Face Recognition with PCA and KPCA using Elman Neural Network and SVM

Hossein Esbati and Jalil Shirazi

Abstract—In this paper, in order to categorize ORL database face pictures, principle Component Analysis (PCA) and Kernel Principal Component Analysis (KPCA) methods by using Elman neural network and Support Vector Machine (SVM) categorization methods are used. Elman network as a recurrent neural network is proposed for modeling storage systems and also it is used for reviewing the effect of using PCA numbers on system categorization precision rate and database pictures categorization time. Categorization stages are conducted with various components numbers and the obtained results of both Elman neural network categorization and support vector machine are compared. In optimum manner 97.41% recognition accuracy is obtained.

Keywords—Facial recognition, Principal Component Analysis, Kernel Principal Component Analysis, Neural network, Support Vector Machine.

I. INTRODUCTION

FACE recognition is a pattern recognition process, which is applied especially on faces. This process include recognition of a face as ‘unknown’ or ‘known’ following a comparison with known people stored faces. Face recognition computation models should be responsive to such a complicated problem. This Complexity is a result of this fact that faces should be represented in a way that the existence information about faces are used optimally for recognizing a special face between other faces. In this case, faces cause a difficult problem because all the faces have a set of similar characteristics as eyes, nose and mouth [1]. Identify and face recognition various systems and technologies are represented and analyzed in [2] and [3] Which in this cases. Elman neural network method can be named which in recent years are used extensively in various pattern recognition applications. Elman neural network essentials are represented in section three. This method is used for identify and recognize faces in [4] and [5] resources. In the present research, principle Component analysis (PCA) and Kernel Principal Component Analysis (KPCA) [5] to dimension reduction and Elman neural network to categorization are used and as a research work PCA and KPCA effects on recognition precision and recognition time are reviewed. Also in order to compare Elman neural network efficiency for face recognition, its obtained results are compared with Support Vector Machine (SVM) [6] results.

The present study is organized as follows: PCA and KPCA algorithm and characteristics are represented in section 2, Elman neural network is introduced in section 3 and demonstrated empirical tests are determined and discussed in section 4 and section 5 is concluding part.

II. CHARACTERISTICS EXTRACTION

Following facial pictures disclosure and standardization, facial characteristics are derived. In system training stage, user model is obtained by characteristics extraction for every user, then in identify approval stage, input picture characteristics vector is compared with user model in database for making decision about accepting or rejection the claim. Aside from trait kind and extraction method this stage output won't have a picture quality, but it is a vector or matrix of characteristics. Until now various methods are used for characteristics extraction and dimension reduction as Principle Component Analysis (PCA) and Linear Discriminate Analysis (LDA). Also various methods are used for categorization as Support Vector Machine (SVM) and neural networks [4]. In this section two methods which are used in this study are explained briefly.

A. Principle Component Analysis

PCA main idea is reducing the existent dimensions of data in a set. So that the existent changes in these data are maintained as much as possible. In PCA special vectors or the same perpendicular, to each other vectors are in a shape that the first component of PCA includes the most information and this trend is continued to the last component and the last component includes the least information. PCA linear conversion, suppose the data as a space in which data have the most variances. Suppose that data include N observations, \( X_k \in \mathbb{R}^M, k=1,...,N \) and \( \sum_{k=1}^{N} X_k = 0 \) Data set covariance matrix is obtained from following formula [7]:

\[
C = \frac{1}{N} \sum_{k=1}^{N} X_k X_k^T
\]  

(1)

Principle components are obtained by C diagonal building, principle are components obtained which these components are orthogonal picture on special vectors which are calculated by solving special quantity equation [7]:

\[
\lambda V = CV
\]  

(2)

Where \( \lambda \geq 0 \) is special quantity and \( V \in \mathbb{R}^M \setminus \{0\} \) (i.e. \( \mathbb{R}^M \) except \{0\}) are special vectors on one hand\( CV = \frac{1}{N} \sum_{k=1}^{N} (X_k V) x_k \). All the responses for V should be a linear combination observations we have [7]:

\[
V = \sum_{k=1}^{N} \alpha_k X_k
\]  

