Face Recognition With Pre-Moment Processing

Abdelatif Hussein A.ALI,

High Institute of Computers and
Information Technology,
Computer Dept.,
El-Shorouk Academy, Cairo, Egypt,
E-mail: aabouali@hotmail.com

Abstract

Face recognition is an active research area because of the importance of its application areas. Face recognition as almost all digital image processing based systems begins with elementary processing on the acquired images. The elementary processing is intended to add more to the abilities of consequent processing steps. The elementary processing phase is in most of the cases processing the image by some filters. In this study, an elementary phase of Sobel derivative filters and the derivatives followed by thresholding by Otsu's method effect on features extracted by moments for face recognition is explored. The preprocessing step with this setup aims to highlight face local features. The moments used in the study are Hu and Legendre. Feed forward neural network used as the classification facility. The results of the study indicated that edges do not form a major differential component in the values of moments. Consequently, moments could be aided with feature vectors that focus on edges. Also, the study indicated that Legendre is superior compared to Hu and the union of Hu and Legendre increases correct recognition probability.

Keywords: **Face Recognition; Biometric Identification; Feature extraction; Image processing; Derivative filters; Legendre moments; Hu moments;**

1. Introduction

Face recognition is an eminent research area in digital image processing. The importance of face recognition comes from the fact that it impacts many applications such as security, robotics, human-computer-interfaces, digital cameras, games, entertainment, authentication,

intelligence image indexing, and as well it feeds other applications with methodologies and algorithms. Face recognition is one of the most visible. In fact, Humans, as well as many of the animals, do it perfectly. However, it is yet a challenging problem in computer vision and pattern recognition. Face detection and recognition is anonymous. The subject face is not matched against a database of known previously stored individuals. In contrast, face detection and recognition is the science of facial features locating, extraction, and classifications. The challenges in face recognition include the variations in scale, illumination, pose, occlusion, clutters, orientation, expression variations, and computational complexity [1]. The popular approaches used for face recognition are based on i) structural ii) statistical and iii) neural networks.

In structural, facial main features such as eyes, nose, mouth and chin their attributes such as: areas, relative distances and angles between them are used [2][18]. In general, such methods use local features as well as the relationships among them. The second approach is statistical, basically uses image transforms where features are extracted from the whole images. So, the features base is global instead of local. The Principal Component Analysis (PCA), multi-linear principal component analysis (MPCA) and linear discriminant analysis (LDA) [2-9] are examples of statistical approach. The third approach uses neural networks for feature extraction as well as recognition but the training time needed to learn raw images is very high beside the complexity and scalability of the network needed[10]. So, feature extraction is required ahead of the neural network [11-13].

Moments are from the famous and the earliest methods used for features extraction from images. Moments, as feature extraction mean, made the neural approach realistic [14-16]. Image moments that are invariant with respect to the transformations of scale, translation and rotation applications include pattern recognition, object identification and template matching. Orthogonal moments have additional properties of being robust in the presence of image noise and have a near zero redundancy measure as a feature set.

The efficiency of a face recognition algorithm depends on: (i) an efficient invariant feature representation with respect to illumination, scaling, rotation, and pose (ii) a classification technique that maps the feature vectors into their appropriate classes with minimal misclassification. (iii) the

time complexity of the algorithm together with mapping ability to parallel platforms.

Image processing system, normally, begins with the application of some filters. Filters are one of the most famous and widely used tools in images processing. Filters could be applied to digital images both in spatial and frequency domains. They are used in image processing for several purposes from these image enhancement and edge detection. The enhancement filters target better perception and enables correct consequent processing steps such as image segmentation. Derivative filters used, mainly, to highlight changes in the image that could be used to enhance edges or directly used in the image segmentation to objects.

This study focuses on the effect of preprocessing of faces ahead of features extraction using moments. The preprocessing step of the study is using the well-known Sobel derivative filters and these filters followed by thresholding based on Otsu's method [17]. The moments used in the study are HU and Legnedre [4-16] to extract global image features. The feed forward neural network used as a classifier based on the extracted features.

The remaining of this paper is organized as following: Section two outlines the face recognition methodology under the study, and summaries the used moments. Section three contains tests and results. Section Four is the conclusion of the study.

