International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.8 pp. 6-11

Face Recognition Using Echo State Networks

Fatimah Fathi Hammad[#1], Mabroukah Hamid[#2], Nadia Hmad[*3]

[#]*Department, College of Computer Technology, Tripoli- Libya*
[1]Fatf_2012@yahoo.com
[2]hamdmarwa7@gmail.com
[*]*Aljafara University, Aljafara-Libya*
[3]nadiaf_hammad@yahoo.com

***Abstract*— A face recognition system can be defined as a set of developing computer vision algorithms find to determine the similarity between images. In this paper, Hu Moment, Legendre Polynomial and Exact Legendre Moments algorithms are used to extract facial features from the Olivetti Research Laboratories (ORL) database. For human face recognition system, an Echo State Network (ESN) Neural Network (NN) model that developed for face recognition system. ESN model have not been explored for any face recognition systems previously. This allows us to claim priority for adopting this ESN model for the face recognition system. A Support Vector Machine (SVM) statistical model is also used to evaluate its adoption in developing face recognition system using the proposed database. The SVM model is mainly used as a baseline for a direct comparison. These models are evaluated on the extracted Olivetti Research Laboratories (ORL) features databases. The results show, the Echo State Networks (ESNs) gave a higher performance compared with the SVM model with 98.55% and 78.00% face recognition accuracy respectively. These results obtained with the Legendre Moment feature extraction technique that gave higher performance compared with other feature extraction techniques used in this paper. Also, our results are comparable with the other researches results.**

***Keywords*— face recognition, neural networks, echo state network, support vector machine, hu moment, legendre polynomial and exact legendre moments, artificial intelligent**

I. INTRODUCTION

Face recognition is an active subject due to extensive practical applications. Many recent events, such as terrorist attacks, exposed serious weakness in most sophisticated security systems. Various government agencies now motivate to improve security data systems based on body or behavioural characteristics, often called biometrics, (fingerprints , iris , finger/palm geometry, voice, signature and face) [1].

Face recognition seems a quite instinctive behaviour for human beings by which the brain interprets, identifies and verifies human faces, but it is really a tough and complex task for a machine-based system and for a computer vision system [2]. The face plays a main role in carrying identity of persons for automatically identifying or verifying a person from a digital image or a video frame from a video source. Face recognition has a great deal of attention by researchers due to its applications in different fields. Face recognition system is a branch of the digital image processing. A survey on face recognition techniques is in [3]

Digital image processing focuses on digital coding of images and find ways to address the digital data until these pictures or information can be used by machine, which can be a computer or a robot or other machines. Digital image processing is of great importance to realize any field images, t's very important when we try to understand image meaning or recognize patterns or shapes [4].

Automatic facial feature extraction is one of the most important and attempted problems in computer vision. It is a necessary step in face recognition, expression recognition, face detection, facial image compression systems. Various studies in features extraction field conducted in [5-11]. This paper investigates Hu Moments, Orthogonal moments, and Legendre moments features that extracted from the Olivetti Research Laboratories (ORL) Database.

Neural networks model is one of the most widely known models used for pattern recognition systems. The pattern is required to be recognized even when it is distorted. This paper investigates a facial recognition system using ESN and SVM neural networks models on the proposed features.

Echo State Network (ESN) was introduced by Jaeger in 2001 and has been applied to different real world applications where it proved to achieve a superior performance, This success has led to a wide acceptance of this technique in this application and encouraged researchers to conduct studies that aim to explore the fundamental properties and behaviour of (ESN) that lies behind its high performance [12]. ESN have already been successively applied to many real world tasks such as speech recognition [13, 14, 15], natural language task [16, 17] recognition, multi-machine power system [18], robot motor control [19], and Evaluation of Information-Theoretic Measures [20]. ESN model have not been explored for any face recognition systems previously. This allows us to claim priority for adopting this ESN model for the face recognition system.

Support Vector Machines (SVM), are supervised learning machines based on statistical learning theory that can be used for pattern recognition and regression [21]. This model is a powerful machine method developed from statistical learning and has made remarkable accomplishments in some fields. A special property of the SVM model is that, it simultaneously minimizes the empirical classification error and maximizes the geometric margin. So model called Maximum Margin Classifiers. This model is based on the Structural Risk Minimization (SRM), it maps input vector to a higher dimensional space where a maximal separating

ISSN 1737-930X

hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separate the data. The separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalization error of the classifier will be [22].

The SVM model have been utilized in a wide range of real world problems such as text categorization, hand written speaker recognition [23], computer vision [24], medical diagnosis [25], text cclassification [26], bioinformatics [27, 28, 29], and document classification [30].

