

Face recognition methods analysis

Shima Zarei^{a,1,*}

^a Università degli studi di Siena , Rettorato-via Banchi di Sotto 55, Siena and 53100, Italy

¹ shima.zarei@student.unisi.it*

* corresponding author

ARTICLE INFO

Article history

Received : 2018-07-14

Revised : 2018-07-19

Accepted : 2018-08-23

Keywords

Face Recognition

Eigen Face

Fisher Face

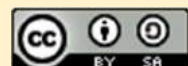
LBPH

PCA

ABSTRACT

Face Recognition is one of the most important issues in Image processing tasks. It is important because it uses for various purposes in real world such as Criminal detection or for detecting fraud in passport and visa check in airports. Face book is a nice example of Face recognition application, when it sends notification to one user's friends who are recognized by their images that user uploaded in face book page. To solve Face Recognition problem different methods are introduced such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Hidden Markov Models (HMM) which are explained and analyzed. Also algorithms like; Eigen face, Fisher face and Local Binary Pattern Histogram (LBPH) which are simplest and most accurate methods are implemented in this project for AT&T dataset to recognize the most similar face to other faces in this data set. To this end these algorithms are explained and advantages and disadvantages of each one are analyzed as well. Consequently, the best method is selected with comparison between the results of face reconstruction by Engine face, Fisher face and Local binary pattern histogram methods. In this project Eigen face method has best result. It should be noted that for implementing face recognition algorithms color map methods are used to distinguish the facial features more precisely. In this work Rainbow color map in Eigen Face algorithm and HSV color map in Fisher Face algorithm are utilized and results shows that HSV color map is more accurate than rainbow color map.

This is an open access article under the [CC-BY-SA](#) license.



1. Introduction

Face Recognition is a task of recognizing face [1] in an image by comparing the discrimination of features of a face in an image with other faces features inside other images. This task is so interesting not just in mathematical equations but also in practical aspect. Face Recognition is a fascinating issue which is used in different practical applications. As an example detecting faces in different pictures is so prominent, because the impact of many enhancement and noise reduction methods corresponds on the picture content. Face recognition [18] is a complex process because of fact that its computational model is difficult and faces are complex and multidimensional and meaningful visual stimuli. Therefore, Face Recognition [1] is a very high level task in terms of computational approaches. The goal of this project is to create a computational model for Face Recognition that is quite fast and reasonable and accurate.

The most prominent issues correspond to face and facial features detection are as follows:

- Intensity: Fundamentally intensity is categorized into 3 type; type-binary, color and gray.

- Pose: Because of variation of pose in image, main features especially eyes may appear completely or partially obstructed or closed. Structural Components: Some main objects
- glasses or beards and mustaches may or may not present always in an image.
- Image Rotation: Face in images may rotate in different directions or fluctuate. Poor
- Quality images: Images which contains noise or are blurred or disordered can cause to low quality face recognition result.
- Facial Expression: The facial expression may differ for various persons.
- Unnatural Intensity: 3-Dimensional face like rendered face, animated movie have unnatural intensity.
- Occlusion: Objects such as scarf, hand may cause to occlusion partially or fully.
- Illumination: The position of light source may cause to variation in face recognition result.

Generally, algorithms like Eigen face and Fisher face which uses Principal Component Analysis (PCA) technique [15] for face recognition are among the most effective method for this task [6]. It is noted that Local Binary Pattern Histogram (LBPH) algorithm is so useful for Face recognition because it is so accurate for even database with low amount of image and it is so sensitive to features discrimination and illumination of images. Moreover, Support Vector Machine (SVM) method is also one of the most effective techniques for face recognition, because it can divide the facial features in different classes by hyper-planes very good. In the next part different algorithms for face recognition in images are explained precisely [19].