2. Face Recognition Using Pre-Moments Processing

The recognition methodology under the study operates in two phases. During the first phase, which is called the training or learning phase, the system learns the recognition. In the second phase, which is called the operation or testing phase, the system recognizes unknown faces classes. The system has to pass successfully the first phase to be usable as a recognition facility. The success of the first phase depends on many parameters that include: selection of a proper set of features, neural network structure, learning algorithm and its parameters, and the selection of the training dataset. From the famous problems in neural classifiers are the over-fitting, inability to converge to acceptable solutions, stuck to local

minima, and incorrect generalization. In the training phase, the set of faces that are the subjects of the recognition run through three stages: a preprocessing, feature extraction, and supervisor class code assignment to form the neural network training dataset. Then, the training dataset is applied to the neural network for the iterative supervised training process. By the end of the training process, the network will be, hopefully, ready to operate in the recognition mode. Normally, the available data is engineered to training and testing data sets. The testing data set is used to verify the network performance in the generalization ability. In the operation phase, an unknown face image runs through three stages: a preprocessing, the feature extraction, and application of the features to the network for recognition of the face class. The output of this phase is faced the class signature. Figure (1) depicts the system under the study.
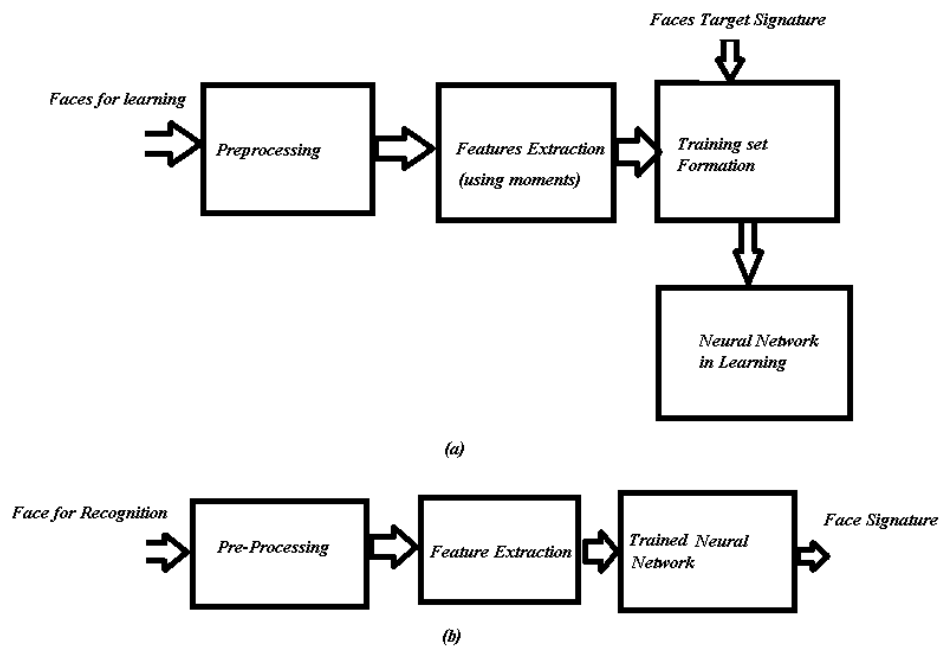
Figure (1) Overall moments based recognition system, (a) Training (b) Testing

Preprocessing

The preprocessing of the image aims at highlighting local facial features. The face images supplied assumed to be passed an enhancement step that is out of the scope of the study. The preprocessing used in this study are Sobel derivative filters (x-derivative and y-derivative), and Sobel derivative filters followed by Otsu's thresholding. The output from this step either gray values, the first two cases, or binary values, the last two cases. The study, which is a comparative, uses the raw images as a reference.

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Y-derivative   X-derivative
Figure (2) Sobel derivative filters

Sobel derivative filters are simple filters as shown in figure (2). Those filters are widely used in edges detection and enhancement. A typical application of these filters is depicted in Figure (3). Figure (3)-a is the original face. Figure (3)-b, c are the same of (a) processed by Sobel derivatives (Y,X) consequently.

Otsu's method for thresholding [26] searches for a threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes, equation (1):-

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \qquad\qquad (1)$$

Weights w_i are the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes.

