it may, the reliant variable is a numerical amount, while for LDA it is a clear cut variable (i.e. the class name).

LDA is additionally firmly identified with chief part investigation (PCA) and component examination in that both search for straight blends of variables which best clarify the information. LDA unequivocally endeavors to demonstrate the distinction between the classes of information. PCA then again does not check any distinction in class, and component examination assembles the element blends in light of contrasts instead of similitudes. Discriminant examination is likewise not quite the same as component investigation in that it is not a relationship strategy: a qualification between autonomous variables and ward variables (additionally called foundation variables) must be made.LDA works when the estimations mentioned on every objective fact are constant amounts. At the point when managing straight out variables, the proportional procedure is Discriminant Correspondence Analysis

VII.5 LDA for two classes

Consider a set of observations **x** (also called features, attributes, variables or measurements) for each sample of an object or event with known class *y*. This set of samples is called the training set. The classification problem is then to find a good predictor for the class *y* of any sample of the same distribution (not necessarily from the training set) given only an observation **x**.

LDA approaches the problem by assuming that the probability density functions $p(\vec{x}|y=1)$ and $p(\vec{x}|y=0)$ are both normally distributed. Under this assumption, the Bayes optimal solution is to predict points as being from the second class if the likelihood ratio is below some threshold T, so that

$$(\vec{x}-\vec{\mu}_0)^T\Sigma_{y=0}^{-1}(\vec{x}-\vec{\mu}_0) + \ln|\Sigma_{y=0}| - (\vec{x}-\vec{\mu}_1)^T\Sigma_{y=1}^{-1}(\vec{x}-\vec{\mu}_1) - \ln|\Sigma_{y=1}| < T$$

Without any further assumptions, the resulting classifier is referred to as QDA (quadratic discriminant analysis). LDA also makes the simplifying homoscedastic assumption (*i.e.* that the class covariance's are identical, so $\Sigma_{y=0} = \Sigma_{y=1} = \Sigma$) and that the covariance's have full rank. In this case, several terms cancel and the above decision criterion becomes a threshold on the dot product

$$\vec{w} \cdot \vec{x} < c$$

for some constant c, where

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

This means that the probability of an input **x** being in a class *y* is purely a function of this linear combination of the known observations.

VII.6 Fisher's linear discriminant

The terms *Fisher's linear discriminant* and *LDA* are often used interchangeably, although Fisher's original article *The Use of Multiple Measures in Taxonomic Problems* (1936) actually describes a slightly different discriminant, which does not make some of the assumptions of LDA such as normally distributed classes or equal class covariances.Suppose two classes of observations have means $\vec{\mu}_{y=0}, \vec{\mu}_{y=1}$ and covariances$\Sigma_{y=0}$,$\Sigma_{y=1}$. Then the linear combination of features $\vec{w} \cdot \vec{x}$ will have means $\vec{w} \cdot \vec{\mu}_{y=i}$ and variances $\vec{w}^T \Sigma_{y=i} \vec{w}$ for $i = 0,1$. Fisher defined the separation between these two distributions to be the ratio of the variance between the classes to the variance within the classes:

$$S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(\vec{w} \cdot \vec{\mu}_{y=1} - \vec{w} \cdot \vec{\mu}_{y=0})^2}{\vec{w}^T \Sigma_{y=1} \vec{w} + \vec{w}^T \Sigma_{y=0} \vec{w}} = \frac{(\vec{w} \cdot (\vec{\mu}_{y=1} - \vec{\mu}_{y=0}))^2}{\vec{w}^T (\Sigma_{y=0} + \Sigma_{y=1}) \vec{w}}$$

This measure is, in some sense, a measure of the signal-to-noise ratio for the class labelling. It can be shown that the maximum separation occurs when

$\vec{w} = (\Sigma_{y=0} + \Sigma_{y=1})^{-1} (\vec{\mu}_{y=1} - \vec{\mu}_{y=0})$ When the assumptions of LDA are satisfied, the above equation is equivalent to LDA.

VII.7 Multiclass LDA

In the case where there are more than two classes, the analysis used in the derivation of the Fisher discriminant can be extended to find a subspace which appears to contain all of the class variability. Suppose that each of C classes has a mean μ_i and the same covariance Σ. Then the between class variability may be defined by the sample covariance of the class means

$$\Sigma_b = \frac{1}{C} \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T$$

Where μ is the mean of the class means. The class separation in a direction $\vec{w}$ in this case will be given by

$$S = \frac{\vec{w}^T \Sigma_b \vec{w}}{\vec{w} \Sigma \vec{w}} = \frac{\vec{w}^T (\Sigma_b \Sigma^{-1}) \Sigma \vec{w}}{\vec{w} \Sigma \vec{w}}$$

This means that when $\vec{w}$ an eigenvector of ΣbΣ− 1 is the separation will be equal to the corresponding eigenvalue. Since Σ_b is of most rank C-1, then these non-zero eigenvectors identify a vector subspace containing the variability between features. These vectors are primarily used in feature reduction, as in PCA.

Applications

1. Face recognition
2. Marketing

VII.8 Experimental Results with the Principal Component Analysis.

Experiment 1.

Consider the data in Table 5.1 which is displayed in Fig. 5.2.