2. The Proposed Algorithms

2.1. Eigen Face Method

Principal Component Analysis (PCA) [9] is a method in which it transforms the face images into a set of characteristics feature images which is called Eigen faces that are principal components of the initial training of data. In this way test image is projected in lower-dimensional face space and classified by a classifier or statistical theory. It should be mentioned that PCA method [10] is used to find directions with greatest variance which are called as principal components. This method is used in Eigen face method to detect the face. Equation (1) indicates the Mean value calculation related to Eigen face algorithm [11]:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

And Equation (2) implies calculation of Co-variance matrix S:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (2)$$

Also equation (3) indicates computation of Eigen values λ_i and Eigen vectors v_i of Co-variance matrix:

$$Sv_i = \lambda_i v_i, \quad i = 1, 2, \dots, n \quad (3)$$

And consequently K principal components of vector x are calculated in equation (4):

$$Y = W^T(x - \mu), \quad W = (v_1, v_2, \dots, v_k) \quad (4)$$

Equation (5) indicates the reconstruction of input from PCA:

$$x = WY + \mu \quad (5)$$

Then Eigen face algorithm first projects all training samples into PCA subspace and then project query image into the PCA subspace. In the last step it finds the nearest neighbours between the

projected training samples and projected query image. It should be noted that if there are M images in data set then N training samples are considered by Eigen face algorithm where number of N is lower than number of M. Actually, this method makes Eigen faces in training set of data with PCA mathematical process. The Eigen faces which are created are consist of dark and light areas that are arranged in a specific pattern. This pattern indicates the way that different features of face are evaluated and scored singularly. Then this algorithm try to detect the differences between features of Eigen faces in training set and faces in sample set of data [6]. While Eigen face algorithm is a powerful method and detects a linear combination of features which maximizes the total variance but it may lost a lot of discriminative information when does not consider components. Also because component that finds by PCA method do not contains discriminative features, the classification become impossible. To solve this problem a Linear Discriminative Analysis (LDA) performs a class specific dimensional reduction which detects the combination of features that separates the best between classes the linear discriminant Analyses the ratio of between-classes to within-classes scatter, instead of maximizing the overall scatter. The idea is that same classes are clustered together and various classes are as far away as possible from each other in the lower-dimensional representation [6].

2.2. Fisher Face Method

Linear Discriminative Analysis (LDA) [20] is a method of Face recognition which extracts features of faces and also project faces into lower-dimensional space. It cause lower computational complexity which avoids any iterative search or computation. In this method feature computation can be done at high speeds and recognition may be done in almost real time. In this way initially a training set which is composed of large set of face with various features is needed. It should be noted that the appropriate selection of training set has great impact on the validity of the result. Moreover, the dataset should contains several face image of one subject in training set and at least one image in sample set which represent different frontal views of subjects with minor variations in view angle. They should also contains different features such as lighting and background conditions and examples with and without glass [3]. Algorithm which use LDA method is Fisher Face algorithm. In Fisher face algorithm random vector of sample drawn from c classes is considered and scatter matrices S_b , S_w are calculated. Equations (6) and (7) implies this computation:

$$S_b = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (6)$$

$$S_w = \sum_{i=1}^c \sum_{x_j \in x_i} (x_j - \mu_i)(x_j - \mu_i)^T \quad (7)$$

Equation (8) implies the mean value μ correspond to scatter matrices is the total mean:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (8)$$

and μ_i is mean class $i \in \{1, 2, \dots, c\}$.

Fisher face method looks for projection W that maximizes the class reparability criterion. Equation (9) implies this projection:

$$W_{opt} = \underset{W}{\operatorname{argmax}} \frac{|W^T S_b W|}{|W^T S_w W|} \quad (9)$$

And in equations (10) and (11) the problem is solved by general Eigenvalue problem:

$$S_b v_i = \lambda_i S_w v_i \quad (10)$$

$$S_w^{-1} S_b v_i = \lambda_i v_i \quad (11)$$

Other problem is that the number of samples is less than the dimension of input data. Therefore, the scatter matrix is singular. This problem is solved by implementing PCA method and projecting samples into (N-C) dimensional space. Then a LDA was performed on the reduced data, because scatter matrix is not singular anymore. Equation (12), (13) implies the optimization problem:

$$W_{pca} = \operatorname{argmax}_W |W^T S_T W| \quad (12)$$

$$W_{fld} = \operatorname{argmax}_W \frac{|W^T W_{pca}^T S_b W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|} \quad (13)$$

Finally equation (14) indicates the transformation matrix which projects data into (c-1) space:

$$W = W_{fld}^T W_{pca}^T \quad (14)$$

This algorithm is one of the best method for face recognition [3].

