GENERALIZED 2DFLD METHOD BY INCORPORATING ENTROPY FOR FACE RECOGNITION

project submitted to

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

In partial fulfilment of the requirement for the degree of

Master in Computer Application

Submitted By

Sukanya Goswami

Class Roll No: 001610503006

Examination Roll No: MCA196005

Reg. No: 137314 *of* 2016-2017

Under the guidance of

Dr. Jamuna Kanta Sing

Professor, Department of Computer Science and Engineering

Jadavpur University

2019

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

TO WHOM IT MAY CONCERN

I hereby recommend that the project entitled 'GENERALIZED 2DFLD METHOD BY INCORPORATING ENTROPY FOR FACE RECOGNITION' was prepared under my supervision and guidance at Department of Computer Science and Engineering, Jadavpur University, Kolkata by **Sukanya Goswami (**Reg. No: 137314 (2016-17), Roll No: 001610503006, Examination Roll no: MCA196005), may be accepted in partial fulfillment for the degree of Master in Computer Application in the Faculty of Engineering and Technology, in the academic year 2018-2019. I wish her every success in life.

………………………

Prof.(Dr.) Mahantapas Kundu

Head of the Department

Dept. of Computer Science & Engineering

Jadavpur University, Kolkata- 700032

………………………...

Prof. (Dr.) Jamuna Kanta Sing

Thesis Supervisor

Dept. of Computer Science & Engineering

Jadavpur University, Kolkata- 700032

………………………

Prof. (Dr.) Chiranjib Bhattacharjee

Dean, Faculty Council of Engineering and Technology

Jadavpur University, Kolkata- 700032

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC PROJECT

I hereby declare that the project contains literature survey and original research work by the undersigned student, as part of her MCA final year course. All information have been obtained in accordance to the academic and ethical rules. I also declare that, I consulted research papers and some references, which are not original in this work.

…………………………………….

NAME: SUKANYA GOSWAMI

ROLL No: 001610503006

EXAM ROLL No: MCA196005

REGISTRATION No: 137314 of 2016-2017

PROJECT TITLE: GENERALIZED 2DFLD METHOD BY INCORPORATING ENTROPY FOR FACE RECOGNITION

…………………………………

(SIGNATURE WITH DATE)

JADAVPUR UNIVERSITY

FACULTY OF ENGINNERING AND TECHNOLOGY

CERTIFICATE OF APPROVAL

The foregoing project is hereby accepted as a credible study of a subject of Computer Science, which is carried out and presented in a manner satisfactory to warrant its acceptance as a pre-requisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the project only for the purpose for which it is submitted.

FINAL EXAMINATION FOR EVALUATION OF THESIS:

1. …………...................

2. ……………………….

ACKNOWLEDGEMENT

I, hereby, express my honest and sincere thanks and humble gratitude to my respected teacher and guide *Prof. (Dr.) Jamuna Kanta Sing*, Professor of the Department of Computer Science and Engineering, for his excellent guidance and support.

Finally, I convey my real sense of gratitude and whole hearted thanks to my parents, my maternal uncle and my cousin brother for their unconditional support and being by my side throughout this entire huge work. Last but not the least, I would also like to show my humble gratitude especially to my father, who is a professor and hence, helped me numerously in the field of study and work and he is an endless source of positive thoughts for me.

……………………………..

SUKANYA GOSWAMI

Class Roll No: 001610503006

Examination Roll No: MCA196005

Reg. No: 137314 of 2016-2017

TABLE OF CONTENTS

Declaration

Certificate of Recommendation

Certificate of Approval

Acknowledgement

1. INTRODUCTION 1-5

1.1 Classification of face recognition procedures 2

1.2 Present work under study 5

2. A BRIEF DESCRIPTION OF GENERALIZED 2DFLD ALGORITHM FOR FEATURE EXTRACTION 6-16

2.1 Alternate Fisher's Criteria 9

2.2 Feature Extraction 15

2.3 Calculating Fisherfaces 16

3. GENERALIZED 2DFLD METHOD BY INCORPORATING ENTROPY 17-20

3.1 Entropy, the entry of *randomness* concept in Feature Extraction 17

3.2 Classification of Entropy 18

3.3 Use of Entropy in present study 19

3.4 Proposed method 20

4. EXPERIMENTS 21-31

4.1 Database used for the proposed study 21

4.2 Results and discussions 22

4.3 Table for experimental results 30

4.4 Comparison with different methods 31

5. CONCLUSION 32

6. REFERENCES 33

CHAPTER 1

INTRODUCTION

Human face recognition has become one of the most active and useful area of study and numerous research works in the field of image processing and pattern recognition for the last decade. An acceptable face recognition system can be developed by considering both the facial's features representation and classification issues. Since 'Face' is actually a point in high dimensional space, thus if we consider the entire raw space and don't use feature extraction for face recognition, then obviously it will be an extremely tedious job and the entire process will become hectic to be handled further [1].

Now, when feature extraction is performed, the entire high dimensional space is reduced to a particular low dimensionality space by reducing dimensions by means of feature extraction methods. Eigenface approach was the first method implemented successfully for feature extraction [4, 5]. For this purpose, the first invented procedure was Principal Component Analysis (PCA) (discussed later).

Face recognition has a wide range of applications. In any examination hall, it is used for identity authentication. In case of any identity card like Passport, Aadhar card, face recognition and biometry is necessary. Not only here, for surveillance systems, security purpose, face recognition has wide applications. Since the range of applications is rising day by day, numerous studies regarding face recognition are also going on eventually. As per the studies are going on, face recognition procedures are classified into three categories (based on the psychological study of how human use holistic and local features).

1.1. CLASSIFICATION OF FACE RECOGNITION PROCEDURES

On the basis of human psychological studies and behavioural aspects, face recognition procedures are classified into three categories, viz.

- HOLISTIC MATCHING METHODS
- FEATURE BASED MATCHING METHODS
- HYBRID METHODS

1.1.1. HOLISTIC MATCHING METHODS

In Holistic matching method, the whole face region is used as the raw input in the recognition procedure. In this method, the most commonly used method is Eigenface Approach. Eigenface approach is based on Principal Component Analysis (PCA). In case of Eigenface or eigenvalues, we always use matrices to calculate them and those calculated eigenvalues generate a class of orthonormal bases. The main function of these orthonormal bases is to capture all the directions of maximum variance in the set of training images. The Principal Component Analysis (PCA) method, Linear Discriminant Analysis (LDA) and Fisher's Linear Discrimination (FLD) [11.13]. All three methods fall under subspace problem. The scatter matrices method was first induced in PCA. In PCA, the scatter matrices are maximized for overall face images [9]. But in this case, we are not considering the external disturbance or randomness surrounding the system. In many cases of real time scenario, undesirable problems like lightning, unintentional distorted facial expression can take place. In fact, experimentally it has been proved that PCA method only reduces the dimension, it can't perform any procedure for class discrimination. Now, to avoid this problem, the LDA arrives in the field of Feature Extraction. In case of LDA, the inter-class separation is maximized whereas the intra-class separation is minimized. These two simultaneous procedures are responsible to obtain the best discrimination of vectors [12]. But whenever, the case of SSS (Small Sample Size) problem takes place, PCA individually will lose its significance [18] .

Another method, which is widely used now, is known as Fisher's Linear Discriminant Method or FLD method. FLD method is also widely used for feature extraction. In case of FLD method, there are two very important terms, 'Between class' and 'Within class' scatter matrices of the projected samples [14]. The goal of FLD method is to calculate the optimal projection that maximizes the ratio of between class and within class scatter matrices. Holistic method can be very easy to handle and user friendly, but it has one major disadvantage, known as, Small Sample Size (SSS) [1]. This case arises when sample size becomes less than sample dimension. Since we all know that a face belongs to a high dimensional space and hence if we analyse this situation mathematically, we must obtain that the original matrix has a high level of trace [8]. Thus the within-class scatter matrix used in FLD method becomes almost singular (determinant value=0, non-invertible). Now if the matrix becomes singular, it is impossible to calculate the optimal projection as we can't perform any mathematical operation whenever the matrix is non-invertible (mathematical operations are done later).