(3)

Where \( \alpha_k \) are coefficients which their quantities are selected in a way that above relationship established [7].
B. Kernel Principal Component Analysis

KPCA main technique is calculating PCA conversion in a mapping space by a Non-linear mapping function which for estimating this mapping, kernel idea is used. Consider \( \phi(X_1), \ldots, \phi(X_N) \) are mapped data which their mean is not zero. First mapped data mean becomes zero following formula:

\[
\hat{\phi}(X_k) = \phi(X_k) - \frac{1}{N} \sum_{j=1}^{N} \phi(X_j)
\]

(4)

Covariance matrix is calculated by (5) formula:

\[
\Sigma = \frac{1}{N} \sum_{j=1}^{N} \hat{\phi}(X_j) \hat{\phi}(X_j)^T
\]

(5)

\( \hat{\lambda} \Sigma \hat{\nu} \) is special quantity equation for covariance matrix where \( \lambda \geq 0 \) is special quantity and \( \hat{\nu} \in \mathbb{R} \) is special vectors. Special vectors equation equivalent can be written in the form of (6) formula:

\[
\hat{\lambda} \left( \hat{\phi}(X_k), \hat{\nu} \right) = \left( \phi(X_k), \Sigma \hat{\nu} \right), k = 1, \ldots, N
\]

(6)

Where \( \alpha_k \) are coefficients which their quantities are selected in a way that (9) formula is established.

\[
\hat{\nu} = \sum_{k=1}^{N} \alpha_k \phi(X_k)
\]

(7)

Where by substituting (7) formula by (6) formula we have:

\[
\hat{\lambda} \alpha = K \alpha, \; (\alpha = (\alpha_1, \ldots, \alpha_N)^T)
\]

(8)

Where \( K \) is kernel matrix which is in the form of \( N \times N \) square matrix by \( \text{K}_{ij} = \left( \phi(X_i), \phi(X_j) \right) = k(X_i, X_j) \) elements. For \( \lambda_k \alpha_k = (\alpha_k^2, \alpha_k^2) = 1 \) formula is applied in mapping space, also like every other PCA algorithms, data should be concentrated in mapped space, so kernel matrix should be replaced by following formula:

\[
\hat{R} = K - 1_N K - K 1_N + 1_N K 1_N
\]

(9)

Where \( (1_N)_{ij} = 1/N \) [8].

III. ELMAN NEURAL NETWORK

A Dynamic Neural Network (DNN) can adjust its structure with continuous changes and never changes topology and training not stopped [9]. Elman network is recommended as a recurrent network for modeling storage system. Elman network is a three layered network with a feedback from hiding layer output to input state. Its advantage in comparison with complete recurrent systems is that it can be used for network training after dispersion, because the links to the context units are constant. In a way that Context units act a sampling cycle delay. Context layer store in itself a copy of hidden neurons output and context layer neuron quantities are applied as an extra input signal in the hidden layers [10].

In Elman network, loads from hidden layer to the context layer are set and fixed on 1, because context neurons quantities must be hidden layer output copies, moreover context neurons output preliminary loads are equal to half of output limits of other neurons in the network. One interpretation of this network is that hidden layer outputs show network state. Network outputs are functions from present state, previous state (as it is provided by context units) and the current outputs that is when a set of outputs are shown to the network. The network can learn to represent suitable outputs in network previous states context. Elman network is invented for speech recognition originally, but in other fields as system recognition and short-term predictions which are applied in robots Movement programming also are mostly used. Elman network includes tensing neurons in hidden layer and purelin neurons in its output layer, this is a special combination, because two layers networks with these conversion functions can estimate each functions (with a limited series of discontinuities) with desirable accuracy. A view of a sample of Elman neural network Structure is shown in figure 1.

![Fig.1 A view of a sample of Elman neural network Structure](image)

In this research a 3 layered 8 neurons Elman neural network in hidden layer and 1 neuron in output layer suitable neurons. In input layer with optimum characteristics numbers are used. Various tests on various neurons of hidden layer are conducted to obtain the best results. Because this function is non-linear and derivative for Elman network training, there are a lot of training functions which in this research. Levenberg-Marquardt post-dispersion error algorithm because of higher convergence than other training function is used and error rate for training stoppage is considered 0.001.

