



RESEARCH ARTICLE

A Face Recognition System Based on Principal Component Analysis-Wavelet and Support Vector Machines

Laith R. Fleah^{1*}, Shaimaa A. Al-Aubi²

¹Department of Computer Science, Cihan University-Erbil, Iraq, ²Department of Computer Science Department, College of Science, Salahaddin University, Erbil, Iraq

ABSTRACT

Face recognition can represent a key requirement in various types of applications such as human-computer interface, monitoring systems, as well as personal identification. In this paper, design and implement of face recognition system are introduced. In this system, a combination of principal component analysis (PCA) and wavelet feature extraction algorithms with support vector machine (SVM) and K-nearest neighborhood classifier is used. PCA and wavelet transform methods are used to extract features from face image using and identify the image of the face using SVMs classifier as well as the neural network are used as a classifier to compare its results with the proposed system. For a more comprehensive comparison, two face image databases (Yale and ORL) are used to test the performance of the system. Finally, the experimental results show the efficiency and reliability of face recognition system, and the results demonstrate accuracy on two databases indicated that the results enhancement 5% using the SVM classifier with polynomial Kernel function compared to use feedforward neural network classifier.

Keywords: Face recognition, feedforward backpropagation, neural network and K-nearest neighborhood, support vector machines

INTRODUCTION

Facial recognition is considered as one of the biometric systems and the importance of this field as a topic of research has increased in recent years. Face recognition system uses biometric information which takes from human, but the use of this system is more easily than in another biometric system such as fingerprints, iris, and signature, especially with the uncooperative people. Face recognition systems are usually used in security systems to monitor people and to discover any security breach, for example, face recognition system uses in the airport for monitoring and arrests the wanted. Therefore, face recognition system is used in many security applications such as avoiding crime, monitoring through video, and also for verification of the identity of persons.

Principal component analysis (PCA) is one of the linear transformation techniques that are used to obtain features from data or to decompress the information. It is also known as one of the famous techniques that are used to take global structures from the high-dimensional information set as well as it is applied to decompose dimensionality of the data and take unique features from the faces picture. This technique can be used also to distinguish patterns in data and express the data so as to highlight their differences and similarities between them.^[1]

Wavelet transforms are a very effective method and specifically in image compression and many other types of

data. As many coefficients of wavelet transform method become zero or very small. For this reason, wavelet transforms can consider as a very helpful tool for image compression. The primary benefit of wavelet transforms compared than other decomposition methods is that the basic functions that related with a wavelet decomposition typically have short and long support. The short support can effectively act on sharp transitions as the edges of the image; however, long support is operative in representing slow variations of the image.^[2]

Support vector machine (SVM) is one of important learning algorithms that are applied in many of applications and it is currently represented one of the very used methods in many application fields. The SVM theories were introduced at first in the sixties and seventies from the previous century by the scientists Vapnik and Chervonenkis, but the SVMs method is still not implementations practical until the early.^[3]

Corresponding Author:

Laith R. Fleah, Department of Computer Science, Cihan University-Erbil, Iraq. E-mail: laith.flaih@cihanuniversity.edu.iq

Received: Apr 18, 2019

Accepted: Apr 25, 2019

Published: Aug 20, 2019

DOI: 10.24086/cuesj.v3n2y2019.pp14-20

Copyright © 2019 Laith R. Fleah, Shaimaa A. Al-Aubi. This is an open-access article distributed under the Creative Commons Attribution License.