1.1.2. FEATURE-BASED MATCHING METHODS

This method is very trivial and is used in early invented face recognition procedure. Hidden Markov Model (HMM) method in this category, is used to cover human face parts like chin, forehead, mouth- as those parts are individual salient features of a human body and hence, could be identified distinctly [2, 3, 4]. Under Feature Based method, we can also use Graph Matching Technique, which is slightly different from HMM and it uses Dynamic Link Architecture Procedure for face detection [3]. The main characteristic of feature based approach is- it mainly uses facial structure and the obtained information from those raw structure can be used for face recognition. Now since, facial structure is only the key idea for this approach, hence the total procedure is entirely heuristic [10, 11]. The main disadvantage of this procedure is- the profile images, images with partial or side view and illumination variations can affect the complexity and make it tedious [3].

1.1.3. HYBRID METHODS

Hybrid method is basically based on human perception. The term 'Hybrid' signifies the hybrid of local features and whole face images. This method catches various human perceptions by integrating both the previous methods. Thus in this method, the behavioural aspect of human faces are detected and hence, feature extractions take place. On the basis of those captures, the face recognition procedure is done. Thus, hybrid method basically catches Modular Eigenface Method, Local feature analysis, Shaped Organized method [3, 5, 6]. As per example, the

Modular Eigenface method works with the combined features of eigenfaces and the individual modules like eye, mouth and nose [16]. Thus this method works with integrating both holistic method and feature based approaches. Since using additional human perceptions, this method is considered much superior than the holistic matching method. Another hybrid method is Hybrid LFA method, which is a hybrid of Principal Component Analysis (PCA) and LFA method. The Shape normalized method uses both shape and gray-level information. For this hybrid characteristic, it is used in case of automatic face detection process. In spite of the huge field of application, hybrid method has one major disadvantage. It needs a huge number of training images, which are randomly chosen from different viewpoints and different conditions, which scenario may not be available always.

Now there are so many soft computing based feature extraction procedure. Approaches like fuzzy logic, swarm intelligence, genetic algorithm are used now for feature extraction in case of face recognition procedure [18]. As various experiments are taking places now, numerous methods are actually developed, which are based on the concept of fuzzy logic and they have a large field of application. Various problems in computer vision and detection of a certain zone problem are solved with the help of fuzzy image segmentation, fuzzy face detection etc. [15]. Here in our entire work, we have used generalized 2DFLD method, which is an improved form of 2DFLD method. In case of conventional 2DFLD method, we used between class and within class scatter matrices $G_b \ and \ G_w$ to calculate the optimal projection. But in case of generalized 2DFLD method, the optimal projections will be based on image row between-class, image row within-class, image column between-class and image column within-class scatter matrices. Each matrix will have comparatively very low dimension, relative to the original high dimension face and hence, it will be convenient and accurate for feature extraction.

PRESENT WORK UNDER STUDY

The main focus of this paper is to insert the idea of entropy (Shannon Entropy) in the Generalized 2DFLD method. The Generalized 2DFLD method calculates the feature extraction and Fisherfaces on the basis of between-class and within-class scatter matrices; but it doesn't include all the external disturbances like flickering lights, or unwanted voice during face recognition procedure. In this study, a term called 'entropy' (symbolized as p_{ik}) is inserted. This term is nothing but the Shannon entropy used in the field of Thermodynamics to measure the randomness or unsteadiness of the adiabatic system considered. This term is multiplied with the between-class and within-class scatter matrices. Those modified matrices will be responsible to get the new feature extractions. With the help of the modified equations obtained after insertion of entropy, we calculate maximum projections, hence the feature extraction and finally, calculating the Fisherfaces. Since, those equations are modified on the basis of external disturbances, hence the results of the average recognition rates (discussed later) will be purer than the original one.

CHAPTER 2

A BRIEF DESCRIPTION OF GENERALIZED 2DFLD METHOD FOR FEATURE EXTRACTION

Generalized 2DFLD method is mainly based on 2D image matrix. It has a major advantage over 2DFLD method as this method maximizes class separability from both row and column directions simultaneously. In 2DFLD method, the scatter matrices were not separately used for both row and column direction. But in G-2DFLD method, the optimization procedure for class separability is considered in both row and column direction of the 2D matrix. Now the scatter matrices must have comparatively more low dimension than those in 2DFLD method. Thus along with the optimization procedure, it will also be easier to deal with low dimensional scatter matrices for feature extraction procedure.

In order to implement the G-2DFLD method, we have to introduce the image matrix *X*, which is basically an $m \times n$ random matrix (Fig. 2.1).

The linear transformation used in this purpose is-

$$Z = U^T X V \tag{2.1}$$

U and *V* are the projection matrices of dimension $m \times p$ ($p \leq m$) and $n \times q$ ($q \leq n$) respectively (Fig. 2.2 & Fig. 2.4). Eventually we have to calculate U^T (Fig. 2.3). Our goal is to find the optimal projection directions *U* and *V* so that optimality is achieved, i.e the maximum class separability is obtained.

Let us consider the total number of images in a training set be *N*, where each image can be signified by a $m \times n$ matrix; say X_i (*i=1,2,…,N).* The total number of classes taken into account is *C*. Let us consider the c^{th} class which contains a total number of N_c images.

Then we define the total mean and class wise mean as-

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{2.2}$$

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N} X_i \tag{2.3}$$

The representations of Z, μ, μ_c are given in Fig. 2.5, Fig. 2.6, Fig. 2.7 respectively.

TWO DIMENSIONAL EMPIRICAL DIAGRAMS OF THE ABOVE TERMS

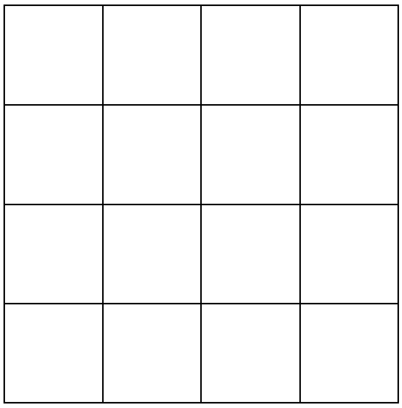

Fig. 2.1 Two dimensional empirical diagram for image matrix $X\ (m \times n)$

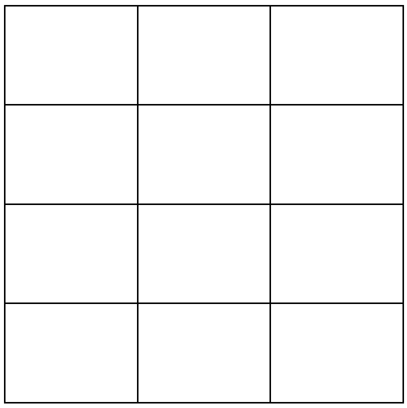

Fig. 2.2 Two dimensional empirical diagram for projection matrix $U\ (m \times p)$

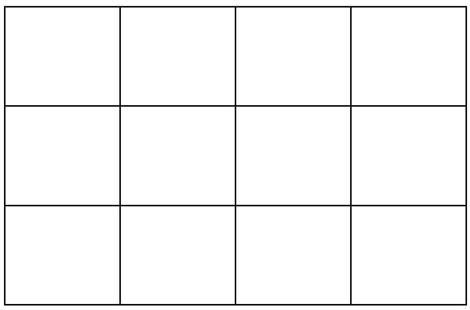

Fig. 2.3 Two dimensional empirical diagram for $U^T\ (p \times m)$

Fig. 2.4 Two dimensional empirical diagram for projection matrix $V\ (n \times q)$

Thus the dimension of Z can be calculated as-

$$Z = U_{p\times m}^{T} \times X_{m\times n} \times V_{n\times q} \tag{2.4}$$

Thus the linearly transformed matrix Z will be of dimension $p \times q$ (from the rule of matrix multiplication) and can be represented as-