2.3. Local Binary Pattern Histogram

Local Binary Pattern Histogram (LBPH) describes only local features not all image in a high dimensional vector and concentrates on shape and texture. It exactly compare each pixel with its neighborhood in a way that take a pixel as a center and threshold its neighbours against; If the intensity of the center pixel is greater or equal to its neighbour then indicate it with 1 and if not set 0. Therefore, there is a binary number corresponding to each pixel [8]. Equation (15) indicates that algorithm is represented in mathematical form as following:

$$LBPH(x_c, t_c) = \sum_{p=0}^{p-1} 2^p S(i_p - i_c) \quad (15)$$

Where $(x_c - y_c)$ is central pixel with intensity i_c and i_p is intensity of the neighbor pixel. Equation (16) indicates the computation of S which is the sign function:

$$\begin{cases} +1 & \text{if } \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (16)$$

In fact this description can give a very useful details in images. LBPH [13] incorporates spatial information in the face recognition model in which it divides images into M local regions and extract the histogram from each one. In this way the spatially enhanced feature vector is gained by concatenating the local histograms not merging them [8]. In order to implement the LBPH the input of model is a complete array of probability image and the output will be a complete feature vector of probability image. First of all LBPH [13] value of each pixel must be calculated which is defined as above. Secondly, the bin of histogram should be located and increment its result [8].

2.4. Independent Component Analysis Method

In ICA algorithm [16] second ordered and higher ordered statistics achieves and projects the input data into the basis vectors which are statistically independent. It is an unsupervised data that attempts to detect the original components by simple assumptions of statistical properties. Also the process is independent of each other and it use non Gaussian structure of data. The basic idea of ICA is as follows [5]: The observation variables are denoted as $x_i(t)$, $i = 1, 2, \dots, n$, $t = 1, 2, \dots, T$. Moreover, i is index of observed data variable and t is the time index, or some other index of different observations. The $x_i(t)$ are signals measured by a scientific device. They can be modeled as Linear combinations of hidden variables $s_j(t)$, $j = 1, 2, \dots, m$ with some coefficients a_{ij} [5]. Equation (17) shows these signal computation:

$$X_{it} = \sum_{j=1}^m a_{ij} s_j(t), \text{ for all } i = 1, 2, \dots, n. \quad (17)$$

The point is that the variable $x_{i(t)}$ is only observed, whereas both a_{ij} , $s_{i(t)}$ are to be estimated or inferred. The s_i are independent components whereas the coefficients a_{ij} are called the mixing coefficients. This estimation issue is referred as Blind source separation. This model can be indicated in different ways, but in typical model t is dropped and x_i , t_i are considered as random variables. Furthermore, x_i is usually into vector x with n dimension, the same is performed for the s_i and the

coefficients a_{ij} are collected into a mixing matrix A with size $n * n$. Equation (18) denotes the model [5]:

$$x = As \quad (18)$$

where x and s are random variables and A is a matrix of parameters. Also, it is possible to move to matrix notation where the observed $x_i(t)$ are collected into a $(n * T)$ matrix X , with i as the row index and t as column index and likewise for $s_i(t)$, equation (19) indicates:

$$X = As \quad (19)$$

A is still a same matrix as in previous equation. The main drawback of this model is that it can be made identifiable by making the unconventional assumption of the non Gaussian of the independent components. There are 3 conditions in which the model is identifiable under them [5]:

1) The components s_i are mutually statistically independent. Equation (20) implies that their joint density function is factorable:

$$P(s_1, \dots, s_m) = \prod_j P(s_j) \quad (20)$$

2) The components s_i have nonGaussian distributions.

3) The A matrix of component is square and invertible.

Most of ICA algorithms [16] divide the estimation of the model into two steps: a preliminary whitening and the actual ICA estimation. Whitening means that the data are initially linearly transformed by a matrix V such that $z = Vx$ is white. Equation (21) is an example of whitening:

$$\frac{1}{T} Z Z^T = I \quad \text{or} \quad \frac{1}{T} \sum_{t=1}^T z(t) z(t)^T = I \quad (21)$$

where I is the identity matrix. Matrix V can be computed by PCA method in a way that normalizing the principal components to unit variance is one way to whitening data but not the only one. After whitening to reduce the number of free parameters in model estimation of the mixing matrix to the space of orthogonal matrices is performed [5]. After this two steps PCA still holds:

$$Z = VX = VA \quad S = \hat{A}S \quad \text{or} \quad z = \hat{A}s \quad (22)$$

Where with refer to equation (22) the transformed mixing matrix $\hat{A} = VA$ is now orthogonal. Therefore, after whitening, constrain estimation of mixing matrix to orthogonal space will cause to reduction of number of parameter in model. Consequently numerical optimization in faster and more stable in orthogonal matrices space than general space of matrices space. Equation (23) noted that the objective function is created by inverse of $\hat{A}$, whose rows are denoted by W_i^T [5]:

$$L(W) = \sum_{i=1}^n \sum_{t=1}^T G_i(W_i^T z(t)) \quad (23)$$

where G_i is the logarithm probability density function of s_i , or its estimate $W_i^T Z$. Moreover, function $G(u) = -\text{Logcosh}(u)$ which is a smoothed version of the negative absolute value function $|u|$, works well in many applications. After estimating PCA, it would be very useful to assess the reliability or statistical significance of the components [5].

2.5. Support Vector Machine

Support Vector Machine (SVM) is recognized as a very powerful method for general aims of pattern recognition. SVM detects the hyperplane that separates the largest fraction of points correspond to similar class on the same side, while maximizing the distance from either class to the hyperplane, this hyperplane is called Optimal Separating Hyper-plane (OSH). OSH reduce misclassifying not only the examples in the training set but also the examples of the test set. Face Recognition application train a model with SVM where the discrimination between face and non-face classes are considered with thousands of examples. Because the difference between two classes of

objects is significant, the discrimination between them is an easy task. Moreover, it is proven that even by applying simpler algorithm the same accurate result can be obtained as SVM [4]. In SVM method there are several kernels, but the main four kernels are: Linear, Polynomial, Sigmoid and Radial Basis Function RBF. Also, SVM with RBF kernel is selected for Face Recognition due to fewer numerical difficulties. Equation (24) indicates the RBF kernel:

$$K(x_i, y_i) = \exp(-\lambda \|x_i - x_j\|^2), \quad \lambda > 0 \quad (24)$$

where λ is a kernel parameter and parameterized using: $\lambda = \frac{1}{2\sigma^2}$ [19]. Then representation of training set is changed by mapping data to a feature space F that has OSH. In this case of issue study is limited to two-class discrimination and with consideration of training set S with one vector of features of n dimension, where each data x_i is belong to the one of two class identified by label 1 or -1. Equation (25) shows that this problem is addressed by solving the quadratic optimization problem with linear constraints that can be interpreted in terms of Lagrange multipliers computed by quadratic programming:

$$\max(\alpha_i) \tilde{l}(\alpha) = \sum_{i=1}^n \alpha_i \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (25)$$

α_i are Lagrange multipliers parameters to be adjusted, c is the penalty parameter of the classification error term it must be adjusted because the data are rarely completely separable, x_i are training examples. Also, in equation (26) the solution of optimization problem is a vector $W \in F$, that can be written as a Linear combination of the training inputs:

$$W = \sum_{i=1}^n \alpha_i y_i w_i \quad (26)$$

(w, b) define the hyperplane $OSH = x: wx + bc = 0$ and b is the bias. Separating OSH is used once we have trained it on the training set, the OSH divides R^n into two regions: One where $w_i + b \geq 0$ and one where $w_i + b < 0$ [19]. However, Face Recognition is a multi-class classification problem. There are two methods for face recognition by SVM: One _against_ one and one _against_ all. The One _against_ one method is classification between each class and all the rest classes. In this experiment one _against_ all is used. Let's explain it briefly; In this method n number of SVM is trained which everyone separates a single class from all remaining classes. In base of comparison of several multi-class techniques favors the one _vs_ all approach because of its simplicity and excellent classification performance. In terms of training effort one _vs_ all with n SVM is preferred rather than one _vs_ one approach with $n(n-1)$ SVM_s [19]. The procedure of generating n -class classifier using two _class_ discrimination methods is as follows:

Equation (27) construct n two-class decision functions $d_k(x), k = 1, \dots, n$ which separates examples of class K from the training points of all other classes:

$$d_k(x) = \begin{cases} +1 & \text{if } x \text{ belongs to class } k \\ -1 & \text{otherwise} \end{cases} \quad (27)$$

The face AT&T database is consist of 10 images per each person in which 5 of them are considered for training set and the rest 5 ones are set as test set of data. Five image of individual are signed as positive and all other ones are signed as negative samples. Both positive and negative samples were taken as input samples in order to train a SVM classifier for achieving related support vectors and optimal hyperplane. Then SVM was labeled as SVM_1 and in turn we can allocate SVM for every individual and label them as $SVM_1, \dots, SVM_n$ respectively [19]. When samples are enter to each SVM there should be a several cases:

- If the sample was considered as positive by SVM_i and to be negative by other SVM_s at the same time, then the sample was classified as class i .
- If the sample was considered to be negative by several SVM_s and synchronously to be positive by other SVM_s then the classification was false.