It has been applied in many of applications in many various fields such as medical diagnosis, text or image classification and also categorization, spam categorization, detection and recognition of an object, face detection, verification and recognition, bioinformatics, signal processing, prediction, information, and image retrieval.^[4]

K-nearest neighbor [KNN] is one of the most important algorithms used in face recognition system. This algorithm is easy to understand, but in another hand, it is useful in practice. Furthermore, it is used in several of applications such as vision, proteins, computational geometry, and graphs. It also considered as one of the best 10 algorithms in data mining.^[5,6]

KNN is considered as one of the non-parametric learning algorithms. This means that it does not produce any supposition on the implicit data distribution. This is very helpful in the real-world application, many of the practical data do not conform the classical theoretical supposition made such as linearly separable and Gaussian mixtures.^[5,6]

The overall structure of the paper is organized as follows: Section II: Introduced the structure of the system, this is included pre-processing methods, feature extraction methods, and classification methods. In Section III, the experimental results and its analysis are introduced while Section IV concludes the paper.

FACE SYSTEM STRUCTURE

Face recognition system that introduced in this paper consists of the following steps: Pre-processing, feature extraction, classification, and in the final steps is identification. The stages of the system are shown in Figure 1.

Pre-processing

Pre-processing stage is the first process that done in the proposed system. It is applied to improve the image and also to improve the recognition rate through remove the effect of the light on and also to correct the gradient of illumination in the image. In the following, the pre-processing techniques that applied at this stage to get the good results:

Normalization size of the image: Two types of database are used in this thesis and each database has different size of an image; therefore, this technique is used to make all images have the same size. This is because the next stages have an arithmetic operation such as multiplication and division of array. For all these reasons, the size normalization is very significant.

Contrast stretching: It is also known as histogram stretching. It is considered as one of the techniques that used to enhance an image. The goal of this technique has improved the contrast in an image through stretch the range of intensity values.

In other words, this is applied by raising the dynamic range of the gray levels in the image. For example, in an 8-bit system, the 256 gray levels can be shown only but if the number of gray levels in the captured image prevalence through a lesser range, the images can be improved by extended this number to a wider range.

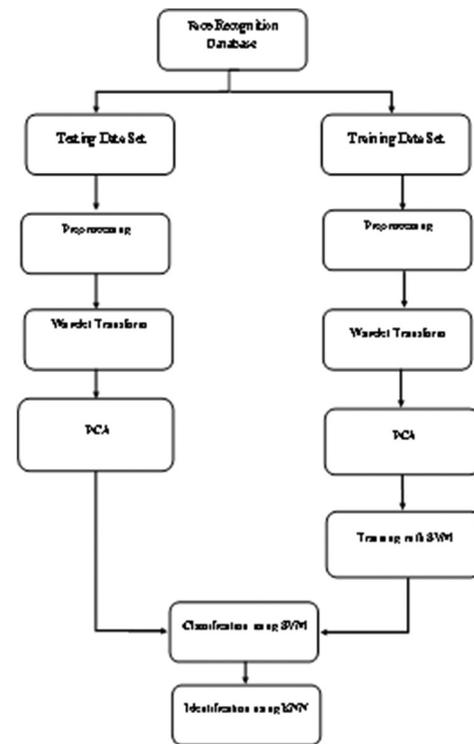


Figure 1: The structure of a proposed system

Features Extraction

The second stage of this system is feature extraction. It is used to obtain the significant features from the picture of the face and then applied these features to the next stage (identification and classification stages). Without this stage, the function of recognition and classification becomes very hard and very difficult because the picture consists of many features and most of these features cannot be applied to implement classification.

In this paper, each of PCA and wavelet transform techniques is used to obtain the significant features that are important to identify and classify the face pictures. After pre-processing stage is implemented, the face picture $I(x, y)$ that is a two-dimensional matrix $(N \times N)$ is then transformed to a vector and that length is had dimension N^2 . Thus, the picture with dimension 64×64 will transform to a vector of dimension 4096, but this vector cannot be applied directly to recognition directly because the face of human is very comparable from each other and another reason the 4096-dimension space is big and it is needed much time to process. Therefore, wavelet and PCA techniques are applied to obtain a better feature from face picture and also to enhance the capability of the recognition rate.