The preprocessing step takes as an input enhanced face image matrix and produces one of the following matrices (a) Sobel-x derivative (b) Soble-y derivative (c) Sobel-x followed by Otsu's thresholding (d) Sobel-y followed by Otsu's thresholding. Figure (3) shows the outcomes of the preprocessing on the original face of (a), the Sobel filters are re-scaled to (0-255). The preprocessing step, as shown in figure, highlights the facial features areas.

Figure (3) Cases under the study: (a) Original (b),(c) Sobel derivative-x,y (d),(e) Sobel derivative-x,y followed by Otsu's thresholding

Features extractions

Feature extraction is a focal point in biometric recognition systems. The failure in features extraction or the wrong choice of the features leads to the failure of the recognition process. Feature extraction methods, in biometric systems, are either global or local. The global features are extracted from the whole image. So, the feature operators are applied to the entire image to extract the set of desired features which are used directly in the classification process. Local feature methods focus on chucks of the image. The operators, in this case, operate on windows of the image to extract local features. In this case, the classification process uses the windows location as well as the extracted local features. Feature extraction bases include colors, edges, corners, textures, statistical, and combinations of them. The feature extraction is done in spatial and frequency domains [19]. From the famous methods: moments [20], Principal Component Analysis (PCA), Scale Invariant Feature Transform (SIFT) [21], PCA-SIFT [22][23], and SURF (Speeded Up Robust Features)[22].

The study focus is the effects of preprocessing on moments performance. The preprocessing outputs, as shown in Figure (3), produces images more focused on facial local features. The moments used in the study are Hu and Legendre. These moments are summarized in the following:-

Hu moments

Hu moments based on the geometric moments are image texture descriptors. Geometric moments are also known as regular or raw moments. The geometric moment's equations (2) to (5) provide nonnegative values.

$$m_{p,q} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \qquad (2)$$

Where p,q =0,1,2,3,… The moment order is known as (p+q).

For digital image defined by digital function $f(x,y)$ of size M by N the moments becomes:-

$$m_{p,q} = \sum_x \sum_y x^p y^q f(x,y) \qquad (3)$$

The central moments relative to the central gravity are:-

$$\mu_{p,q} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x-\overline{x})^p (y-\overline{y})^q f(x,y)dxdy \qquad (4)$$

where

$$\overline{x} = \frac{m_{1,0}}{m_{0,0}} \qquad , \qquad \overline{y} = \frac{m_{0,1}}{m_{0,0}}$$

For digital image the moments become

$$\mu_{p,q} = \sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q f(x,y) \qquad (5)$$

The normalized central moments are defined as:-

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu^{\gamma}{}_{0,0}} \qquad where \qquad \gamma = \frac{p+q}{2}+1$$

The seven invariant moments are driven from the moments up to the third order moment as following:-

$$\phi_1 = \eta_{2,0} + \eta_{0,2} \qquad (6)$$

$$\phi_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}{}^2 \qquad (7)$$

$$\phi_3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2 \qquad (8)$$

$$\phi_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{2,1} - \eta_{0,3})^2 \tag{9}$$

$$\phi_5 = (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{3,0} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})$$
$$[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \tag{10}$$

$$\phi_6 = (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} + \eta_{0,3}) \tag{11}$$

$$\phi_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{1,2} - \eta_{3,0})(\eta_{2,1} + \eta_{0,3})$$
$$[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \tag{12}$$

These moments equations (6) to (12) are known to be invariant to translation, scale and rotation [17].

Legendre moments

Legendre moments are driven from Teague Legendre polynomial. The polynomial defined by equation (13).

$$P_n(x) = \sum_{k=0}^{n} \left\{ (-1)^{\frac{n-k}{2}} \frac{1}{2^n} \frac{(n+k)! x^k}{(\frac{n-k}{2})!(\frac{n+k}{2})! k!} \right\} \tag{13}$$

The polynomial recurrence relations are defined as in equations (14), (15)

$$P_0(x) = 1 \quad , \quad P_1(x) = x \tag{14}$$

$$P_n(x) = \frac{(2n-1)P_{n-1}(x) - (n-1)P_{n-2}(x)}{n} \tag{15}$$

Legendre moments [15][16] are orthogonal, consequently contains minimal redundancies and caries independent image characteristics. Two-dimensional Legendre moments of order (p+q) is defined by equations (16) to (18).