Sample	X_1	X_2
1	-21.844	-33.68
2	31.92	22.213
3	19.548	23.796
4	-9.068	4.14
5	-30.481	-16.978
6	-22.473	-23.978
7	-35.154	-28.185

8	30	26.162
9	22.465	18.136
10	-36.451	-38.136
11	25.754	17.12
12	-2.104	-8.04
13	-3.407	13.586
14	12.505	2.536
15	47.061	45.289
16	-0.388	11.552
17	38.66	38.084
18	2.264	-3.729
19	-39.355	-34.678
20	-27.962	-16.095
21	26.573	13.994
22	-2.191	1.536
23	-11.531	-19.425
24	-23.147	-32.19
25	-4.763	-16.746

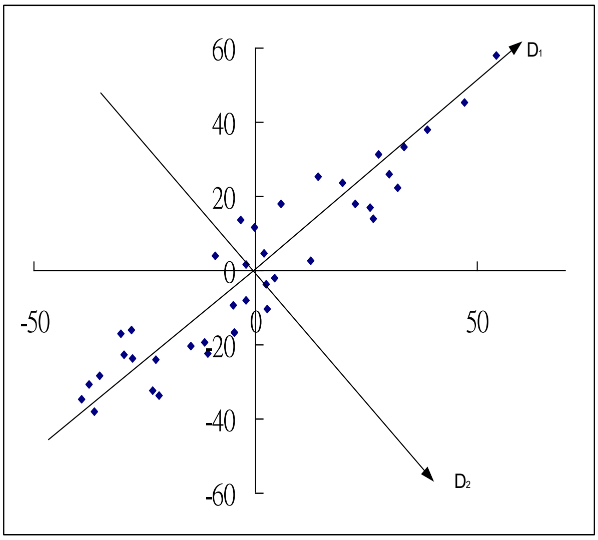


Figure 5.2 Consider the data

The covariance matrix is found to be

$$V=\begin{bmatrix} 604.4 & 561.648 \\ 561.648 & 592.519 \end{bmatrix}$$

The two Eigen value are：

λ_1 =1160.139

and λ_2 =36.78.

The Eigen vectors corresponding to these Eigen values are：

D_1 =(0.710, 0.703),

and D_2 =(-0.703,0.710),

ISSN: 0975-8585

respectively.

The percentage of total variance contributed by D_1 is

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1160.139}{1160.139 + 36.78} = 97\%$$

Both D_1 and D_2 are shown in Fig. 5.2. The samples, represented by the new coordinate system, are shown in Table 5.2.of W_2.

We then calculated the covariance matrix of the data shown in Table 3.2. As expected, the covariance matrix is as follow :

V'= $\begin{bmatrix} 1160.139 & 0 \\ 0 & 36.78 \end{bmatrix}$

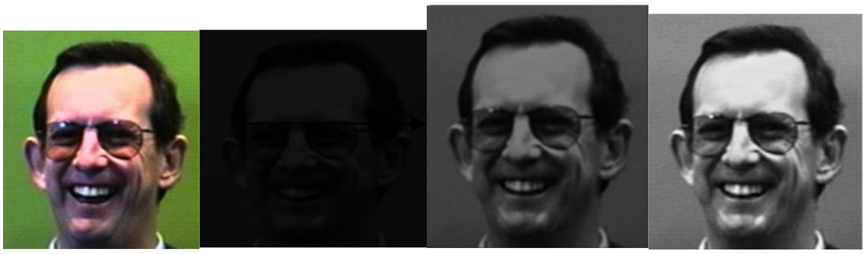
sample	W_1	W_2
1	-383855	-8.206
2	38.674	-6.291
3	30.993	3.535
4	-3.166	9.682
5	-33.247	9.740
6	-32.013	-0.397
7	-44.451	5.060
8	40.087	-2.134
9	29.085	-2.539
10	-52.373	-1.100
11	30.709	-5.574
12	-6.789	-3.865
13	7.484	12.423
14	11.033	-6.622
15	65.667	-0.537
16	8.210	8.854
17	54.628	0.249
18	-0.652	-3.873
19	-52.005	3.400
20	-30.836	8.854
21	29.092	-8.373
22	-0.116	3.002
23	-23.488	-8.373
24	-38.733	-6.231
25	-14.803	-8.183

Based on this tabular column the face can be recognize for the images by this value block diagram shows the procedure for face recognition.

RECOGNIZING FACE

If there any changes in this face ae can also detect the face when compare to original face.

DATA DIMENSIONALITY



If any changes in the sense we can also detect the face based on data dimensionality this diagram shows the data dimensionality reduction also it can read and trained the faces.

VIII.CONCLUSION

We have presented Exponential Embedding and a general framework for dimensionality reduction. Under the framework, we used matrix exponential to extend LPP, UDP and MFA algorithms. These exponential versions can deal with 1) small sample size (SSS) problem, 2) the algorithmic sensitivity to the size of neighbors k and 3) de-emphasizing small distance pairs. The experiments on the synthesized data, UCI and the Georgia Tech face database revealed that the proposed framework can well address above problems.

REFERENCES

[1] Su-Jing Wang and Chun-Guang,"A general exponential framework for dimensionality reduction",vol-23.NO2,*IEEE* February 2014

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[3] 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.

[4] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 780–788, Jun. 2002.

[5] 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.

[6] L. Chen, H. 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, no. 10, pp. 1713–1726, 2000.

[7] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 929–941, Jun. 2005.

[8] P. Howland and H. Park, "Generalizing discriminant analysis using the generalized singular value decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 995–1006, Aug. 2004.

[9] S. Zhang and T. Sim, "Discriminant subspace analysis: A Fukunaga- Koontz approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1732–1745, Oct. 2007.

[10] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. 10th ECCV*, 2008, pp. 725–738.

[11] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[12] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—With application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, Dec. 2001.

[13] J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space?" *Pattern Recognit.*, vol. 36, no. 2, pp. 563–566, 2003.

[14] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 4–13, Jan. 2005.