SVM models are generally capable of delivering higher performance in terms of classification accuracy. Therefore, this model is mainly used as a baseline for effective comparison to the Echo State Network (ESN) Neural Network (NN).

II. Feature Extraction

In our everyday life, each of us receives, processes and analyses a huge amount of information of various kinds. Making the right decisions bases on the quality of this analysis. More than 95% of information we perceive is optical, so images are a powerful source of information to communicate machines [31].

An image can be defined as a two-dimensional function I. Where I = *f*(x, y)

As shown in Fig. 1, x and y are spatial coordinates, and the amplitude of *f* at any pair of coordinates (x, y) is called the intensity or grey level of the image at that point. Note that a digital image is composed of a finite number of elements; each of it has a particular location and value. These elements are referred to as picture elements [32]. Each element of this matrix (2-D array) is referred to as picture element or pixel [33].

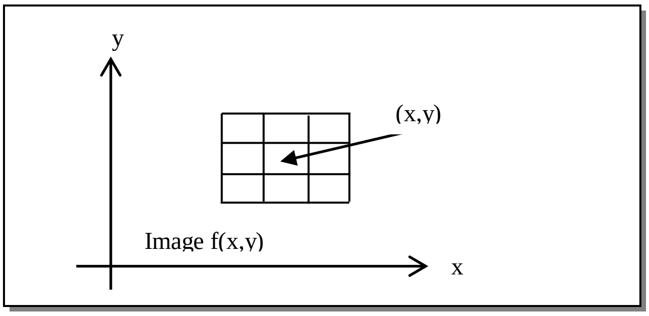

Fig. 1 A Digital Image Representation

The image can be processed by processing their elements directly (spatial domain processing) or indirectly (transform-domain processing) [33].

Analysis and interpretation of an image acquired by a real imaging system is the key problem in many application areas. This thesis attempted to understand the vision systems that can recognize human faces and to achieve the vision process for the computer through the application of several approach to analyse, interpret and extract significant, non-recurring features from grayscale faces of the image.

Extraction of visual features, as colour, texture, and shape is an important component. A shape is one of the visual features. However, shape descriptors can be divided into two categories, region based techniques and contour-based techniques [10]. In region-based techniques, all the pixels within a shape region are taken into account to obtain the shape representation, rather than only use boundary information as in contour-based methods. Common region based methods use moment descriptors that are preferred to represent the shape content of an image as Moment Invariants (MI), Legendre Moments (LM) and Exact Legendre Moments (ELM) to describe shapes. Other region-based methods include grid method, shape matrix [34].

In this paper some types of moments have been selected, then applied to make a direct comparison of them for face recognition system. Below we introduce the theoretical concepts of the moments that have been selected:

A. Hu Moment

Hu Moment or so called Geometric moments (GM) are a global description of a shape, it describes a shape's layout (arrangement of it's pixels), and has the ability to discern, filter, and aimed to avoid the effects of noise. They are usually used in image analysis. Moments can be classified as statistical moments that describe the rate of change in a shape's area [35].

Moment invariants are one of the earliest methods employed to perform object recognition, and they have been continuously developed. Furthermore, they are one of the most used methods in the field. Therefore, studying and applying this type of shape descriptor is a good starting point for providing an appropriate background in object recognition. The two-dimensional Cartesian moment is associated with an order that starts from low (where the lowest is zero) up to higher orders. The general definition of moment functions m_{pq} of order $(p+q)$ for an $x \times y$ continuous image intensity function f (x, y) is as in(1) [36]:

$$m_{pq} = \iint_{-1}^{1} x^p y^q f(x,y)\,dxdy \tag{1}$$

where p, q are integers represent the order of geometric moment, x and y are the coordinates of an image having a size $N \times M$ pixels, and $x^p y^q$ is the basis function.

Similarly, the general definition for an $x \times y$ digital image can be obtained by replacing the integrals with summations:

$$M_{pq} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} x^p y^q f(x,y) \tag{2}$$

These moments are not invariant to geometric transformations: translation, rotation and scaling. To make it invariant to translation, the central moment of order $(p+q)$ is given by (3) [8]:

$$\mu_{pq} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (x-x_0)^p (y-y_0)^q \; f(x,y) \tag{3}$$

Where x_0, y_0 are the coordinates of the center of gravity of the image, M_{00} The zero-order moment, which represents the total mass of a function and the two first-order moments, M_{01} and M_{10}, which represents the center of mass are given using (3). The centered normalized moment of order $(p+q)$ which is invariant to translation and scaling is defined as:

$$x_0 = \frac{M_{10}}{M_{00}}, \; y_0 = \frac{M_{01}}{M_{00}} \tag{4}$$

Second-order moments M_{02}, M_{20} describe the distribution of mass of the image with respect to the coordinate axes. It defines the orientation of the image [31].