- If the sample was considered to be negative by all SVMs, then the sample was decided not belonging to the face datasets [19].
- However, the study indicated that SVM is useful as a classifier with combination to other methods such as PCA [11], LDA [11], ICA and LBPH [19].

2.6. Hidden Markov Model

Hidden Markov Models (HMM) [14] are a set of statistical models used to characterized the statistical properties of the signal. There are two main inter correlated process in HMM; First one is an underlying, unseen Markov Chain with a finite number of states, a state transition probability matrix and an initial state probability density functions correspond to each state [7]. Also, the elements of HMM are:

- N , number of states in the model. If S is the set of states, then $s = s_1, \dots, s_N$. The state of model at time t is given by $q_t \in S, 1 \leq t \leq T$, where T is the length of observation sequence.
- M is number of different observation symbols. If V is the set of all possible observation symbols then $V = \{v_1, \dots, v_n\}$ [7].
- A is a set of state transition probability matrix, i.e, $A = \{a_{ij}\}$ where $a_{ij} = P[q_t = s_j | q_{t-1} = s_i], 1 \leq i, j \leq N$, with the constraint, $0 \leq a_{i,j} \leq 1$ and $\sum_{j=1}^N a_{ij} = 1, 1 \leq i \leq N$, b is the observation symbol probability matrix, i.e $B = \{b_j(k)\}$, where $b_j(k) = P[O_t = V_k | q_t = s_j]$, and O_t is the observation symbol at time t .
- Π the initial state distribution, i.e $\pi = \{\pi_i\}$, where $\pi_i = P[q_1 = s_i], 1 \leq i \leq N$.

In a short form HMM is written as $\lambda = (A, B, \Pi)$ triple [7]. The above equation is correspond to discrete HMM where the observation are the discrete symbols that are selected from a finite alphabet $V = \{v_1, \dots, v_M\}$. In a continuous density HMM, the states are characterized by continuous observation density functions. Equation (28) implies, the most general model of the model probability density function (pdf) which is a finite mixture of the form:

$$b_i(O) = \sum_{k=1}^M c_{ik} N(O, \mu_{ik}, U_{ik}), \quad 1 \leq i \leq N \quad (28)$$

where c_{ik} is the mixture coefficient for the k .th mixture in state i . Without loss of generality $N(O, \mu_{ik}, U_{ik})$ is assumed to be a Gaussian pdf with mean vector μ_{ik} and covariance matrix U_{ik} [7]. HMM [14] also is used mainly for Face Recognition. For frontal face the significant facial features like hair, forehead, eyes, nose, and mouth come in a natural order from top to bottom, even if images are taken under small rotations in the image plane and/or rotations in the plane perpendicular to the image plane. Each of this facial regions is assigned to a 1 dimensional continuous HMM. Each face image of width W and height H is divided into overlapping blocks of height L and width W . Also the amount of overlap between consecutive blocks is P and number of blocks which extracted from face image is equal to number of observation vector T . Equation (29) shows this vector [7]:

$$T = \frac{H-L}{L-P} + 1 \quad (29)$$

3. Method

In this project aim is to recognize the face which is the most similar one to other faces in AT&T data set. To this end 3 main algorithms which are common and simpler and faster than other algorithms for face recognition are implemented on AT&T dataset. Formally ORL dataset is consist of face images captured between April 1992 and April 1994. This dataset is used for collaboration between speeches, vision and Robotics groups of Cambridge University Engineering Department. This dataset contains 10 images corresponded to 40 different faces. Some of face images are taken in different times with different lighting, facial expressions such as open/close eyes, smile/non-smiles and facial details like glasses/no glasses. All the images are captured against a dark homogeneous background with the subjects in an upright and frontal position. The size of each image is 92×112 pixels, with 256 grey levels per pixel and are in PGM format. The images are organized in 40 directories, which have names of the form sX , where X implies the subject number between 1 and 40. In each of these

directories, there are 10 different images of that subject, which have names of the form Y.pgm where Y is the image number for that subject between 1 and 10. Example of AT&T dataset is shown in Fig.1 and different intensity, facial features are explicit like various pose, occlusion, facial expression, poor quality, image rotation, unnatural intensity and structural components. For instance, there are picture of one person in different pose and also with some occasion like closed eye and also there are different structures components in these pictures like glass. It should be noted that in this data all picture have same illumination level but approximately same quality.