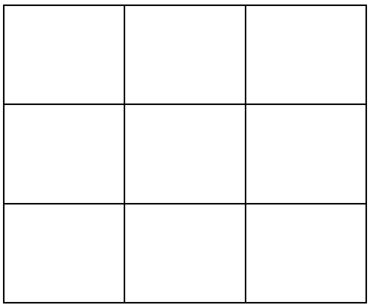

Fig. 2.5 Two dimensional empirical diagram for projection matrix $Z\ (p \times q)$

Let us also find the dimensions of μ_c and μ, the class wise and total mean respectively. Since those means are calculated in original image matrix X, hence their dimension will be $m \times n$ too. Their two dimensional crude diagram will be like-

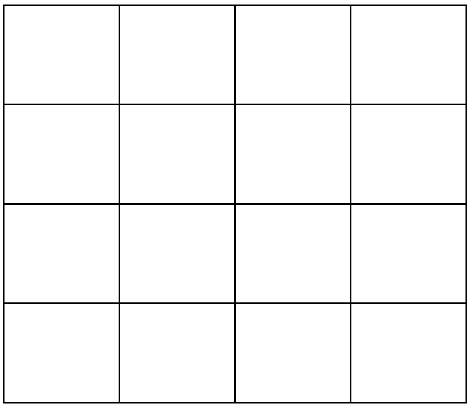

Fig. 2.6 Two dimensional empirical diagram for μ $(m \times n)$

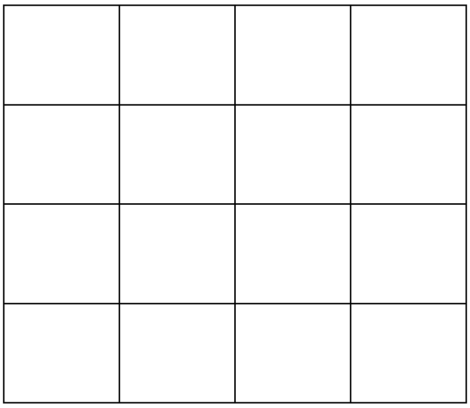

Fig. 2.7 Two dimensional empirical diagram for μ_c $(m \times n)$

2.1. <u>ALTERNATE FISHER'S CRITERIA</u>

Alternative Fisher's criteria are actually two Jacobians *J(U) and J(V)* corresponding to row and column wise projection directions as

$$J(U) = \frac{|U^T G_{br} U|}{|U^T G_{wr} U|} \tag{2.5}$$

$$J(V) = \frac{|V^T G_{bc} V|}{|V^T G_{wc} V|} \tag{2.6}$$

Now the expressions of $G_{br},\ G_{wr},\ G_{bc},\ G_{wc}$ can be written as follows-

$$G_{br} = \sum_{c=1}^{C} N_c(\mu_c - \mu)(\mu_c - \mu)^T \quad (2.7)$$

$$G_{wr} = \sum_{c=1}^{C} \sum_{i \in c}^{N} (X_i - \mu_c)(X_i - \mu_c)^T \quad (2.8)$$

$$G_{bc} = \sum_{c=1}^{C} N_c(\mu_c - \mu)(\mu_c - \mu)^T \quad (2.9)$$

$$G_{wc} = \sum_{c}^{C} \sum_{i \in c}^{N} (X_i - \mu_c)^T (X_i - \mu_c) \quad (2.10)$$

The terms of the LHS of equation (2.7), (2.8), (2.9), (2.10) are known as image row between class scatter matrix, image row within class scatter matrix, image column between class scatter matrix, image column within class scatter matrix respectively (Fig. 2.10, Fig. 2.11, Fig. 2.12, Fig. 2.13). Mathematically we deduce that the dimension of $G_{br}G_{wr}^{-1}$ $is\ m \times m$ and the dimension of $G_{bc}G_{wc}^{-1}$ is $n \times n$. Hence, we obtain that the dimensions of the individual scatter matrices are smaller than the initially developed 2DFLD method, where the dimension of scatter matrices is $mn \times mn$. This phenomenon actually has two positive sides. One is- the between class and within class matrices are separately defined, which will help to fetch the optimality [9] and the other is- the scatter matrices are smaller in dimension than that of in conventional 2DFLD method, which makes easier mathematical operations like inverting the matrices in order to find the optimal projections.

Now let us go back to the Jacobians. The ratios on the right hand side of the Jacobians should be maximized under the condition that the column vectors of U and V are the respective eigenvectors of $G_{br}G_{wr}^{-1}$ and $G_{bc}G_{wc}^{-1}$. From the theory of normalization of a topological space, we obtain the optimal projections U_{opt} and V_{opt} defined as-

$$U_{opt} = argmax|G_{br}G_{wr}^{-1}|$$

$$= [u_1,\ u_{2,} \dots, u_p] \quad (2.11)$$

$$V_{opt} = argmax|G_{bc}G_{wc}^{-1}|$$

$$= [v_1,\ v_{2,} \dots, v_q] \quad (2.12)$$

Here u_i's are the set of normalized vectors obtained in the topological space of $G_{br}G_{wr}^{-1}$ and $v_j's$ are the set of normalized vectors obtained in the topological space of $G_{bc}G_{wc}^{-1}$.

TWO DIMENSIONAL EMPIRICAL DIAGRAMS FOR ALTERNATE FISHER'S CRITERIA

Since the dimension of $\mu_c \; and \; \mu$ is $m \times n$ each, thus the matrix $(\mu_c - \mu)$ will also have the same dimension. Hence the dimension of $(\mu_c - \mu)^T$ is $n \times m$ (Fig. 2.8, Fig. 2.9). The 2D diagrams of those matrices are-

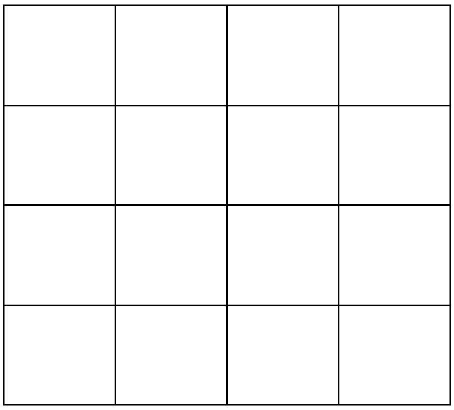

Fig. 2.8 Two dimensional empirical diagram for $(\mu_c - \mu)$ $(m \times n)$

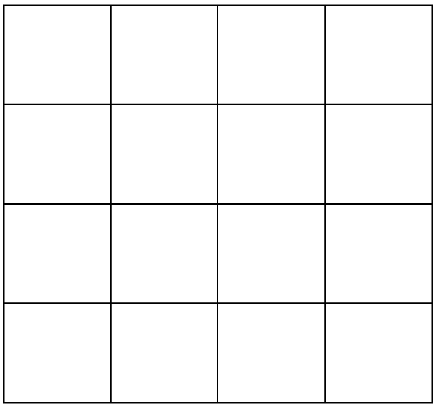

Fig. 2.9 Two dimensional empirical diagram for $(\mu_c - \mu)^T$ $(n \times m)$

From the above expressions, we obtain that the product of $(\mu_c - \mu)$ and $(\mu_c - \mu)^T$ will be of dimension $m \times m$ and the dimension of the product of $(\mu_c - \mu)^T$ and $(\mu_c - \mu)$ will be of dimension $n \times n$.