PCA Algorithm

At first, the training set is generated and this training set consists of a number of images (M) and each image in its to be transformed to the vector (N^2). The step after generated the training set is calculated the mean of images as this is described in Equation 1:

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_i \quad (1)$$

Γ_i in the question (1) refers to the image in the training group. Then, the normalized is implemented in training set to achieve the zero mean that describes the degree of different between the mean of image and image, and this is described in equation:

$$\varphi_i = \Gamma_i - \Psi_i \quad 1, 2, \dots, M \quad (2)$$

The covariance matrix C of the training set can be obtained using Equation 2.

$$\Psi = \frac{1}{M} \sum_{n=1}^M \phi_i \phi_i^T \quad (3)$$

If a vector u satisfies the condition in question (4), the eigenvector of the covariance matrix C is considered a non-zero vector.

$$C u_k = v_k u_k \quad (4)$$

Where, v_k refers to the corresponding eigenvalues.

The size of matrix C with the dimension of $N \times N$ is a large size and needs much computational time and size memory to obtain results from it, to solve this problem, the covariance matrix L that has small size compare than the original covariance matrix C is calculated. Equation 5 describes how the matrix C can be decomposed:

$$C = AA^T \quad (5)$$

Where, A represents zero means matrix $[\varphi_1, \varphi_2, \dots, \varphi_i]$ and the eigenvectors v_i can be consideration according to the following:

$$AA^T v_i = \mu_i v_i \quad (6)$$

Now, by multiply each side of Equation 6 by A , we can obtain the following:

$$AA^T A v_i = \mu_i A v_i \quad (7)$$

From Equation (7), we can see that represents the eigenvectors of the covariance matrix C and the matrix L with dimension $(M \times M)$ can be rewritten as shown in Equation (8):

$$L = AA^T \quad (8)$$

Now, we obtained the eigenvectors v_i of smaller covariance matrix L that has dimension M and this way, the linear combinations of the training data of the images to the form of Eigenfaces u_i can be obtained from equation below:

$$u_i = \sum_{k=1}^M v_{ik} \phi_k \quad (8)$$

This eigenvector that has symbol u_i is also called as Eigenfaces because when transfer these eigenvectors form a vector of length (N) to a two-dimensional matrix with dimension $(N \times N)$ and show it. It is shown as the human face.

Wavelet Transform

It is the first step of this stage and it is used wavelet transform to remove the noise from the image with saving the quality of

an image. The second benefit is to obtain some of a feature from the image and also to decompose the size of the image and by this way, thus reduced the time of the process.

The output of the wavelet transform on an image, it consists of four sub-bands. These sub-bands are the following: Low-low (LL), high-low (HL), low-high (LH), and high-high. The LL sub-band gives the most important features of the image and HL and LH sub-bands give the horizontal edges and vertical edges, and finally, the high sub-band gives the diagonal edges. The results that obtained after applied wavelet transform are shown in Figure 2.

When this technique is applied for 1 time to the image with dimensions (112×92) , the dimensions of the four sub-bands are 56×46 . We can conclude from the previous paragraph after used one stage of the wavelet transform, the size of the image is decreased to the half. In this work, two levels of the wavelet transform are used.

Classification Stage

A classification is the third stage of this recognition system. It is used after each of the previous stages is done. It is a very significant stage. The goals of this stage are classified the interimages based on the information that obtained through the training stage. There are a number of classification methods that can be used to do this job. In this recognition system, SVMs that are considered as one of the famous classification methods are used to do this function.

SVMs are one of the most important learning algorithms which are applied in many of applications. In this work, we also aim to make a comparison between different kinds of the SVMs kernel to analyze its results and determine any kernels of them can give a good result with high accuracy rate and also we make a comparison with a feedforward backpropagation neural network (FFBPNN) that used as a classifier to ensure from the power of our methods.