$$L_{p,q} = \lambda_{p,q} \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} P_p(x_i) \; P(y_j) \; f(i,j) \tag{16}$$

$$x_i = \frac{2i}{M-1} - 1 \quad , \quad y_j = \frac{2j}{N-1} - 1 \qquad (17)$$

$$\lambda_{p,q} = \frac{(2p+1)(2q+1)}{MN} \qquad (18)$$

The parameter $\lambda_{p,q}$ is known as the normalization constant. The Legendre moments define a complete orthogonal basis set over the interval [-1, 1] in which the moments are defined. The Legendre moments used in the study are $L_{2,0}, L_{0,2}, L_{1,1}, L_{2,1}, L_{0,3}, L_{3,0}, L_{4,0}$

Neural networks

The Artificial Neural Network or ANN algorithms are the commonly used as base classifiers in wide range of problems [24]. An artificial neural network is a powerful data modeling and information-processing paradigm that can capture and represent unknown-complex input/output relationships [25]. The advantage of neural networks mainly lies in that they are data driven self-adaptive methods, which can adjust itself to data without any explicit specification of functional or distributional form for the underlying model. Also, they are universal functional approximations and can approximate functions with an arbitrary accuracy [25], [26]. The function of the neural network is transforming inputs into meaningful outputs. It is inspired by the way of biological nervous systems, neural networks act, somehow, like a human brain in two stages learning stage and acting stage. Moreover, neural networks can represent both linear and non-linear relationships and the way it learns these relationships directly from the modeled data. The most common neural network model is the multilayer perceptron (MLP). This type of network is a supervised because it requires the desired output per pattern from the supervisor. The goal of this type of network is to create a model that correctly maps the feature vector of a single sample (input) to the class of the input sample (output) using historical data so that the model can then be used to produce the output when the desired output is unknown.

Input layer (7 neurons)

Output layer (5 neurons)

Hidden layer 1 (15 neurons)   Hidden layer 2 (10 neurons)
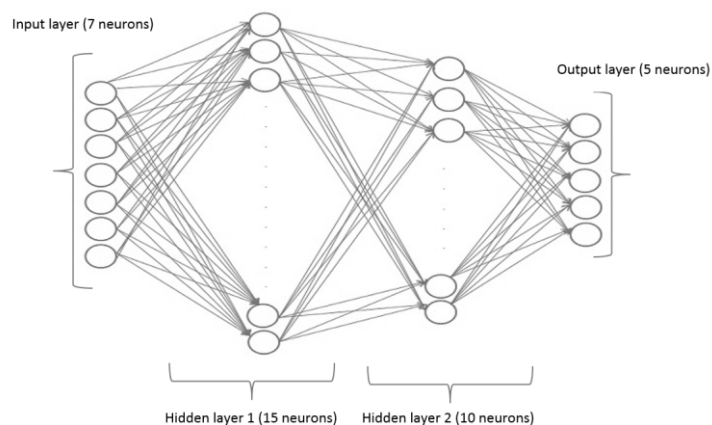
Figure (4) Neural Network Architecture

The well-known perceptron neural network consists of three types of layers input, hidden and an output layer. Each of them consists of a number of, perceptron, neurons. The layered neurons of a network connected according to the network architecture. Each connection has a very important unit called "weight". The weight unit preserves the strength of the link in the learned mapping functions of the neural network. The input layer is the layer that represents the input data so the length of this layer is equal to the length of the input data (feature vector), and there is only one input layer in the neural network. It consists of a set of input values (Xi) and associated weights (Wi). The hidden layer is the kernel of the network because it controls internal mapping to other spaces and by which the result gets better. The neural network may contain several hidden layers which represent the intermediate spaces. The last one is the output layer, which returns the result. The number of neurons in this layer coincides with the definition of classes signatures encoding. There is only one output layer in the neural network. The MLP neural network looks like any neural network so it operates in two modes. The first one is the learning mode, which trains or adapts the network to be able to think and return the best result. The learning process comes by updating the weights interconnecting neurons and the bias values. The second mode is the testing or operational mode in which the network produces signatures for unknown feature vectors.

3. Testing and results

The testing process of the study uses the entire standard ORL database [27]. The ORL database consists of images of forty subjects, persons. The subjects are selected to include a wide range of human's features. There are ten images of each subject in different orientations, expressions or poses. This makes a total of four hundred images. These images are in gray scale of pixel values ranging from 0 to 255. Samples from the used dataset are in figure (5).