Thus from the equation (2.7), (2.8), (2.9), (2.10), it is obtained that-

$$G_{br} = \sum_{c=1}^{C} N_c \times (\mu_c - \mu)_{m\times n} \times (\mu_c - \mu)_{n\times m}^{T} \quad (2.13)$$

$$G_{wr} = \sum_{c=1}^{C} \sum_{i\in c}^{N} (X_i - \mu_c)_{m\times n} (X_i - \mu_c)_{n\times m}^{T} \quad (2.14)$$

$$G_{bc} = \sum_{c=1}^{C} N_c \times (\mu_c - \mu)_{n\times m}^{T} \times (\mu_c - \mu)_{m\times n} \quad (2.15)$$

$$G_{wc} = \sum_{c=1}^{C} \sum_{i\in c}^{N} (X_i - \mu_c)_{n\times m}^{T} \times (X_i - \mu_c)_{m\times n} \quad (2.16)$$

Thus the diagrammatic representations will be like-

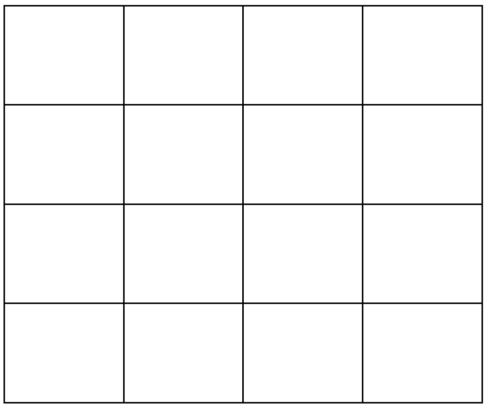

Fig. 2.10 Two dimensional empirical diagram for G_{br} $(m \times m)$

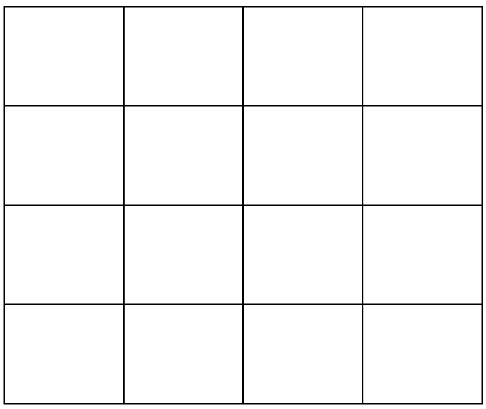

Fig. 2.11 Two dimensional empirical diagram for G_{wr} $(m \times m)$

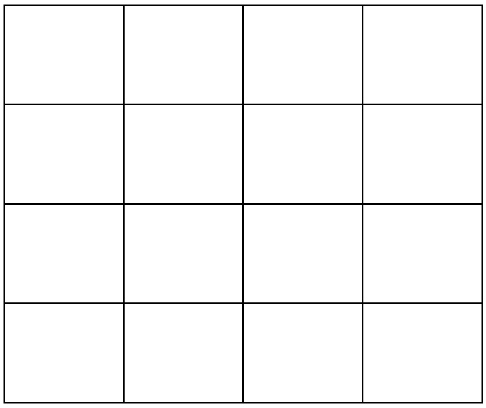

Fig. 2.12 Two dimensional empirical diagram for G_{bc} $(n \times n)$

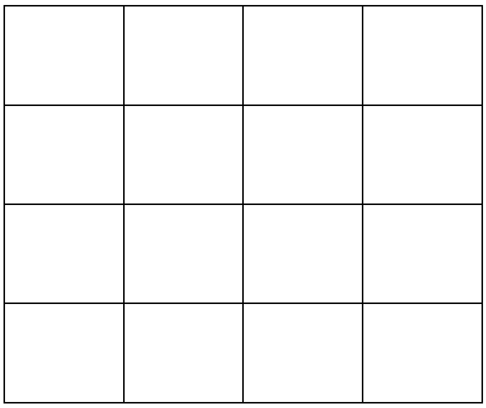

Fig. 2.13 Two dimensional empirical diagram for G_{wc} $(n \times n)$

Now from the mathematical expressions of U_{opt} and V_{opt} , it is evident that their respective dimensions are $m \times p$ and $n \times q$ respectively. Since those matrices are the optimal form of the projection matrices U and V , so their dimensions will also be same (eqn. 2.11 and eqn. 2.12). Their 2D diagrams are shown in Fig. 2.14 and Fig. 2.15.

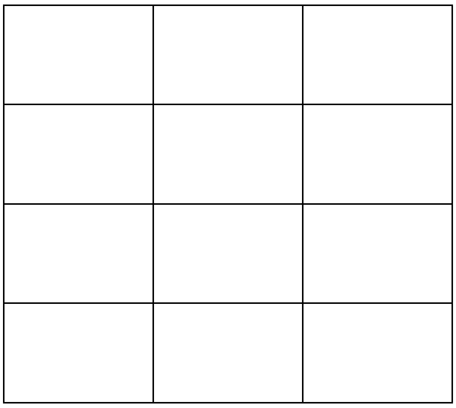

Fig. 2.14 Two dimensional empirical diagram for U_{opt} $(m \times p)$

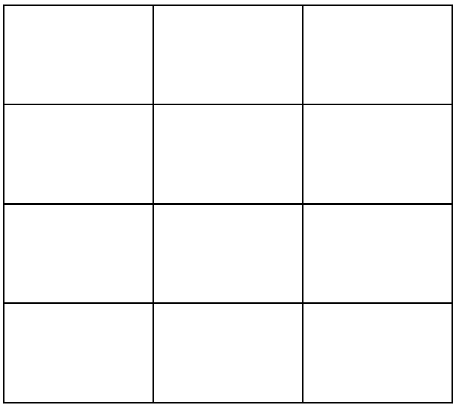

Fig. 2.15 Two dimensional empirical diagram for V_{opt} $(n \times q)$

2.2. FEATURE EXTRACTION

The optimal projection matrices U_{opt} and V_{opt} are used for feature extraction. Let us consider the sample image *X*. Then a relatively low dimensional space image feature can be obtained by the linear projections of the eigenvectors as-

$$z_{ij} = u_i^T X v_j \tag{2.17}$$

This (*i=1,2…p; j=1,2…q*) is known as *'Principal component'* of the sample image; which is nothing but a vector; as originally obtained from a given metric space. Here is the difference with traditional FLD method, where the principal component is a scalar. This principal components are accumulated together to form a relative low dimensional image feature matrix ***Z*,** which has dimension $p \times q$ (Fig. 2.15). Since both *p* and *q* are much smaller than *m* and *n* respectively, hence this image feature matrix is reduced in both row and column directions and is easier to work with.

TWO DIMENSIONAL EMPIRICAL DIAGRAM FOR FEATURE EXTRACTION

From equation (2.12), we obtain that the dimension of the feature matrix z_{ij} will be $p \times q$ (only the rule of matrix multiplication is followed).

$$z_{ij} = {u_i^T}_{p\times m} \times X_{m\times n} \times {v_j}_{n\times q} \tag{2.18}$$

Hence it is diagrammatically represented as-

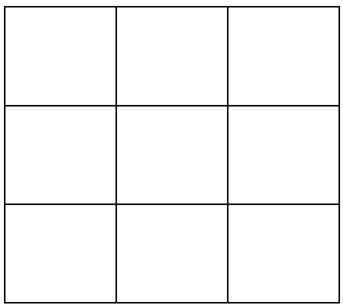

Fig. 2.15 Two dimensional empirical diagram for z_{ij} $(p \times q)$

2.3. FISHERFACE CALCULATION

To calculate fisherfaces, let us consider an image (*i=1,2,...N*), which is an $m \times n$ matrix. Now we know, in order to find the optimal projections, we got the dimension values of the row and column scatter matrices $G_{br}G_{wr}^{-1}$ and $G_{bc}G_{wc}^{-1}$ as $m \times m$ and $n \times n$ respectively. Now those eigenvectors together define a subspace of high dimensional face image. Hence by theory of normalization, if we combine those scatter matrices linearly, we can find the fisherface. The fisherfaces which can be generated by the linear combination of eigenvectors are-

$$I_{ij} = u_i v_j^T, \; i = 1,2, \dots, p; \; j = 1,2, \dots, q \qquad (2.19)$$

Since the dimension of v_j is $n \times q$, hence the dimension of v_j^T is $q \times n$. The dimension of u_i is already known ($m \times p$). But, let us look at the logic of Feature Extraction. The procedure of linear combination of eigenvectors in Fisherface calculation is possible only when optimality is achieved*; i.e.* U_{opt} *and* V_{opt} will be pair of Orthonormal matrices [7]. Now, from the condition of Orthonormal matrices, we find that $p = q$ (Fig. 2.16).