ISSN 1737-930X

International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.8 pp. 6-11

The seven Hu moments which are invariant to translation, rotation and scaling are calculated using the following formula [8]:

$$\varphi_1 = \eta_{20} + \eta_{02} \quad (5)$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (6)$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \quad (7)$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \quad (8)$$

$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (9)$$

$$\varphi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (10)$$

$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)] \quad (11)$$

B. Orthogonal moments

Orthogonal functions have been around for a very long time. The best known are the sine and cosine. Two functions or vectors are orthogonal if their inner product (defined as the sum of the product of their corresponding elements) is zero [6].

An important class of orthogonal functions is orthogonal polynomials, which are orthogonal over various intervals of the real axis. Important orthogonal polynomials include Legendre, Hermite, Chebyshev, …, etc. Legendre polynomials, which are orthogonal over [-1, 1], can be taken as a product P(x).P(y), and the result is an orthogonal set of polynomials over a square. Orthogonal moments are computed similar to regular moments, except that the set of orthogonal polynomials replaces the x^p or $x^p y^q$ monomial. That is, where $h_{pq}(x,y)$ is the pq^{th} orthogonal polynomial, and R is the region over which the polynomials are defined.

C. Legendre moments

Legendre moments (LM) are one of the orthogonal moments that were introduced by Teague (1980).

The kernels of the Legendre Moments (LM) are products of Legendre polynomials defined along rectangular image coordinate axes inside a unit circle. The Legendre Moments (LM) of order $(p + q)$ are defined as:

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{1}\int_{-1}^{1} P_p(x)\, P_q(y) f(x,y)\, dxdy \quad (12)$$

Where the functions $P_n(x)$ denote Legendre polynomial of order (n). the Legendre moments L_{pq} generalizes the geometric moments M_{pq} in the sense that the monomial $x^p y^q$ is replaced by the orthogonal polynomial $P_p(x)\, P_q(y)$ of the same order.

In order to evaluate the Legendre Moments (LM), the Image coordinate space has to be necessarily scaled so that their respective magnitudes are less than one. If the image dimension along each coordinate axis is N pixels, and (i , j) denote the pixel coordinate indices along the axes, then of $(0 \le i , j \le N)$, and the discrete version of the (LM) can be written as :

$$L_{pq} = \frac{(2p+1)(2q+1)}{(N-1)^2} \sum_{i=1}^{N}\sum_{j=1}^{N} P_p(x_i) P_q(y_j) f(i,j) \quad (13)$$

Where x_i, $y_j$denote the normalized pixel coordinates in the range [-1,1] given by:

$$x_i = \left(\frac{2i}{N}\right) - 1; \quad y_j = \left(\frac{2j}{N}\right) - 1 \quad (14)$$

Where the Legendre polynomial $P_n(x)$ of order (n) is defined as :

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

Where $P_0(x) = 1, P_1(x) = x, |x| \le 1$ *and* $n > 1$

The orthogonality property of the Legendre polynomials enables the construction of the independent Legendre moments, providing minimum information redundancy among the feature descriptors. The recursive relation and the integral formula of Legendre polynomials have been effectively utilized in reduction the computation moments of binary images.

The recurrence relation, which can be used for efficient computation of the Legendre polynomials are:

$$P_0(x) = 1 \quad (16)$$

$$P_1(x) = x \quad (17)$$

$$P_{n+1}(x) = \frac{(2n+1)}{(n+1)} x P_n(x) - \frac{n}{n+1} P_{n-1}(x) \quad (18)$$

III. Face Recognition Classifier

A. Echo State Network

Echo State Networks (ESNs) are a recent approach was followed to train recurrent neural networks which showed excellent performance on learning temporal tasks. The training boils down to determining the weights of the connections to the output nodes [37]. When utilizing supervised learning, the weights of the output layer can be trained using linear regression.

The connection weights generated randomly, and not changed during training. Regression computes weights for every output connection. The internal layer of ESN is sparsely connected, and so it seems to be a contradiction that each output node is connected to all internal nodes [37].