Figure (5) Samples of the standard testing dataset.

In our testing for each subject seven randomly selected are used for training the network and the entire subject set for testing. The input matrices to the feature extraction stage are normalized to the range [-1, 1] which is Legendre requirement. The neural network used in our experiments is fixed

to two hidden-layers 40 neurons each. The network learning termination criteria is set to 0.001 rmse, or epoch of 20 iterations. The network training function that updates weight and bias values according is set to Levenberg-Marquardt optimization.

Tables (1), (2) present the seven values vector of the Legendre, and Hue moments for the first three images of the first two persons, subjects, of the testing dataset consequently. The tables represent samples of the output of feature extraction step for both cases. These features are extracted after preprocessing step by Sobel-X.

Table (1) Legendre feature vectors for the first three images of the first two subjects

	Subject					
	1-1	1-2	1-3	2-1	2-2	2-3
$L_{2,0}$	-1.52585	-2.16267	-2.00713	-0.73775	-0.72303	-0.68035
$L_{0,2}$	-1.94712	-2.18926	-2.18054	-1.19314	-1.4577	-1.42579
$L_{1,1}$	0.110292	-0.11757	0.098119	0.204222	0.212857	0.11833
$L_{2,0}$	0.203614	-0.4778	0.342505	0.175914	0.475421	0.348751
$L_{2,1}$	-4.58814	-5.22268	-5.02745	-2.99221	-3.4366	-3.29391
$L_{0,3}$	-3.8673	-5.17718	-4.73071	-2.21299	-2.17949	-2.01838
$L_{4,0}$	-6.64154	-8.72905	-7.93448	-3.98326	-3.92777	-3.62289

Table (2) Hu-moments vectors for the first three images of the first two subjects.

	Subject					
	1-1	1-2	1-3	2-1	2-1	2-3
ϕ_1	0.17128644	0.17537247	0.17781504	0.207842	0.204242	0.21594033
ϕ_2	0.001979277	0.001445499	0.00258315	0.000771	0.001507	0.002730456
ϕ_3	6.88174E-09	1.04882E-08	3.9153E-08	7.3E-09	3.55E-08	8.84915E-08
ϕ_4	1.24848E-10	1.44157E-10	1.175E-09	3.25E-09	1.05E-09	2.32795E-08
ϕ_5	-8.58647E-20	-9.60776E-20	4.7886E-18	1E-17	4.66E-18	-1.79818E-16
ϕ_6	-1.87666E-11	-5.55984E-11	-8.829E-11	4.97E-11	2.16E-11	5.63705E-10
ϕ_7	-1.50432E-15	7.13864E-15	4.6216E-14	1.99E-16	1.71E-13	9.79702E-13

26

Each of the experiments presented in the study was performed several times and the results coincide with the conclusions drawn from each. An experiment is done as a single run and the entire set of results are taken as is to present. A better recognition rate could be presented if a selection is applied over the runs of each experiment. However, that has no effect on the drawn conclusions. Also, a better recognition rates could be presented if the network structure and the learning parameters with set per case but that take the study out of being a comparative one.

Experiment 1:-

The first experiment uses each of Hu-moments and Legendre moments with input source images, raw, preprocessed by Sobel (x-derivative and y-derivative), and Sobel followed by Otsu's thresholding. The experiment is done using samples set 20, 100, 200, 300, and the full set. The tests use 2, 10, 20, 30, and 40 persons consequently. The results of the experiment are summarized in Table (3). The results show the superiority of the Legendre over the Hue moments

Table 3: Results for both Hue moments and Legendre for raw, Otsu's binary and Sobel

Source / Samples count	Raw		Binary-X		Binary-Y		Sobel-X		Sobel-Y	
	Legendre	Hue	Legendre	Hue	Legendre	Hue	Legendre	Hue	Legendre	Hue
20	1.0000	1.0000	0.9500	0.6500	0.8500	0.9500	0.8500	0.5500	0.6500	0.8000
100	0.6300	0.5000	0.7800	0.7300	0.6200	0.5100	0.7700	0.1100	0.6700	0.1800
200	0.7950	0.3200	0.5800	0.5300	0.6550	0.5650	0.6400	0.0600	0.5600	0.0550
300	0.6067	0.2833	0.5600	0.3767	0.5167	0.3867	0.4667	0.0467	0.6067	0.0400
400	0.6275	0.2300	0.4475	0.3325	0.4500	0.2800	0.4150	0.0350	0.4775	0.0300