TWO DIMENSIONAL EMPIRICAL DIAGRAM FOR FISHERFACE

Thus from the rule of matrix multiplication, the dimension of I_{ij} will be of dimension $m \times n$.

$$I_{ij} = u_{i_{m \times p}} \times v_j^T{}_{q(=p) \times n} \qquad (2.20)$$

The two dimensional empirical diagram for this case is given as-

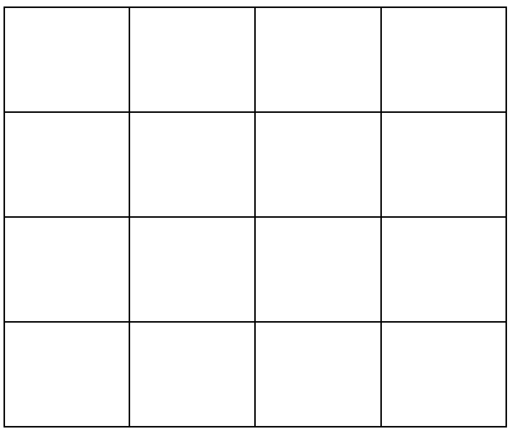

Fig. 2.16 Two dimensional empirical diagram for I_{ij} $(m \times n)$

CHAPTER-3

GENERALIZED 2DFLD METHOD BY INCORPORATING ENTROPY

3.1. Entropy, the entry of *Randomness* concept in Feature Extraction

This part of work actually represents a possible insertion of entropy in feature extraction on the generalized two-dimensional Fisher Linear discriminant (G-2DFLD). The primitive method used for feature extraction is 2DFLD method. Later, generalized 2DFLD method was developed, which was also based on original 2D image matrix. But, in 2DFLD, the class separability is maximised either in row or column direction. But in G-2DFLD method, class separability is maximized both in row and column direction simultaneously. But G-2DFLD method doesn't include one of the most important property, which actually have an influence on the feature extraction. This important property is Entropy or Shannon Entropy. We already know this term in Physics. Entropy is a physical property which signifies 'Modular disorder' or 'Randomness' of a system.

Entropy has numerous application, not only in Physics and applied Mathematics and engineering. In case of face recognition, the term 'Entropy' signifies a meaningful measure for the degree of uncertainty, which can be used to characterize the texture an image and implement that texture.

Mathematically, entropy can be implemented as-

$$-\sum p_i \times log p_i \tag{3.1}$$

Where p_i is the probability of the given symbol.

3.2. CLASSIFICATION OF ENTROPY

There are different kinds of entropy used in the field of image processing and face recognition. Those are briefly discussed below.

3.2.1. SHANNON ENTROPY

a. Shannon entropy is defined only for discrete probability distribution.

b. Shannon entropy is required while measuring how much information is adequate to identify random sample of a distribution.

Shannon entropy is denoted by $H(X)$, where-

$$H(X) = \sum_{i=1}^{N-1} p_i log_2 p_i \quad (3.2)$$

p_i is probability density functions in two-dimensional random variable.

3.2.2. RENYI ENTROPY

a. Important in field of geological survey, ecology and statistics. In case of a measurement of biodiversity, Renyi entropy is a measurement of diversity index.
b. It is also used in XY Spin Chain model, related to Heisenberg's Uncertainty Principle.
c. This entropy is entirely a function of 'α' and calculated in such a way that it will be an automorphic function w.r.t to any particular subgroup of an individual modular group.
d. It is a generalization of Shannon entropy, when used in case of enlargement.

This entropy is defines as the order of α and is mathematically represented as-

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha \quad (3.3)$$

Where $\alpha \geq 0$ and $\alpha \neq 1$

$H_\alpha(X)$ is the Renyi Entropy.

3.2.3. HARVRDA-CHARVEL ENTROPY

a. Used in Statistical mechanics in Physics.
b. It was first used in Bose-Einstein Statistics.
c. Later it was modified by Dracozy.

3.2.4. KAPUR ENTROPY

a. The concept of thresholding for image processing is first inserted here.
b. To reduce the complexity of entropy function, a new threshold algorithm is introduced, denoted by $H_{\alpha}^{K}(p)$, which is in the order of α *,* is mathematically represented as -

$$H_{\alpha}^{K}(p) = \frac{1-\left(\sum_{k=1}^{n} p_k^{1/\alpha}\right)^{\alpha}}{1-\alpha} \tag{3.4}$$

Where $H_{\alpha}^{K}(p)$ is the Kapur Entropy.

3.3. USE OF ENTROPY IN THE PRESENT STUDY

We are already familiar about G-2DFLD method used in feature extraction. At that field, we used between class and within class scatter matrices- which actually helped us to reduce the high dimensional space to comparatively low dimensional space so that the main goal of feature extraction can be fulfilled successfully. But in this procedure, we are considering that the working environment and all the associated instrumentation procedure absolute. But in the real world, it is literally impossible as every working procedure and instrumentation as well as the environment, in which we are pursuing all the experiments, contain some individual disturbances. As per the G-2DFLD method is described, only absolute condition was considered and hence the entire calculations were done accordingly. Thus, the experimental results we are obtaining are not always correct. To make those results more accurate, we consider real life scenario, where entropy is inserted and then, the feature extraction is done. Just like Thermodynamics in Physics, entropy plays the role in Feature Extraction. In case of face recognition also, the term entropy literally signifies the disturbance caused by noise in the surroundings, irregular luminous intensity etc. . Thus, in a nutshell, entropy is the measure of intensity of homogeneity in the performed face recognition procedure.

3.4. PROPOSED METHOD

Since, we need to insert entropy for more accurate results in our experiment procedure, we must introduce a term of entropy in this particular procedure. Here three quantities are considered as the entropy factors, G_{ik}, u_{ik} and p_{ik} which are mathematically defined here-

$$G_{ik} = e^{-\|x_k - v_i\|^2/2\sigma^2} \tag{3.5}$$

$$u_{ik} = \frac{G_{ik}}{\sum_{i=1}^{c} G_{ik}} \tag{3.6}$$

$$p_{ik} = -\sum_{i=1}^{c} u_{ik} \times log u_{ik} \tag{3.7}$$

Now this term p_{ik} is multiplied with all the scatter matrices (image row between-class, image row within-class, image column between-class, image column within-class).

Since those matrices are responsible to find the optimal projection matrices U_{opt} and V_{opt}, hence entropy should be inserted in those matrices and then we have to calculate the optimality on the basis of $G_{br}G_{wr}^{-1}$ and $G_{bc}G_{wc}^{-1}$. Whenever this term is multiplied, all the values afterwards will change in G-2DFLD method and hence, the feature extraction and fisherface calculation should be done.

The modified equations of generalized 2DFLD method are given as-

$$G'_{br} = \sum_{c}^{C} p_{ik} N_c (\mu_c - \mu)(\mu_c - \mu)^T \tag{3.8}$$

$$G'_{wr} = \sum_{c=1}^{C} \sum_{i \in c}^{N} p_{ik} (X_i - \mu_c)(X_i - \mu_c)^T \tag{3.9}$$

$$G'_{bc} = \sum_{c=1}^{C} p_{ik} N_c (\mu_c - \mu)(\mu_c - \mu)^T \tag{3.10}$$

$$G'_{wc} = \sum_{c}^{C} \sum_{i \in c}^{N} p_{ik} (X_i - \mu_c)^T (X_i - \mu_c) \tag{3.11}$$

Now these four new scatter matrices will be responsible to achieve the optimality. Rest of the procedures up to Fisherface calculation will remain unchanged; but the calculations will be based on the modified matrices with entropy inserted.

CHAPTER 4

EXPERIMENTS

4.1. DATABASE USED FOR THE PROPOSED STUDY

The performance of the proposed method has been evaluated on AT&T Laboratory Cambridge Database, formerly known as ORL database. Generally, AT&T Database is used to test performance of the proposed method on the basis of minor variations and scaling [3].