1) Echo State Network Architecture

ESN are the first pioneering method in reservoir computing. It's rounded in the observation that if a Recurrent Neural Networks (RNN) has certain generic properties, only the output layer needs to be trained. The untrained part of an ESN is called a dynamical reservoir, and the states of the reservoir are called echoes of input history [42]. Fig. 2 shows the basic architecture for the ESN.

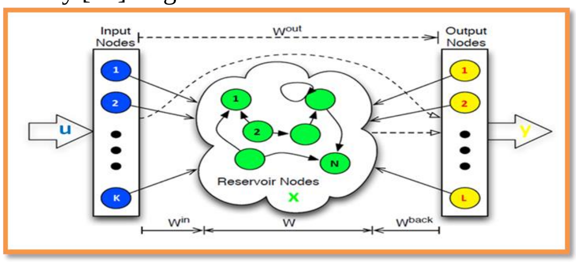

Fig. 2 A basic architecture of echo state network.

2) Echo State Network theory

The network consists of K input nodes connected to N reservoir nodes through a weighted connection matrix W^{in}. The reservoir has an internal connection matrix W . W^{back} is an optional back projection matrix connecting the output to the reservoir. Finally, the weights between the input and reservoir nodes to the L output nodes are collected in the matrix W^{out}. In the Fig. 2, the solid lines are static and

ISSN 1737-930X

randomly generated weights, while the dashed lines are being trained weights.

Echo State Networks (ESN) consist of neurons: input neurons $\mathbf{u} \in R^K$, hidden neurons $\mathbf{x} \in R^N$ *and* output neurons $\mathbf{y} \in R^L$, where, $\mathbf{u}$ represent the input vector $\boldsymbol{u} = (\boldsymbol{u1}, \boldsymbol{u2}, \ldots, \boldsymbol{uK})^T$, $\boldsymbol{K}$ number of input vector elements, $\boldsymbol{x}$ represent the internal state vector $\boldsymbol{x} = (\boldsymbol{x1}, \boldsymbol{x2}, \ldots, \boldsymbol{xN})^T$, $\boldsymbol{N}$ number of hidden units or internal units called (reservoir) and y represent the output vector $\mathbf{y} = (\mathbf{y1}, \mathbf{y2}, \ldots, \mathbf{yL})^T$, $\mathbf{L}$ number of output units [39].

3) *Echo State Network Training*

Assuming that an untrained network ($\boldsymbol{W^{in}}, \boldsymbol{W}, \boldsymbol{W^{back}}$) with state update and with transfer functions $\boldsymbol{tanh}$. Let $\boldsymbol{W}$ have a spectral radius $|\lambda_{max}| > \mathbf{1}$, where $|\lambda_{max}|$ is the largest absolute value of an eigenvector of $\boldsymbol{W}$. Then the network has no echo states with respect to any input/output interval $\boldsymbol{U \times D}$ containing the zero input/output *(0,0)*[12].

An ESNs at each time step $\boldsymbol{t}$ computes its output $\boldsymbol{y(t)}$ based on its internal state $\boldsymbol{x(t)}$ using (1) [37]:

$$\boldsymbol{y}(\boldsymbol{t}) = \boldsymbol{f_{out}}(\boldsymbol{x}(\boldsymbol{t})^T \boldsymbol{W^{out}}) \tag{19}$$

Where $\boldsymbol{W^{out}}$ is the output weight matrix, and $\boldsymbol{T}$ denotes the transpose of a matrix. $\boldsymbol{f_{out}}(.)$ is the output activation function. State $\boldsymbol{x(t)}$ of internal nodes is computed based on the input $\boldsymbol{u(t)}$, and the previous state $\boldsymbol{x(t-1)}$. Optionally the previous network output $\boldsymbol{y(t-1)}$ can be fed back to the net. The state is the internal nodes computed using (20) [37]:

$$\boldsymbol{x}(\boldsymbol{t}) = \boldsymbol{f}(\boldsymbol{W^{in}}\,\boldsymbol{u}(\boldsymbol{t}) + \boldsymbol{W}\boldsymbol{x}(\boldsymbol{t}-\mathbf{1}) + \boldsymbol{W^{back}}\boldsymbol{y}(\boldsymbol{t}-\mathbf{1})) \tag{20}$$

where $\boldsymbol{W}$, $\boldsymbol{W^{in}}$, and $\boldsymbol{W^{back}}$ are weight matrices of the connections, between the internal nodes $\boldsymbol{W}$, between the input nodes and the network $\boldsymbol{W^{in}}$ and between the output nodes and the network $\boldsymbol{W^{back}}$, $\boldsymbol{f}$ is a nonlinear transfer function, commonly sigmoid or $\boldsymbol{tanh}$ function, the work in this paper is used $\boldsymbol{tanh}$ function. So $\boldsymbol{x_{train}}(\boldsymbol{t})$ is computed according to (20):

$$\boldsymbol{x_{train}}(\boldsymbol{t}) = \boldsymbol{f}(\boldsymbol{W^{in}}\,\boldsymbol{u_{train}}(\boldsymbol{t}) + \boldsymbol{W}\boldsymbol{x_{train}}(\boldsymbol{t}-\mathbf{1}) + \boldsymbol{W^{back}}\boldsymbol{y_{train}}(\boldsymbol{t}-\mathbf{1})) \tag{21}$$