Experiment 2:-

The second experiment combines moments both from Legendre and Hue. The feature vector formed of fourteen values aiming that each carries different features. The moments are applied on input source images, raw, preprocessed by Sobel (x-derivative and y-derivative), and Sobel followed by Otsu's thresholding. The experiment is done on samples set of 20, 100, 200, 300, and the full set that correspond to the use 2, 10, 20, 30, and 40 persons consequently. The result of the experiment is in Table (4). The result indicates that an improvement in the recognition is achieved especially when the dataset volume increases.

Table 4: Hue and Legendre moments for raw , Otsu's binary and Sobel

Source Samples count	Raw	Binary-x	Binary-Y	Sobel-X	Sobel-Y
20	0.9500	1.0000	0.9500	1.0000	0.8500
100	0.9100	0.7900	0.5000	0.5800	0.7000
200	0.8000	0.7700	0.7750	0.4950	0.4600
300	0.7800	0.7133	0.6933	0.3733	0.0300
400	0.6700	0.6525	0.6350	0.3500	0.5200

Experiment 3:-

The third experiment setup is to combine moments of x-derivative and y-derivative to form the feature vector aiming that they carry different features. The Hue moments are used and the same setup for the experiments one and two is used. Table (5) summarizes the result of the experiment. The result indicates no significant improvement in the probability of recognition. Table 5: Hue moments on both x and y for Sobel and Otsu's binary

28

Source Samples #	Binary	Sobel
20	0.9000	0.8000
100	0.8000	0.1300
200	0.3350	0.0600
300	0.5133	0.0533
400	0.4825	0.0300

In the results of the experiments, there are few cases they may not fully coincide with the drawn conclusions and this come from the fact that the neural networks optimization is in general either to a deep or shallow local minima [24-26]. Each experiment result is a single run, taken as is, out many runs point to the same conclusion. The recognition rates are not the focus of the study rather presenting a complete run randomly selected for the setup mentioned above that reflect the reality.

4. Conclusion

In this paper, a face recognition method with a preprocessing stage added ahead of moments features extraction was subject to study. The preprocessing used are derivative filters and the derivative filters followed by Otsu's method thresholding. The classification mean used in the study was the feed forward neural network.

The experimental results indicated:-

1) The superiority of Legendre over Hue moments as an image feature extraction method based on the most widely set of moments of each.

2) The union of Legendre and Hue moments improves the probability of recognition compared to each of them. That is none of them is a subset of the other from an information point of view.

3) Combining moment's features based on derivatives-x and derivatives-y is insignificant to the probability of recognition.

29

4) Edges do not form a major component of moments, in general, but rather the texture of the surface dominates its values.

The study, implicitly, points to the fact that classifiers using moments: could be aided with edge-features. Also, likely more surface texture oriented preprocessing could be used rather than x, y derivatives to add more classification ability to moments.

References

[1] Chellapa, R., Wilson, C. L., and Sirohey, S. Human and machine recognition of faces: A survey. Proc. IEEE,83, 705–740.1995

[2] Nese Alyuz, Berk Gokberk, and Lale Akarun, 3-D Face Recognition Under Occlusion Using Masked Projection, IEEE transaction on information forensics and security, vol. 8, No. 5, May 2013.

[3] Berk Gökberk∗ , M. Okan ˙Irfano ˘glu, Lale Akarun, Ethem Alpaydın, " Learning the best subset of local   features   for   face   recognition", Elsevier- Pattern Recognition, Vol. 40, 2007, 1520 – 1532.

[4] C.Magesh Kumar, R.Thiyagarajan , S.P.Natarajan, S.Arulselvi,G.Sainarayanan, "Gabor   features   and   LDA   based   Face   Recognition   with   ANN classifier",Procedings Of ICETECT 2011.

[5] Önsen TOYGAR Adnan ACAN ,"Face recognition using PCA,LDA and ICA approaches on colored   Images", Journal Of Electrical and Electronics Engineering, vol- 13,2003.