Moreover, in this project three algorithm such as Eigen Face algorithm, Fisher Face algorithm and LBPH algorithm [17] are implemented with C++ code and OpenCv library. These algorithms are selected due to the simplicity of them. Also in comparison to other algorithms which are indicated in this paper these 3 methods have lower complexity. It should be considered that simplest algorithms have lower calculation and time complexity rather than complex algorithms. For instance HMM is more complex than Eigen Face, Fisher Face and LBPH methods [17]. Also in this problem with use of a simple algorithm the good result is obtained as well. In this way the result of Eigen Face, Fisher Face and LBPH methods are illustrated and compared with each other and best one is selected in base of accuracy percentage, fig.1 presets the example dataset.



Fig. 1. Example of dataset

3.1. Eigen Face Method Implementation

First of all Eigen Face algorithm [11] implemented by help of OpenCv library [2] and C++ language. In the project code face module is used for recognizing face object in an image. In this project regarding to Eigen face method Rainbow color map [2] were applied on image dataset AT&T that is indicated in Fig.2. It shows how gray-scale values and illumination were distributed within the specific Eigen faces. This color map indicated the features of each image more precisely to distinguish the differences between images features. This led to recognition of the exact face which were in training set and test set and is the most similar face to other ones. So this face is recognized with comparison of all face features in different pictures with each other and then reconstructed with refer to those features. Also the result of image reconstruction from lower dimensional approximation by Eigen face algorithm after recognizing the face is illustrated in Fig.3. It indicated that the face were reconstructed from its lower dimensional level. For a good reconstruction a plot with 10, 30,..., 310

Eigen face is considered. It should be noted that 10 Eigen vector were not sufficient for a good image reconstruction, but 50 Eigen vectors were sufficient to encode important facial features. Regarding to AT&T dataset 300 Eigen vector is sufficient to achieve a good reconstruction result. In order to determine how many Eigen face required to generate a good result from dataset rule of thumbs used, but it strongly depended on the input data.

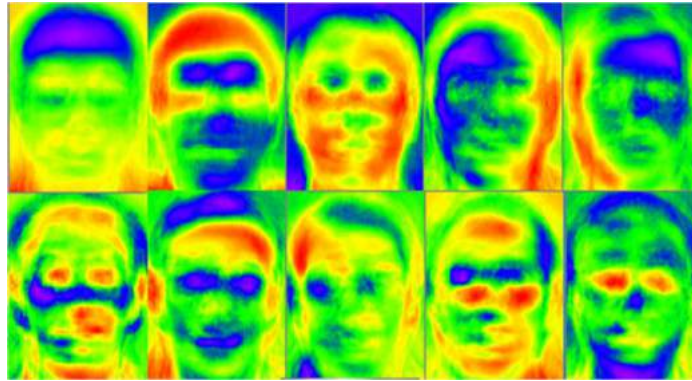


Fig. 2. Rainbow color map



Fig. 3. Eigen face reconstruction result

3.2. Fisher Face Method Implementation

In spite of the fact that Eigen face algorithm is so fast but Fisher face method [12] is more accurate and more reliable. Therefore, to evaluate fisher face method let's implement it. It should be noted that this algorithm like Eigen face algorithm try to find the face that is in both train set and test set of data and is the most similar one to other faces in AT&T dataset [19]. The Fisher face algorithm does not consider illumination as much as Eigen face method, instead it focus on facial feature discrimination which is calculated by LDA. It must be taken into account that Fisher face algorithm result is strongly depend on input data, if this algorithm is learned for well illuminated images and is tested for bad-illuminated data, then algorithm finds wrong components. This is because the model is not quite well to learn illumination [19]. In Fisher Face algorithm HSV color map were used for showing facial features more significant in Fig.5. Also Fig.4 indicates that Fisher Face method can reconstruct the image from lower resolution level. Because in dataset AT&T there were not so suitable images the result is not so obvious as much as it must be in terms of image reconstruction. In Fisher face method sample image was projected into each of Fisher faces [12]. Therefore, a good visualization will be

generated. Reconstruction of Fisher face is illustrated in Fig.5 and it indicated that Fisher face method is not suitable for AT&T dataset because it does not have good reconstruction.