The output weight matrix $\boldsymbol{W^{out}}$ can be computed via regression using (22):

$$\boldsymbol{W^{out}} = (\boldsymbol{X^T_{train}}\,\boldsymbol{X_{train}} + \boldsymbol{\lambda^2 I})^{-1}\,\boldsymbol{X^T_{train}}\,\boldsymbol{Y_{train}} \tag{22}$$

Where $\boldsymbol{I}$ the identity matrix and -1 is pseudo invers matrix. A C++ code was written by the author to implement the algorithm above and free available for the researchers, more details are in [40, 41].

B. *Support Vector Machine*

1) *Support Vector Machine For Binary Classification*

A SVM is primarily a two-class classifier [42]. But it is possible to solve the multi classification problems using support vector machines by converting the multi classification problem to a binary classification problem and deal with each class separately as a binary classification problem. SVM considers the following cases:

- Linearly separable data:

In this case the aim is to separate the two classes by a function which is induced from available examples. The goal is to produce a classifier that will work appropriately on unseen examples, i.e. it generalizes well.

To clarify the linear classification case using support vector machines assume that we have given training data samples S as follows:

$$\boldsymbol{S} = (\boldsymbol{x_1}, \boldsymbol{y_1}), \cdots \cdots (\boldsymbol{x_m}, \boldsymbol{y_m}) \in \boldsymbol{R^n}, [-\mathbf{1}, +\mathbf{1}] \tag{23}$$

Where xi is input vectors and *yi* are their labels. A label with the value of +1 denotes that the vector is classified to class +1 and a label of −1 denotes that the vector is part of class −1. Here there is a need for function: $\boldsymbol{f(x)} = \boldsymbol{y} : \boldsymbol{R^n}, [-\mathbf{1}, +\mathbf{1}]$ that correctly classifying the training data, correctly classifies unseen data samples too, this is called a generalization.

It is imperative that the class of functions should be restricted so the machine can learn, otherwise learning the underlying function is impossible. Thus, for this reason the SVMs are based on the class of hyperplanes.

$$\boldsymbol{w.x} + \boldsymbol{b} = \mathbf{0}, \boldsymbol{w} \in \boldsymbol{R^n}, \boldsymbol{b} \in \boldsymbol{R} \tag{24}$$

Where, the vector w defines a direction perpendicular to a hyperplane while varying the value of *b* move the hyperplane parallel to itself. To find the class of a particular vector *x*, the following decision function is used [44].

$$f(x) = sign(w.x + b) \tag{25}$$

- Non-linearly separable data

A very few data sets in the real world are linearly separable. What makes support vector machines so remarkable is that the basic linear framework is easily extended to include the case where the data set is not linearly separable.

The fundamental idea is to transform the input space where the data set is not linearly separable into a higher-dimensional space called a feature space as shown in Fig. 3, where the data becomes linearly separable. The functions associated with these transformations are called kernel functions, and the process of using these functions to move from a linear to a nonlinear support vector machine is called the (kernel trick [45]).

The Non-linear discriminant function can then be written as:

$$\boldsymbol{\alpha^*} = \boldsymbol{argmin}\,\frac{\mathbf{1}}{\mathbf{2}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}\sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{\alpha_i \alpha_j\, y_i y_j K(x_i, x_j)} - \sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{\alpha k} \tag{26}$$

where $K(x, x_0)$ is the kernel function performing the non-linear mapping into feature space, and the constraints are unchanged [44].

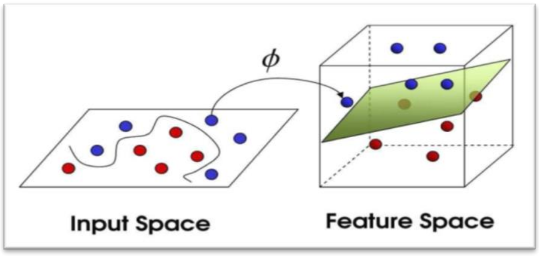

Fig. 3 Mapping non-linear to linear problem with kernel trick

This paper used the kernel function for mapping the input vectors to the feature space. More details for used the kernel function are in [22, 44].