[6] Issam Dagher,"Incremental PCA-LDA algorithm", International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2).

[7] J. Shermina,V. Vasudevan,"An Efficient Face recognition System Based on Fusion of MPCA and LPP", American Journal of Scientific Research ISSN 1450-223X Issue 11(2010), pp.6-19.

[8] Yun-Hee Han,Keun-Chang Kwak, " Face Recognition and Representation by Tensor-based MPCA   Approach",   2010   The   3rd   International Conference on Machine Vision (ICMV 2010).

[9] Neerja,Ekta Walia, "Face Recognition Using Improved Fast PCA Algorithm", Proceedings of IEEE 2008.

[10] Steve Lawrence,C.Lee Giles,A.h Chung Tsoi,Andrew D. Back, " Face Recognition: A Convolutional Neural Network Approach", IEEE Transactions on Neural Networks, Vol. 8, Jan 1997.

[11] Lin-Lin Huang, Akinobu Shimizu, Yoshihiro Hagihara, Hidefumi Kobatake, "Face detection from cluttered images using a polynomial neural network", Elsevier Science 2002

[12] Li, S. Z. and Lu,J. (1999). "Face recognition using the nearest feature line method". IEEE Transactions on Neural Networks,10(2):439-443.

[13] Ajoy Kumar Dey, Susmita Saha, Avijit Saha, Shibani Ghosh, "A Method of Genetic Algorithm (GA) for FIR

Filter Construction: Design and Development with Newer Approaches in Neural Network Platform", International Journal of Advanced Computer Science and Applications, Vol. 1, No. 6, pp. 87-90, 2010.

[14] Dr.S.Annadurai, A.Saradha. " Face recognition using Legendre Moments", Indian Conference on Computer Vision, Graphics and Image Processing 2004.

[15] D. Sridhar, I.V. Murali Krishna ,Combined Classifier for Face Recognition using Legendre Moments , Computer Engineering and Applications Vol. 1, No. 2, December 2012.

[16] Rajiv Kapoor, Pallavi Mathur, Face Recognition Using Moments and Wavelets, International Journal of Engineering Research and Applications , Vol. 3, Issue 1, January -February 2013, pp.632-635.

[17] R.C.Gonzalez, R.E.Woods Gonzalez, Digital Image Processing, Addison-Wesley, 2012.

[18] V.V. Starovoitov, , D.I Samal, D.V. Briliuk, "THREE APPROACHES FOR FACE RECOGNITION", The 6-th International Conference on Pattern Recognition and Image Analysis, October 21-26, 2002, Velikiy Novgorod, Russia, pp. 707-711

[19] S. Dabbaghchian, M. P. Ghaemmaghami, A. Aghagolzadeh, Feature Extraction Using Discrete Cosine Transform and Discrimination Power Analysis with a Face Recognition Technology, Pattern Recognition 43(4), 2010, 1431–1440.

[20] T. Lo and J. Siebert, "SIFT keypoint descriptors for range image analysis,", Annals of the BMVA X, 2009, pp. 1–18.

[21] J. Shen, W. Shen and D. Shen, "On Geometric and Orthogonal Moments", Inter. Journal of Pattern Recognition and Artificial Intelligence, Vol.14, No.7, pp.875-894, 2000.

[22] Dilipsinh Bheda□, Mahasweta Joshi□, Vikram Agrawal, A Study on Features Extraction Techniques for Image Mosaicing, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2014

[23] Yan Ke ; Sukthankar, R., PCA-SIFT: a more distinctive representation for local image descriptors, Computer Vision and Pattern Recognition, 2004.Proceeding of IEEE CVPR 2004 Computer Vision and Pattern Recognition Conference,Volume2.

[24] F. Roli, G. Giacinto, and G. Vernazza. "Methods for Designing Multiple Classifier Systems", In of 2nd International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, Cambridge, UK, Springer-Verlag, vol. 2096, pp. 78, 2001.

[25] G. P. Zhang, "Neural Networks for Classification: A Survey", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications And Reviews pp. 451-462, November, 2000.

[26] K. Hornik. "Approximation Capabilities of Multilayer Feedforward Networks", Neural Networks pp.251–257, 1991.

[27] Collection of Facial Images: Faces94, http://cswww.essex.ac.uk/mv/allfaces/faces94.html