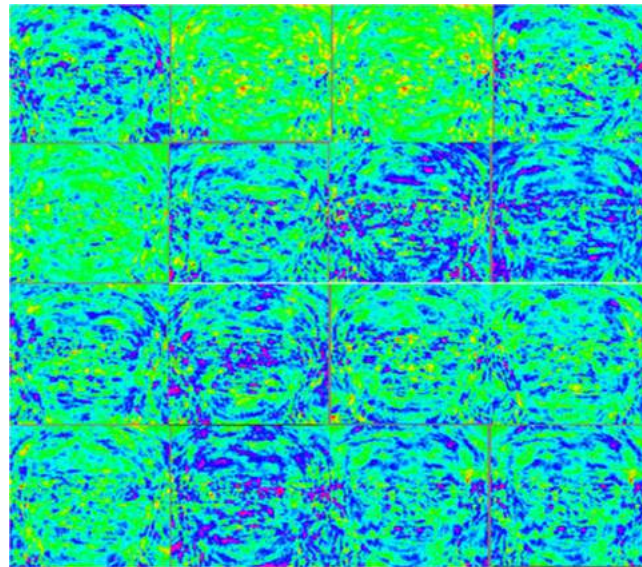


Fig. 4. HSV color map result



Fig. 5. Fisher fae reconstruction

3.3. Local Binary Histogram Method Implementation

It is indicated that Eigen Face method maximizes the total scatter which can lead to some issue if the variance is made by an external source, because these components are not useful for classification. Therefore, to maintain some important discrimination information a Linear Discriminant Analysis is applied and analyzed. It created Fisher Face method in which this method worked not appropriately for this project. Nevertheless, because it were not possible to guarantee perfect light setting or 10 image per each person in dataset the problem raised. Other problem were that if there were just one image per person how the issue could be solved. In order to get the best result it was important to have (8×10^4) images for each person, so Fisher face method in this project did not work well. The idea is to use LBPH method [13]. This method implemented because it had the most accuracy among other method for Face Recognition. This methods like Eigen face and Fisher Face methods aim to detect the most similar face to other faces in AT&T dataset. It were noted that LBPH method [13] is robust against monotonic gray scale transformations. Because there were not good data model for LBPH method in AT&T dataset we cannot expect a very good result and it just released some information about model as implies in Table 1.

Table 1. LBPH result

LBPH Result	Local Binary Pattern Histogram	
	Parameters	Values
	Radius	1
	Neighbours	8
	$Grid_x$	8
	$Grid_y$	8
	Threshold	0.00
	Size of histogram	16384

^a. This table indicates useful parameters for further processing of face recognition method

4. Results and Discussion

As it were mentioned there were different types of algorithm for face recognition task which were based on PCA [11] like Eigen face [11] and Fisher face [12] methods and other algorithms which were base in LDA [20] and also LBPH method which used Local features in its Binary Pattern Histogram [17] to find the face object in images or SVM [15] method which used Hyper-plane to separate the class of features of a face. Nevertheless, it should be taken into consideration that selecting the best method of Face recognition task were dramatically related to type of images of dataset. As a fact for images with illumination Eigen face method were among the best one, because it considered illumination and extracted face features which were important in face recognition task. Also the main benefit of this method were that, it could trim down the data required to recognize the entity to 1/1000 of the data existing. Fisher face method had advantages over Eigen face method because some important features of images could lost by Eigen face method while it compressed the data, but Fisher face method by applying LDA preserved some of important facial features. The disadvantages of Eigen face and Fisher face methods were that perfect light could not be guaranteed for 10 images of person and also if there were one image for each person these methods cannot work. Therefore, to come up with his problem LBPH method were introduced which were more suitable for database with low amount of images per each individual [19].

5. Conclusion

In conclusion, the task of face recognition is related to dataset images and consequently choosing the suitable algorithm to implement the application. In this work with consideration of AT&T dataset, the most common methods are selected and implemented like; Eigen Face method, Fisher Face method, LBPH method. Also, the best result in this case is corresponded to Eigen Face method with more than 96% accuracy. The result of Eigen Face method is indicated in Fig.3 and the result of Fisher Face method is illustrated in Fig.5 which implies that Eigen Face method were more accurate for face recognition than Fisher Face method in this project. It should be noted that HSV color-map in Fig.4 is more accurate than Rainbow color-map in Fig.2 in terms of feature face discrimination recognition. In addition, LBPH method in this project could not give very good result due to lack of appropriate images of dataset for this method [19].