In AT&T database, there exist 400 gray-scale image of 40 persons. Each person has 10 gray-scale image, each having a resolution of 112×92 pixels. The images can be taken in various conditions like dark or dimmed light, open or closed eyes, smiling or not smiling etc. against a dark homogeneous background with scale variation up to 10%. The sample images in AT&T database are shown below-

Fig. 4.1. Sample images of AT&T Database

In AT&T database, the recognition rate is defined is the percentage of ratio for the number of correct recognition by the proposed method to the total number of images in the test set for a certain experimental sample.

The average recognition rate, R_{avg} is mathematically expressed as-

$$R_{avg} = \frac{\sum_{i=1}^{l} n_{cls}^{i}}{l \times n_{tot}} \times 100 \quad (4.1)$$

Where

l is the number of experimental run, performed by random partition of the database into two sets- training and test.

n_{cls}^{i} is the number of correctly recognized faces for i^{th} run.

n_{tot} is the total number of faces in test set.

4.2. RESULTS AND DISCUSSIONS

In this experiment, we considered five scenarios. Let us look through them-

4.2.1. Scenario 1

Orl_trn_5_i (*i*=1,2..10) as training set.

Orl_tst_5_i (*i*=1,2…10) as test set.

Orl_tst_5_i (*i*=1,2…10) as training set.

Orl_trn_5_i (*i*=1,2…10) as test set.

Here each training and test set has 10 image files. First we are considering Orl_trn_5_i as Training set and Orl_tst_5_i as Test set. After calculation of recognition rates for all the feature spaces, we swap those sets; i.e. Orl_tst_5_i as Training set and Orl_trn_5_i as Test set. Thus for each feature space, we get total 20 values. Now we calculate average of values for each feature space, and plot those average recognition rates vs. feature spaces graph.

The average recognition rate along with the feature spaces are given as-

TABLE 4.2.1. Average recognition rates for (Orl_trn_5_i, Orl_tst_5_i)

F_SPACE	AVERAGE
8*8	97.35
10*10	97.475
12*12	97.6
14*14	97.875
16*16	97.925
18*18	97.9
20*20	97.8

Corresponding to this result, the graphical representation will be like-

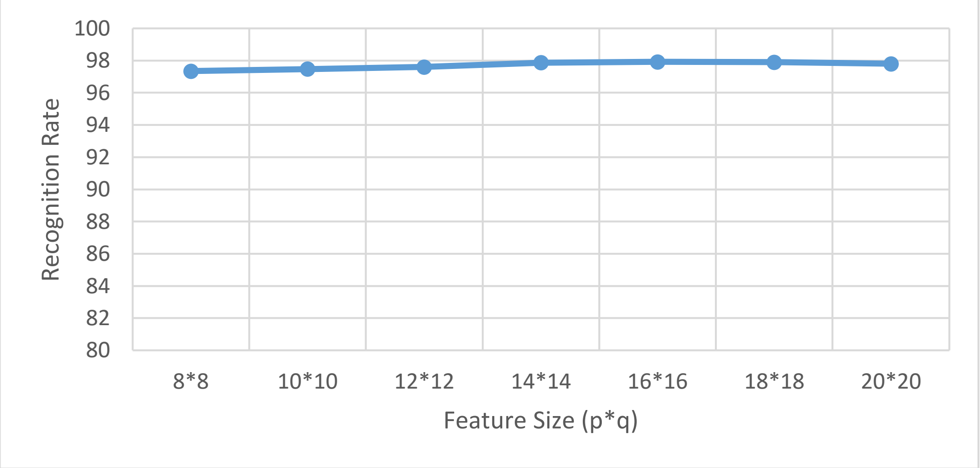

Fig. 4.2.1. Average recognition rate vs. Feature space graph for (Orl_trn_5_i, Orl_tst_5_i)

4.2.2. <u>Scenario 2</u>

Orl_trn_4_i *(i=1,2…20*) as Training set

Orl_tst_6_i *(i=1,2…20)* as Test set

Here each training and test set each has 20 image files. We are considering Orl_trn_4_i as Training set and Orl_tst_6_i as Test set. Thus for each feature space, we get total 20 values. Now we calculate average of values for each feature space, and plot those average recognition rates vs. feature spaces graph.

The average recognition rate along with the feature spaces are like-

TABLE 4.2.2. Average recognition rates for (Orl_trn_4_i, Orl_tst_6_i)

F_SPACE	AVERAGE
8*8	95.28669
10*10	95.71602
12*12	96.13216
14*14	96.54075
16*16	96.76088
18*18	96.11189
20*20	96.06215

Corresponding to this scenario, the graphical representation will look like-

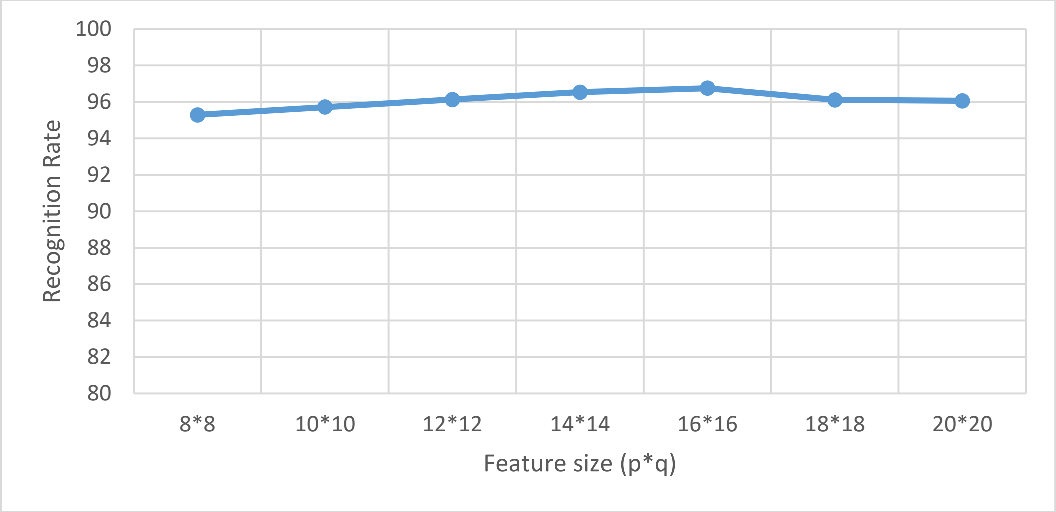

Fig. 4.2.2. Average recognition rate vs. feature space graph for

(Orl_trn_4_i, Orl_tst_6_i)

4.2.3. <u>Scenario -3</u>

Orl_trn_3_i *(i=1,2…20)* as training set.

Orl_tst_7_i *(i=1,2…20)* as test set.

Here each training and test set each has 20 image files. We are considering Orl_trn_3_i as Training set and Orl_tst_7_i as Test set. Thus for each feature space, we get total 20 values. Now we calculate average of values for each feature space, and plot those average recognition rates vs. feature spaces graph.