2) *Solving SVMs with Sequential Minimal Optimization (SMO):*

SMO is an efficient batch numerical algorithm that has been developed to solve the SVM quadratic programming (QP) problem [46].

SMO is an algorithm for training support vector machines. Training a support vector machine requires the solution of a very large quadratic programming (QP) optimization problem. SMO breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically [47].

The SMO algorithm is composed of 3 main procedures:

- Run: which iterates over all points until convergence to a tolerance threshold
- Examine Example: which finds two points to jointly optimize
- Take Step: which solves the 2-dimensional optimization problem analytically.

SMO algorithm gives an efficient way of solving the dual problem desired [47]. The SMO algorithm and its pseudo code and C++ source code, are in [40]. More details of SVM are in [45].

IV. Experimental Results

A. Data

The ORL (Olivetti Research Laboratories) database [48] have been used by many researchers and proven quality in the classification process. This database makes available 400 pictures were taken between April 1992 and April 1994 in the laboratory, and these face images were taken in different times, varying in lighting, facial expressions ((open / closed) eyes, smiling / not smiling) and details of the face (wearing glasses / no glasses). All the images were captured on a homogeneous dark background, and the faces of the people were in the front projection (while allowing some lateral movement (rotation)). This database contains pictures of faces for 40 different people. Every person has ten different shots, available in various sizes (112x92, 64x64, 32x32). The image size (64x64) was chosen this paper work.

B. Experimental Work For Feature Extracting Phase

Ttheoretical concepts that described in studies [6-8, 10] were addressed on the ORL database, and each image was conducted the following processes:

- Reading the image and storing it in a square matrix.
- Normalizing image by dividing each pixel or element at the highest pixel value or element in an image matrix.
- Hu Moment of order 3 and order of order 4 for the Legendre Polynomial and Exact Legendre Polynomial were used for extracting the facial features.
- Computing features matrix. It is a square matrix due the degree of methodology specified. This matrix corresponds to the original image matrix and called Image Moment.

7 features of Hu Moment, 25 features for Legendre Polynomial and for the Exact Legendre Polynomial were obtained.

The source code for these algorithms, in web and windows applications, was written using Visual Studio.Net Technology by C# language and is available free for the researchers. More details are in [40].

C. Experiment's Results

The features dataset was divided randomly into 70% for training and 30% for testing. The ESN number of neurons and learning rate were experimentally optimized and gave the best results when they were 25 input neurons, 400 reservoir neurons, 40 output neurons, and 0.9 learning rate. Evaluating the efficiency of the networks is used (27).

$$Recognition\ rate = \frac{No\ of\ correct\ classification\ sample}{No\ of\ samples} \quad (27)$$

The results from our experiments compared to other published work on the same ORL database. Table 1 shows the results of the ESN and the SVM proposed algorithms compared to other researchers work used the same face database (ORL) with different classifier techniques.

TABLE I
THE RESULTS OF THE ESN AND THE SVM ALGORITHMS COMPARED TO OTHER RESEARCHERS RESULT USED THE ORL DATABASE WITH DIFFERENT TECHNIQUES

Model (Classifier Type)	**Feature Extraction technique**	**Accuracy**	**Reference**
PCA, SVM	magnitude and phase of Gabor	99.90%	[49]
nearest neighbour	legendre polynomial	98.25%	[50]
nearest neighbour	hu moment	46.8 %	[50]
SVM	wavelet	98.1 %	[51]
SVM	wavelet	94.8 %	[52]
SVM	ICA	96%	[53]
convolutional multilayer perceptron (MLP) neural network.	self-organizing map (SOM)	88.2%	[54]
SVM	hu moment	53%	this study
	exact legendre moment	54%	
	legendre moment	78 %	
ESN	hu moment	50%	this study
	exact legendre moment	95.83%	
	legendre moment	98.55%	

Considering previous studies [49-52, 54] on ORL database we obtained comparable face recognition accuracy rates with Legendre Moment feature extraction technique and Echo State Networks (ESNs) as it is shown in Table 1.

The nearest results are obtained for the ORL dataset compared to our work are [50, 51, 49]. The work of [50] was done on one classifier with different extraction technique. This gave us indication that the features type has strong effect on the results. Next work will be done to examine the ESN with different feature extraction technique.