References

- [1] Kong, Seong G and Heo, Jingu and Abidi, Besma R and Paik, Joonki and Abidi, Mongi A, Recent advances in visual and infrared face recognition a review, Computer Vision and Image Understanding, Elsevier, vol.97, pp.103-135, 2005.
- [2] G. Bradski and A. Kaehler. Learning OpenCV: Computer vision with the OpenCV library. " O' Reilly Media, Inc.", 2008.
- [3] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. JOSA A, vol.14, pp.1724-1733, 1997.
- [4] G. Guo, S. Z. Li, and K. Chan. Face recognition by support vector machines. In Automatic Face and Gesture Recognition, Proceedings Fourth IEEE International Conference on, IEEE, pp. 196-2011, 2000.

-
- [5] A. Hyvärinen. Independent component analysis: recent advances. *Phil. Trans. R. Soc. A*, vol. 371, pp.20110534, 2013.
 - [6] P. Navarrete and J. Ruiz-del Solar. Analysis and comparison of eigenspacebased face recognition approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, vol.16, pp.817-830, 2002.
 - [7] A. V. Nefian and M. H. Hayes. Hidden markov models for face recognition, In *Acoustics, Speech and Signal Processing*, 1998. Proceedings of the 1998 IEEE International Conference on, IEEE, vol.16, pp. 2721-2724, 1998.
 - [8] N. Stekas and D. van den Heuvel. Face recognition using local binary patterns histograms (lbph) on an fpga-based system on chip (soc), In *Parallel and Distributed Processing Symposium Workshops*, IEEE International, IEEE, pp.300–304, 2016.
 - [9] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, vol.3, pp.71-86, 1991.
 - [10] Zhao, Wenyi and Krishnaswamy, Arvinth and Chellappa, Rama and Swets, Daniel L and Weng, J, Discriminant analysis of principal components for face recognition, *Face Recognition*, pp.73-85, 1998.
 - [11] Belhumeur, Peter N and Hespanha, Joao P and Kriegman, David J, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, 1997.
 - [12] Cevikalp, Hakan and Neamtu, Marian and Wilkes, Mitch and Barkana, Atalay, Discriminative common vectors for face recognition, *IEEE Transactions on pattern analysis and machine intelligence*, vol.27, pp.4-13, 2005.
 - [13] Shan, Caifeng, Learning local binary patterns for gender classification on real-world face images, *Pattern recognition letters*, vol.33, pp.431-437, 2012.
 - [14] Nefian, Ara V and Hayes, Monson H, Maximum likelihood training of the embedded HMM for face detection and recognition, *Image Processing, Proceedings. 2000 International Conference on*, IEEE, vol.1, pp.33-36, 2000.
 - [15] Deniz, Oscar and Castrillon, M and Hernandez, Mario, Face recognition using independent component analysis and support vector machines, vol.24, pp.2153-2157, 2003.
 - [16] Bartlett, Marian Stewart and Movellan, Javier R and Sejnowski, Terrence J, Face recognition by independent component analysis, *IEEE transactions on neural networks a publication of the IEEE Neural Networks Council*, Howard Hughes Medical Institute, vol.13, pp.1450, 2002.
 - [17] Ahonen, Timo and Hadid, Abdenour and Pietikainen, Matti, Face description with local binary patterns: Application to face recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, pp.2037-2041, 2006.
 - [18] Zhao, Wenyi and Chellappa, Rama and Phillips, P Jonathon and Rosenfeld, Azriel, Face recognition: A literature survey, *ACM computing surveys (CSUR)*, ACM, vol.35, pp.399-458, 2003.
 - [19] A. Bouzalmat, J. Kharroubi, and A. Zarghili, Comparative study of pca, ica, lda using svm classifier, vol.2, *Journal of Emerging Technologies in Web Intelligence*, pp.64-68, 2014.
 - [20] Yu, Hua and Yang, Jie, A direct LDA algorithm for high-dimensional data—with application to face recognition, *Pattern recognition*, Pergamon, vol.34, pp.2067-2070, 2001.
-