SVM classifier

SVM is considering as learning algorithm method depended on statistical learning algorithms. It is used to

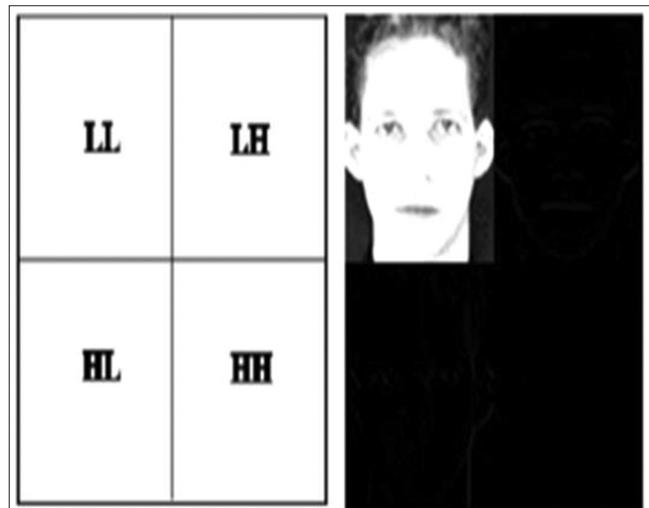


Figure 2: The result after used one level of the wavelet transform

classify the feature of the data by finding a maximum marginal hyperplan that separating between two classes. It also can be applied for linear and non-linear classification using different types of kernel functions.

Basic SVM formulation

SVMs are used to make classification among two classes to give a decision surface. This decision surface is represented a maximum distance that can depend on it to separate closest points in the classification’s class. The closest points are known as support vectors. Figure 3 describes the hyperplane of support vector separating.

In this part of the chapter, we will consider SVMs as maximum margin classifiers. To facilitate the condition that supposed, we will consider that input data can separate linearly. Under this supposition, we can separate any two classes by obtained the linear hyperplane between them. Figure 3 describes the good and bad separation between two classes.

Let us summed that, a binary classification with x_i as the input data, where $i = 1, 2, 3, \dots, L$. with identical labels for the two class, now let assumed that the function of decision knows as shown in the question below:

$$f(x) = \text{sign}(w x + b) \tag{9}$$

Where, symbol represented the inner product and symbol b represent the bias of the function and point x take position directly on the hyperplane and it achieves the condition.

$$w x + b = 0 \tag{10}$$

From calculating the left side of Equation (10), we can obtain the label y_i of a data, where sign refers to the label of class. The point’s x_i on both sides of the hyperplane must achieve the following conditions:

$$x_i w + b < 10 \tag{11}$$

$$x_i w + b > 10 \tag{12}$$

Equations (11) and (12) can be rewritten as follows:

$$y_i (x_i + b \geq 10) \tag{13}$$

Using the formulation of Lagrangian, the SVMs prediction can obtain using the following equation:

$$f(x) = \sum_{i=0}^m y_i \alpha_i < x, x_s \tag{14}$$

Where, m is represented the number of support vectors, x_i is a support vector, and α_i is represented the corresponding Lagrange multiplier, and finally, a sign of $f(x)$ is classified every test vector.

We can say that the output of classification is correct if Equation (14) holds for all input points. The number of possible combinations of both weights and bias are big and sometimes not optimal. In general, in the binary classification, there is one ideal separating hyperplane is offer. On the ideal hyperplane, the margin among two data is maximized. As shown in Figure 4, the margin among two sets of data is represented by $d+$ and $d-$ which represented the distances between them. Figure 5 describes the two separating lines, one of them is ideal and another is random.^[7]