The average recognition rate along with the feature spaces are given as-

TABLE 4.2.3. Average recognition rates for (Orl_trn_3_i, Orl_tst_7_i)

F_SPACE	AVERAGE
8*8	92.385
10*10	93.0355
12*12	93.25
14*14	93.4055
16*16	93.4425
18*18	93.284
20*20	93.171

Hereby, the graphical representation is given as-

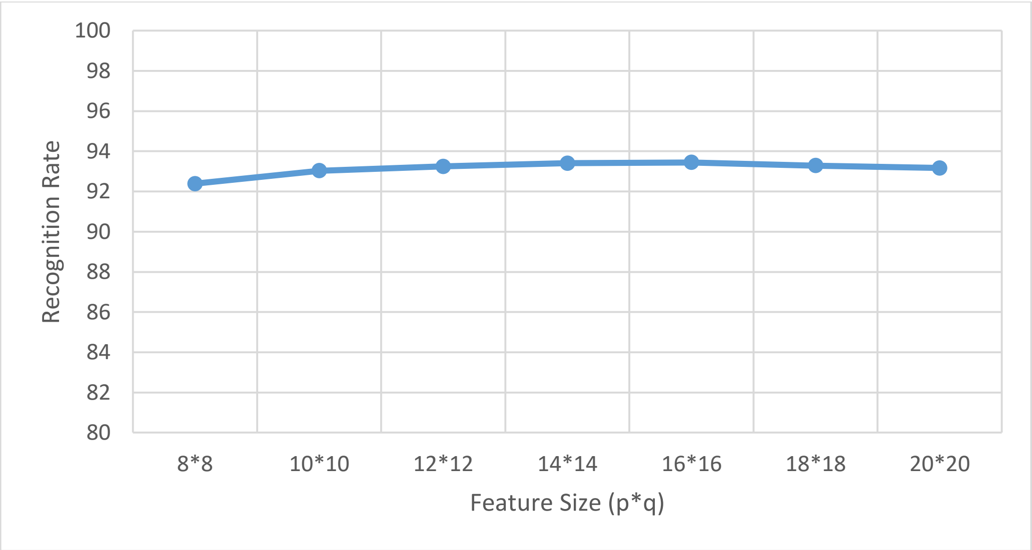

Fig. 4.2.3. Average recognition rate vs. feature space graph for (Orl_trn_3_i, Orl_tst_7_i)

4.2.4. <u>Scenario 4</u>

Orl_tst_6_i *(i=1, 2,…,20)* as training set.

Orl_trn_4_i *(i=1, 2,…,20)* as test set.

Here each training and test set each has 20 image files. We are considering *Orl_tst_6_i* as Training set and *Orl_tst_4_i* as Test set. Thus for each feature space, we get total 20 values. Now we calculate average of values for each feature space, and plot those *average recognition rates* vs. *feature space* graph.

The average recognition ratio along with the feature spaces are given as-

TABLE 4.2.4. Average recognition rates for (Orl_trn_6_i, Orl_tst_4_i)

F_SPACE	AVERAGE
8*8	98.814
10*10	98.8425
12*12	98.9935
14*14	99.5675
16*16	99.4375
18*18	99.3945
20*20	99.337

The graphical representation for this scenario is given as-

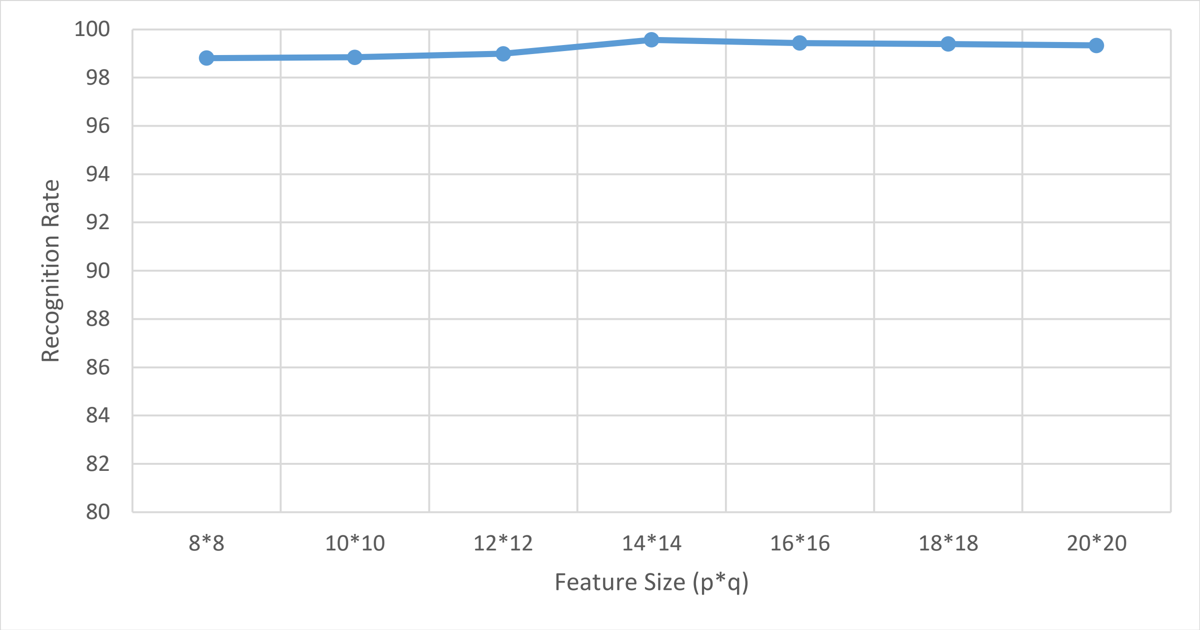

Fig. 4.2.4. Average recognition rate vs. feature space graph for (Orl_tst_6_i, Orl_trn_4_i)

4.2.5. <u>Scenario 5</u>

Orl_tst_7_i *(i=1,2…20)* as Training set.

Orl_trn_3_i *(i=1,2…20)* as Test set.

Here each training and test set each has 20 image files. We are considering Orl_tst_7_i as Training set and Orl_tst_3_i as Test set. Thus for each feature space, we get total 20 values. Now we calculate average of values for each feature space, and plot those average recognition rates vs. feature spaces graph.

The average recognition ration along with the feature spaces are given as-

TABLE 4.2.5. Average recognition rates for (Orl_tst_7_i, Orl_trn_3_i)

F_SPACE	AVERAGE
8*8	98.4175
10*10	98.2915
12*12	98.3335
14*14	98.3675
16*16	98.251
18*18	98.0425
20*20	98.0015

The graphical representation of the above data set is given as-

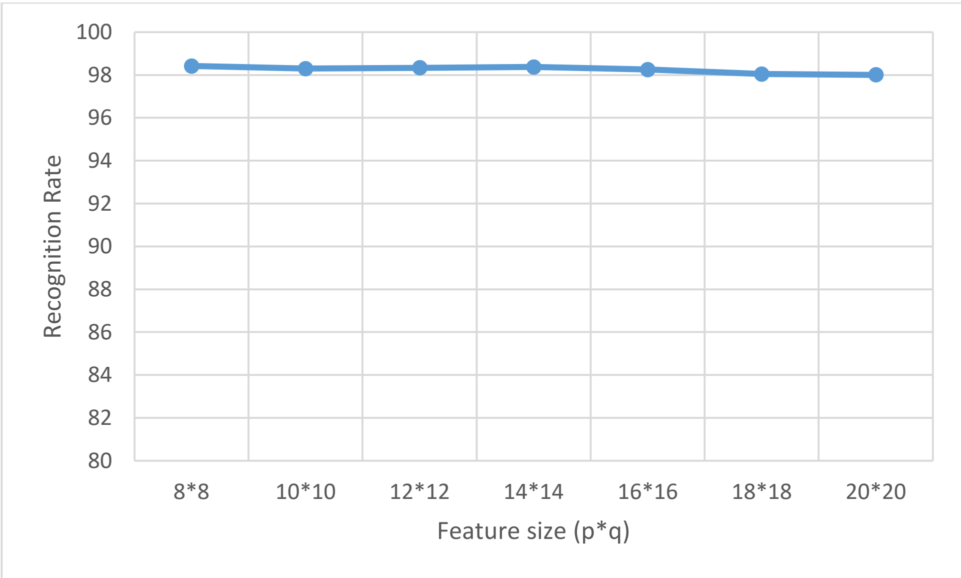

Fig. 4.2.5. Average recognition rate vs. feature space for

(Orl_tst_7_i, Orl_trn_3_i)

4.3. TABLE FOR EXPERIMENTAL RESULTS FOR THIS STUDY

From these scenarios, we can get the average recognition ratios as-

TABLE 4.3.1. Experimental results for various pairs of Training and Test sets

Training and Test Sets	Average Recognition rate
Orl_trn_5 , Orl_tst_5 Orl_tst_5, Orl_trn_5	97.925
Orl_trn_4, Orl_tst_6	96.76088
Orl_trn_3, Orl_tst_7	93.4425
Orl_tst_6, Orl_trn_4	99.5675
Orl_tst_7, Orl_trn_3	98.4175

Thus from those representations, we get that the average recognition ratio is maximum when we consider Orl_tst_6 and Orl_trn_4. It is slightly less when we consider Orl_tst_7, Orl_trn_3 and lesser when we consider Orl_trn_5, Orl_tst_5. The recognition ratio is smallest when we consider Olr_trn_3, Orl_tst_7. Now these results must by more accurate and purer as entropy is inserted here, which is multiplied with the mathematical formulae of the between-class and within-class scatter matrices. Thus the optimal projections obtained from those matrices will be more accurate and represent the real time scenario here.