V. Conclusion

The objective of this paper was to obtain an accurate face recognition system. In order to achieve this aim, the Hu Moment, Exact Legendre Moment, Legendre Polynomial algorithm used to extract the face features. Thus, the ORL face database was used to explore the potential gain in building a face recognition system using the SVM model as a baseline to our work using the ESN model. To the best of our knowledge, ESN model models have not been explored

ISSN 1737-930X

for any face recognition system. This encouraged us to investigate the ESN model for face recognition system using the ORL database. This allows us to claim priority for adopting this ESN model for face recognition system. This ESN model gave a higher performance for face reception system on the proposed ORL database compared to the SVM model. Also, our results are comparable with the other researches results. Our best result was on the ESN model with 98.55% face recognition performance.

VI. References

[1] S.S., Al-atrash, Robust Face Recognition. 2011.

[2] D. Mou, "Machine-based Intelligent Face Recognition. Springer," 2010.

[3] M. Lal , K. Kumar, R. Arain, A. Maitlo, " Study of Face Recognition Techniques: A Survey," (IJACSA) International Journal of Advanced Computer Science and Applications, 2018.

[4] http://en.wikipedia.org/wiki/Facial_recognition_system.

[5] B.G. Bhatt, and Z.H. Shah, "Face Feature Extraction Techniques: A Survey," National Conference on Recent Trends in Engineering & Technology, 2011.

[6] S. Sharma, A. Dhole, "Content Based Image Retrieval Based on Shape Feature using Accurate Legendre Moments and Support Vector Machines," IJCSET, 2013. 3(5): p. 194-199.

[7] T. Arif, et al., "Object Classification Via Geometrical, Zernike And Legendre Moments," Journal of Theoretical and Applied Information Technology, 2009.

[8] D.S. Annadurai, A. Saradha, "Face Recognition Using Legendre Moments," ICVGIP, 2004.

[9] Sabhara, R.K., C.-P. Lee, and K.-M. Lim, "Comparative Study of Hu Moments and Zernike Moments in Object Recognition. Smart Computing Review," 2013.

[10] R. Kapoor, P. Mathur, "Face Recognition Using Moments and Wavelets. International Journal of Engineering Research and Applications," in IJERA, 2013.

[11] Sivakumar, R. and D.M. Sridhar, "Multi Class Different Problem Solving Using Intelligent Algorithm. International Journal of Engineering Research and Applications," in IJERA, 2012.

[12] H. Jaeger, A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. German National Research Center for Information Technology, 2003.

[13] A. Alalshekmubarak, "Towards A Robust Arabic Speech Recognition System Based On Reservoir Computing," PhD. Thesis, UK, 2014.

[14] N. Hmad, "Deep Neural Network Acoustic Models for Multi-dialect Arabic Speech Recognition," PhD. thesis, Nottingham Trent University, UK, 2015.

[15] N. Hmad, A. Alalshekmubarak, T. Allen, "A comperative study between and Echo State Networks and Support Vector Machine models on the Levantine Arabic corpus," in LICEET, March 2018, Libya.

[16] M.H. Tong, et al., "Learning Grammatical Structure with Echo State Networks," Neural Networks, Elsevier, 2007. 20: p. 424–432.

[17] V. Sakenas, "Distortion Invariant Feature Extraction with Echo State Networks," in School of Engineering and Science2010, School of Engineering and Science, Jacobs University Bremen gGmbH.

[18] G.K. Venayagamoorthy, "Online Design of an Echo State Network Based Wide Area Monitor for a Multimachine Power System," Neural Networks 2007. 20: p. 404–413.

[19] M. Salmen, P.G. Ploger, "Echo State Networks used for Motor Control," IEEE, 2005: p. 1953-1958.

[20] M. Torda, I. Farkaš, "Evaluation of Information-Theoretic Measures in Echo State Networks on the Edge of Stability," International Joint Conference on Neural Networks (IJCNN) ,IEEE, 2018.

[21] H. Girma, A Tutorial on Support Vector Machine. 2009.

[22] D.K. Srivastava, L. Bhambhu, "Data Classification Using Support Vector Machine," Journal of Theoretical and Applied Information Technology, 2009.

[23] T. Tashan, "Biologically Inspired Speaker Verification," PhD. thesis, Nottingham Trent University, UK, 2012.

[24] B.D. Barkana, I. Saricicek, B. Yildirim, "Performance analysis of descriptive sta- tistical features in retinal vessel segmentation via fuzzy logic, ann, svm, and classifier fusion," Knowl.Based Syst. 118 (2017) 165–176 .