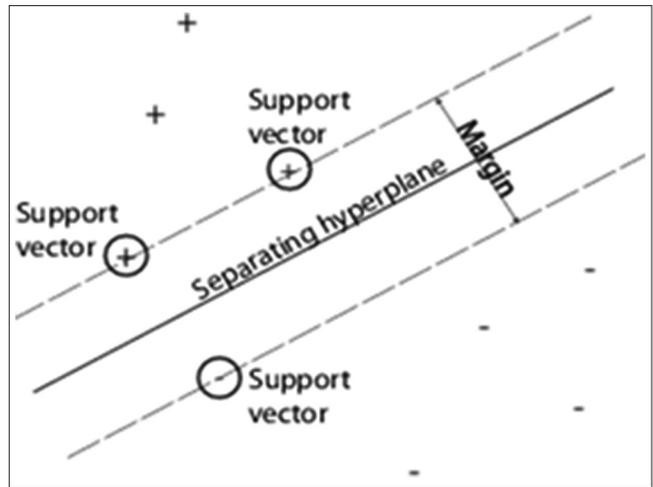


Figure 3: The support vector and the separating hyperplane

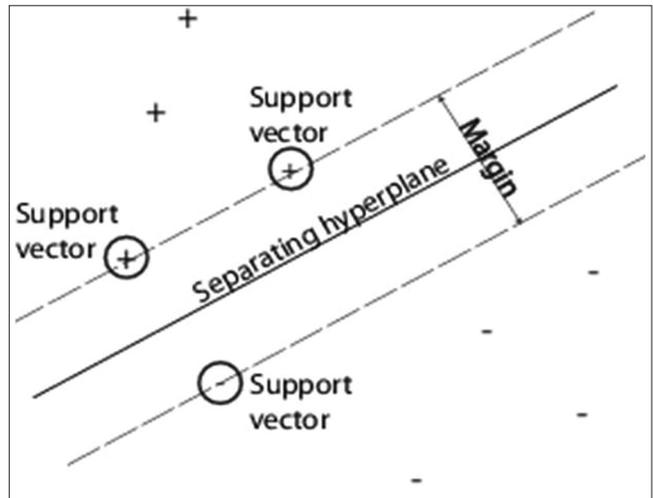


Figure 4: The support vector and the separating hyperplane



Figure 5: Samples of Yale database

Identification

The final stage of the system is identification. The aim of this stage is to find the nearest image for the new image

that classified before that using SVM or ANN. KNN is used to implementing that KNN is not like SVM or ANN. It does not need to the training phase, and therefore, all the data are needed for the testing phase. The data that need for testing consist of a set of vectors and class label associated with each vector only.

KNN for classification

In the section, we will see how can be used KNN as a classifier. To do that, two data sets will create. One of them is training set and another set is testing set. The goal of this algorithm is to obtain the class label of the new point. The algorithm has various conducts depended on the value of k parameter.

The first hypothesis: The value of k parameter is equal to one ($k=1$)

To explain the first hypothesis, let consider that x is the point to be labeled and to obtain the point that closest to it. Let consider the closest point y . In this way, the NN rule asks to assign the label of y to x . This appears very simplistic, but this way is correct only when the number of points is not very large.

However, in other situations, when the number of data points is very big. The label of x and y may be same. To simplest the explain – let assume that the potentially biased coin. Then, maybe the head will appear in next call. The same argument can be applied in this situation.

Now, lets us apply another case here – suppose that every point is represented in a D -dimensional plane and the number of points is very huge. This led to conclude that the density of the plane on each point of data is high. In other description, in each subspace, a large number of points is offered and consider x point is in the subspace and is many of neighbors, but in this time, consider y be the NN to the point x .

We can say that, if x point and y point are very close together, then the following can propose, the probability of point x and point x related to the one class is very big, then using decision theory, the following is obtained, both points related to the one class.

Bellhumer *et al.*^[8] provided a very good argument about the rule of NN. One of the important results is to obtain a very small error bound. This bound is described in question:

$$p^* \leq p \leq p^* \left(2 - \frac{c}{c-1} p^*\right) \quad (15)$$

Where the Bayes error rate, the symbol c is represented the number of classes and the symbol P is represented the error rate of NN. The result is given the following conclusion if the number of points very huge, then the error rate of NN becomes less than twice time from the Bayes error rate and this is very good results obtained from algorithm as KNN.