Some points:

In case of not choosing suitable structure the network develops overtraining problem on the basis of empirical relations, two factors of network input vectors and network loads total quantities are important for preventing form this problem. For instance, if we have three layers network with five neurons in input layer, 30 neurons in middle layer and 2 neurons in output layer, network loads quantities are \( 5 \times 3 + 3 \times 2 = 21 \), so in real situation input vectors number should be selected about \( 4 \times 21 = 84 \) or more.

\[
> 3 \text{ or } 4 \text{ (Real), 10 (Ideal)}
\]

(10)
by considering that neural network acts on the basis of random implementation performance, usually many times median is considered as percent of final parting, but for reaching network precision, many times implementation criterion deviation rate is considered in data parting which in most of the reports is in the form of ± signs. For instance, 82.2 ± 5.2 percent means classifier able to separate data with 82.2 percent median and with a 5.2 percent precision.

IV. EMPIRICAL RESULTS AND THEIR ANALYSIS

In this section we review used database and algorithm.

A. Face Bank

Experimental pictures are form ORL database, these data includes 400 personal pictures of 40 people [3]. For each individual 10 various picture in various times and light variations various gestures (open and closed eyes a simile or without a smile) and other situations as glasses or without glasses, are provided.

From each individual 10 pictures, seven pictures are used in training stage and three other images are used in test stage. In each experiment, the system tested with 10 programs run and average of error calculated. We used 3 random pictures for testing in every programs run.

B. Algorithm Implementation

The conducted research includes following stages:

1-Reading images collection and creating matrix aim: 172 x 92 x 400 and 400 x 1
2-PCA implementation: Selecting data quantities for reducing dimension by using PCA. Converting images into 70 x 90 x 400 dimension and then converting into 400 x 6300 vectors. Calculating characteristics vectors median and reducing each characterization vectors from median vector. Finding especial rates and arranging them reducing the dimension equal to predetermined rate in stage 2-1.

3-Selecting training data and test as well as each or target vector: 70 % and 30 % total data for training data and test data are selected from each class in a random and independent manner.

4-Neural network creation: an Elman two lined neural network is selected by input line neuron numbers equal to the selected characteristics quantities by PCA. Also epoch numbers 200, \( \mu = 0.95 \) and mean square minimum error performance function (MSE) and training by gradient decent back propagation with adaptive learning rate are considered. Network training with training data and training data target vector. Test data or trained network simulation in figure 2 network convergence error curves and in figure 3 network training curves by using 100 characteristics are shown.

5-Calculating algorithm implementation time and calculating training data correctness percent and test error rate: for relying on the obtained results algorithm implementation 10 times median is considered as correctness percent.

In order to review the effects of PCA output components in applied neural network classification correctness, we conduct the above steps in line of components numbers various rates. Also in order to review Elman network efficiency in face classification, we conduct similar stages by using SVM categorization method which for it we use 40 SVM and one against all method [6]. The obtained results of implementation of some conduct test are shown in table 1 and figures 4 and 5. The results show that in all of using similar quantities, SVM precision characteristics are more than others in all states. While, neural network implementation rapidity is more than SVM. The obtained results in table 1 show that it is possible to distinguish between a set of images with 110 PCA characteristics quantities by using SVM till 96.16% and by using Elman neural network till 77.17% and implementation duration by the same characteristics is about 126 seconds for SVM and about 19 seconds for Elman neural networks. Then we conduct all the above experimentations by using kernel principle components analysis in table 2 and figure 7 and 8. The obtained results show that using kernel principle components analysis in support vector machine increases recognition accuracy.
Fig. 4 Comparing obtained results from algorithm implementation and classification by two methods: 1-Elman neural network (ANN-Elaman) and 2- support vector machine with one against all method

Fig. 5 Comparing algorithm implementation time both methods 1-Elman artificial neural network (ANN-Elman) and 2- support vector machine with one against all method