[25] R. Ren, "ANN Vs. SVM: which one performs better in classification of MCCs in mammogram imaging, Knowl," Based Syst. 26 (2012) 144–153.

[26] M. Ayyash, "Hybrid Support Vector Machine based Feature Selection Method for Text Classification," International Arab Journal of Information Technology 15(3A):599 · April 2018.

[27] E. Byvatov, G. Schneider, "Support vector machine applications in bioinformat- ics.," Appl. Bioinf. 2 (2) (2003) 67–77.

[28] W. Chen, H. Yang, P. Feng, H. Ding, H. Lin, "IDNA4mc: identifying DNA n4-methylcytosine sites based on nucleotide chemical properties," Bioinformat- ics 33 (22) (2017) 3518–3523.

[29] S. Maldonado, R. Weber, J. Basak, "Kernel-penalized SVM for feature selection," Inf. Sci. (Ny) 181 (1) (2011) 115–128.

[30] W. Zhang, T. Yoshida, X. Tang, "Text classification based on multi-word with support vector machine, Knowl," Based Syst. 21 (8) (2008) 879–886.

[31] Flusser, J., T. Suk, and B. Zitová, eds. "Moments and Moment Invariants in Pattern Recognition" A John Wiley and Sons. 2009, Ltd, Publication.

[32] Gonzalez, R.C. and R.E. Woods, eds. "Digital Image Processing, Second Edition. Prentice Hall Upper Saddle River," 2002: New Jersey 07458.

[33] V.E. Duró, "Face Recognition By Means of Advanced Contributions," in Machine Learning. 2013.

[34] Zhang, D. and G. Lu, "Review Of Shape Representation And Description Techniques," Elsevier Ltd, 2003.

[35] Nixon, M.S. and A.S. Aguado, "Feature Extraction and Image Processing," Ltd. 2002.

[36] Cheriet, M., et al., "Character Recognition Systems," John Wiley & Sons, Inc. 2007.

[37] K. Diamantaras et al., Lecture Notes in Computer Science. Springer 2010.

[38] D. Stuan, "Evolving Echo State Networks for Minimally Cognitive Unsupervised Learning Tasks," NTNU In novation and creativity, 2013.

[39] T. Strauß, W. Wustlich, and R. Labahn, Design strategies for weight matrices of Echo State Networks. 2012.

[40] F. Hammad, "Face Recognition Using Legendre Polynomial & Echo State Network," M. Eng. thesis, School of Applied Sciences and Engineering, The Libyan Academy, Libya 2016.

[41] F. Hammad, M. Hamid, N. Hmad, "A comparative study between Echo State Networks and Support Vector Machine models on the Olivetti Research Laboratories Face Database," CSP-2019 conference, 30-31 March 2019, Tunisia.

[42] Lecture 18: Support Vector Machines, in Intelligent Data Analysis and Probabilistic Inference.

[43] Y. Ahuja, S.K. Yadav, "Multiclass Classification and Support Vector Machine," Global Journal of Computer Science and Technology, 2012.

[44] S.R. Gunn, Support Vector Machines For Classification and Regression. 1998.

[45] L. Hamel, "Knowledge Discovery With Support Vector Machines," A John Wiley & Sons, Inc., Publication, 2009.

[46] A. Bordes, "New Algorithms for Large-Scale Support Vector Machines," devant le jury compos_e de, 2010.

[47] J.C Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization," Advances in Kernel Methods. MIT Press, pp. 185–208. 1999.

[48] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[49] F. Bellakhdhar, K. Loukil, M. ABID. "Face recognition approach using Gabor Wavelets, PCA and SVM," IJCSI International Journal of Computer Science Issues 10.2 (2013): 201-206.

[50] N. Seo, "A Comparison of Multi-class Support Vector Machine Methods for Face Recognition," Department of Electrical and Computer Engineering, The University of Maryland (2007).

[51] M.S. Nixon, A.S. Aguado, Feature Extraction and Image Processing. Ltd. 2002.

[52] B. Luo, Y. Zhang, Y.H. Pan, "Face recognition based on wavelet transform and SVM," in Proceedings of the 2005 IEEE international conference on information acquisition, June 27–July 3, 2005, Hong Kong and Macau, China.

[53] K. Rui, B. Zhang, "A New Face Recognition Method Based on Fast Least Squares Support Vector Machine," Physics Procedia 22 (2011): 616-621.

[54] S. Lawrence, C. Giles, A. Tsoi, A. Back, "Face recognition: A convolutional neural network approach," IEEE Transactions on Neural Networks, 8, 1997, 98–113.

ISSN 1737-930X