4.4. COMPARISON OF RECOGNITION RATES WITH DIFFERENT METHODS

On the basis of AT&T database, a tabular form is constructed with various pairs of training and test sets. The results are given below.

TABLE 4.4.1. Comparison of recognition rates with various methods

METHOD	AVERAGE RECOGNITION RATES				
	s=3	s=4	s=5	s-=6	s=7
WFG-2DFLD $\beta = 0.4$	93.25 (14*14)	96.16(14*14)	98.00(14*14)	98.91(18*18)	98.82(16*16)
$MMSD(\beta = 0.4)$ [15]	90.23 (39)	---	95.89 (39)	---	98.72 (39)
$MSD\ (\beta = 0.4)$ [15]	89.15 (39)	---	94.68 (39)	---	98.51 (39)
G-2DFLD [9]	92.82(16*16)	95.94(16*16)	97.98(14*14)	98.72(14*14)	98.42 (8*8)
F-2DFLD [9]	92.08 (56*3)	95.04 (56*3)	---	---	---
G-2DFLD(ASSOCIATED WITH ENTROPY)	93.4425 (16*16)	96.76088(16*16)	97.925 (16*16)	99.5675(14*14)	98.4175 (8*8)
2DFLD [9]	92.30 (112*16)	95.08 (112*16)	97.50 (112*14)	98.26 (112*14)	
FUZZY FISHERFACE [12]	82.32 (39)	88.67 (39)	---	---	---
2DPCA [9]	91.27 (112*16)	94.33 (112*16)	96.83(112*14)	97.72 (112*14)	97.79 (112*8)

The above table signifies the comparative study of average recognition rates of various methods. Since the values for all *s* can't be obtained in each procedure, thus it is only an empirical study of the recognition rates; but this rough study is a depiction of the values obtained in the studies so far. From the above chart, we find that- the Fuzzy Fisherface approach delivers the least values of recognition rates. As per the studies went further, the primitive methods were upgraded and better values for recognition rates are obtained eventually. MSD method, MMSD method and 2DPCA method gives better results than the primitive procedures; but it is noticeable that, whenever the 2DFLD method comes, better results are obtained. All the 2DFLD methods give high values of average recognition rate. In our proposed study, since entropy is inserted in our study; hence the feature extraction will be purer and that will affect the values of the recognition rates. In this entire chart, the G-2DFLD method associated with entropy row gives the best result for average recognition rates.

CHAPTER 5

CONCLUSION

The goal of this project was to insert the concept of a thermodynamical term 'Entropy' in the field of image processing and face recognition. We know that the term entropy is known as the measurement of Molecular Randomness of a system. Then what was the significance to insert a term of Physics in the sector of computer science? Like Thermodynamics, it also measures the randomness or disturbance of the system while the feature extraction procedure. While using generalized 2DFLD method, we have considered the scatter matrices to achieve the optimality in feature extraction and calculating fisherfaces. But we haven't considered the disturbances which can affect the scatter matrices value and hence, the optimality. Here, we have inserted that particular term p_{ik} and hence, multiplied it with the scatter matrices. The presence of entropy will affect the recognition ratios for each training and test set pair and gives purer results corresponding to each pair. Now the term entropy can be inserted by numerous types of formulae. In this project, one formula is used and corresponding results are noted for a number of training and test files.

We considered five scenarios here, each scenario having one pair of training and test set. Now after collecting all the recognition rates corresponding to the feature spaces, the averages are calculated and the graphs are plotted (average recognition rate vs. feature space). We noticed that the average recognition rates are lowest when we considered (Orl_trn_3_i, Orl_tst_7_i) and then the rates increase gradually from (Orl_trn_4_i, Orl_tst_6_i) to (Orl_tst_6_i, Orl_trn_4_i). When we consider the last scenario of (Orl_tst_7_i, Orl_trn_3_i), we notice that the values are less than that of (Orl_tst_6_i, Orl_trn_4_i) and the first value is the highest value in this case.

In article 4.4, a comparative study is provided which actually gives the recognition rates for various values of s as well as from those values, the actual significance of inserting entropy in Generalized 2DFLD method can be clear.

CHAPTER 6

REFERENCES

[1] K. Fukunaga, Introduction to Statistical Pattern recognition, Academic Press, New York, 1990.

[2] M.D Kelly, Visual Identification of people by computer, *Tech rep. AI-130,* Stanford AI Project, Standford, CA, 1970.

[3] Shiladitya Chowdhury, Jamuna Kanta Sing, Dipak Kumar Basu, Mita Nasipuri, Face recognition by generalized two-dimensional FLD method and multi-class Support Vector Machines, 2010.

[4] T.Kanade, Computer Recognition of Human Faces, Birkhauser, Basel, Switzerland and Stuttgart, Germany, 1973.

[5] L.Wiskott, J.M.Fellous, C. Von Der Malsburg, Face recognition by elastic bunch graph matching, 1997.

[6] P.Penev, J.Atrick, Local Feature analysis: a general statistical theory for object representation, *Netw. Comput. Neural syst. 7*, 1996

[7] G. P. Teja and S. Ravi, “Face recognition using subspace techniques”, in *Proc. Int. Conf. on Recent Trends In Information Technology*, India, pp.103-107, 2012.

[8] J. Yang, D. Zhang, A.F. Frangi and J.Y. Yang, “Two-dimensional PCA: a new approach to appearance-based face representation and recognition”, *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 26, pp.131– 137, 2004.

[9] S. P. Bahurupi and D. S. Chaudhari, “Principal component analysis for face recognition”, *International Journal of Engineering and Advanced Technology*, vol. 1(5), pp. 91-94, 2012.

[10] J. M. Keller, M. R. Gray, and J. A. Givens, “A fuzzy k-nearest neighbor algorithm,” *IEEE Trans. Syst. Man Cybernet.*, vol. 15, no. 4, pp. 580–585, August 1985.

[11] W. Yang, J. Wang, M. Ren, L. Zhang, and J. Yang, “Feature extraction using fuzzy inverse FDA,” *Neurocomputing*, vol. 72, no. 13- 15, pp. 3384–3390, August 2009.

[12] H. Xiong, M.N.S. Swamy, M.O. Ahmad, “Two-dimensional FLD for face recognition, Pattern Recogn.”, 38 (2005) 1121–1124.

[13] F. Yang, M. Paindovoine, “Implementation of an RBF neural network on embedded systems: real-time face tracking and identity verification”, *IEEE Trans. Neural Netw.* 14 (2003) 1162–1175.

[14] J.C. Bezdek, J. Keller and R. Krishnapuram and N. Pal, “Fuzzy models and algorithms for pattern recognition and image processing”, *Kluwer Academic Publishers, Dordrecht,* vol. 4. 1999.

[15] B. Weyrauch , J. Huang, B. Heisele, and V. Blanz, “Component-based face recognition with 3D morphable models,” in Proceedings of the International Conference on Computer Vision and Pattern Recognition Workshops, July 2004.

[16] J.K.Sing, D.K.Basu, M.Nasipuri, M.Kundu, “Face recognition using point symmetry distance-based RBF network, Appl. Soft Comput.” 7(2007) 58-70.

[17] J.K.Sing, S.Thakur, D.K.Basu, M.Nasipuri, M.Kundu, "High-speed face recognition using self-adaptive radial basis function neural networks", *Neural Comput. Appld*. 18(2009) 979-990.
[18] R. Chellappa, C.L.Wilson, S.Siohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (1995) 705-740.
[19] The Database of Faces, AT&T Laboratories, Cambridge, U.K [online] Available http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.tar.Z