TABLE I

<table>
<thead>
<tr>
<th>Number of features from PCA</th>
<th>ANN-Elman</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy percent</td>
<td>Elapsed time (second)</td>
</tr>
<tr>
<td>110</td>
<td>77.1708</td>
<td>28.6527</td>
</tr>
<tr>
<td>105</td>
<td>77.0042</td>
<td>28.5128</td>
</tr>
<tr>
<td>100</td>
<td>75.0042</td>
<td>28.2053</td>
</tr>
<tr>
<td>95</td>
<td>76.1708</td>
<td>27.9715</td>
</tr>
<tr>
<td>90</td>
<td>75.2542</td>
<td>27.7101</td>
</tr>
<tr>
<td>85</td>
<td>74.9208</td>
<td>27.5134</td>
</tr>
<tr>
<td>80</td>
<td>75.0042</td>
<td>27.1956</td>
</tr>
<tr>
<td>75</td>
<td>75.1708</td>
<td>27.1736</td>
</tr>
<tr>
<td>70</td>
<td>74.4208</td>
<td>26.9438</td>
</tr>
<tr>
<td>65</td>
<td>76.9208</td>
<td>26.4097</td>
</tr>
</tbody>
</table>

Fig. 6 Comparing obtained results of algorithm implementation with characteristics resulted from kernel- PCA and classification by both methods: 1-Elman artificial neural network (ANN-Elman) and 2- support vector machine with one against all method.

Fig. 7 Comparing algorithm implementation time with characteristics results from using Kernel-PCA by both methods: 1-Elman artificial neural network (ANN-Elaman) and 2- support vector machine with one against all method.

TABLE II

<table>
<thead>
<tr>
<th>Number of features from KPCA</th>
<th>ANN-Elman</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy percent</td>
<td>Elapsed time (second)</td>
</tr>
<tr>
<td>120</td>
<td>81.5000</td>
<td>19.3939</td>
</tr>
<tr>
<td>115</td>
<td>82.5833</td>
<td>14.6133</td>
</tr>
<tr>
<td>110</td>
<td>82.1667</td>
<td>13.9877</td>
</tr>
<tr>
<td>105</td>
<td>81.8333</td>
<td>13.8478</td>
</tr>
<tr>
<td>100</td>
<td>79.8333</td>
<td>13.5403</td>
</tr>
<tr>
<td>95</td>
<td>79.0000</td>
<td>13.3065</td>
</tr>
<tr>
<td>90</td>
<td>78.0833</td>
<td>13.0451</td>
</tr>
<tr>
<td>85</td>
<td>77.9333</td>
<td>12.8484</td>
</tr>
<tr>
<td>80</td>
<td>77.9333</td>
<td>12.3506</td>
</tr>
<tr>
<td>75</td>
<td>78.0000</td>
<td>12.5106</td>
</tr>
<tr>
<td>70</td>
<td>79.2500</td>
<td>12.2788</td>
</tr>
<tr>
<td>65</td>
<td>81.7500</td>
<td>11.7447</td>
</tr>
</tbody>
</table>

World Academy of Science, Engineering and Technology
International Journal of Materials and Metallurgical Engineering
Vol:5, No:10, 2011
V. CONCLUSION

In present study for ORL face bank image classification, Principle Components Analysis (PCA) and Kernel Principle Components Analysis (KPCA) methods by Elman neural network categorization and SVM methods are used. Also the effect of using PCA and KPCA from 65 components to 110 components is reviewed for system recognition rate and program implementation time. The similar tests are conducted by using same characteristics number with neural network and SVM and the obtained results from neural network and SVM are compared. The obtained results show that in similar test, SVM recognition correctness rate is more than that of neural network and also its implementation duration is longer than that of neural network. It is shown that by using PCA 110 components, recognition accuracy is 77.17% in neural network and 96.16% in SVM. The obtained results show that using KPCA increases SVM classification accuracy to 97.451% in the best states. For continuing this work it is considered to use other methods characteristics selection and dimension reduction as Generalized Discriminate Analysis (GDA) and other classification methods.

REFERENCES