The second hypothesis: The value of k parameter is equal to two ($k=2$)

In general, we aim to obtain the KNN and then make a majority voting. As we mentioned previously, the value of k is odd when there are two numbers of classes. Let's assumed that k is equal 5 and there are three instances of C1 and two instances of C2. In this situation, KNN method will label the new point as C1. This is similar when there is multiclass want to label.

One of the KNN techniques is not given 1 to every neighbor and another common technique is give weight to each point according to the distance that calculated.

In another way, the values of each weight of point are inversely proportional to the distance of the point that will classify. From the sentence above, the neighboring points have higher values compared than the ultimate points.

EXPERIMENTAL RESULTS

A number of experiments on the two different databases are used. The first experiment is applied on the Yale database^[8] and the second experiment is applied on the ORL database.^[9] Both these database are very famous databases and are used by many researchers. In every database, two experiments are done. The first experiment is compared between different SVMs kernels to show any kernel give best results and the second experiment is between SVM classifier and neural network classifier to calculate the performance of the classifier that used the system.

Yale Experiment

Yale database is used in this experiment. This database is created by Yale University and it is provided free on the internet and anyone can download it. This database contains from 15 subjects and each subject has 11 images take under various conditions. From that mention above, this means that the database consists of 165.

This database is applied because it consists of a number of images captured at various conditions; this is included lighting condition or other things; therefore, it can be applied to test the recognition rate of the proposed system and it is also very widely used by many of researchers.

In our experiment, the database is divided into two groups; the first group is for train and the second group is for the test. The training group is applied to learn the system and the training group consists of five images of the first 12 classes. The testing group is applied to calculate recognition rate of the proposed system and it consists of the other five images of the same 12 classes as well as all images of the last three classes. This is meant that 40% of the database is applied for training and 60% of the image is applied for the test.

In this experiment, the combinations of the PCA-wavelet transform are applied for pre-processing and feature extraction, and comparison between various types of the SVMs kernel is done. Figure 6 describes the results of the SVMs classifier using various kernel functions in Yale database.

In the second experiment, FFBPNN is applied as classifier instead of the SVMs. This is done to make a comparison between the results of SVMs and ANNs classifiers as well as to calculate the performance of the system. Figure 7 describes the recognition rate results of the SVMs classifier using polynomial kernel function and artificial neural network.

ORL Experiment

ORL database is used in the second experiment. This database is created by Cambridge University computer and anyone can download it from the internet without fees. This database

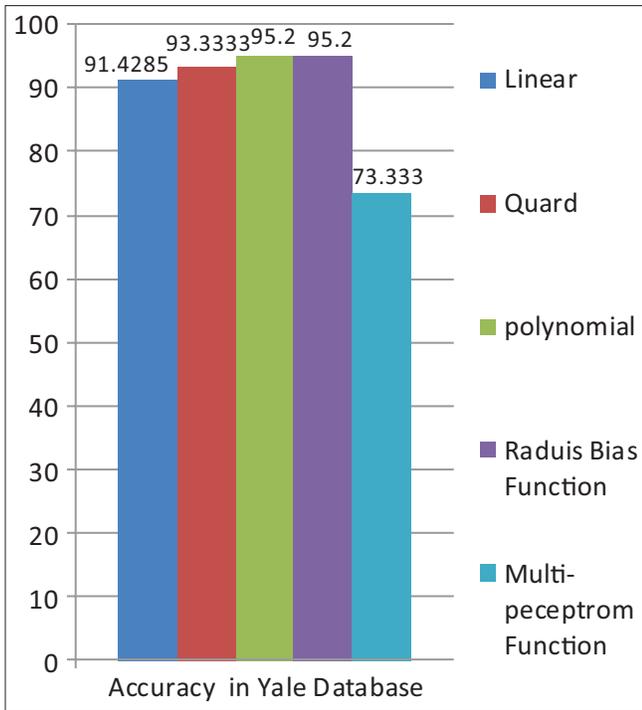


Figure 6: The results of the support vector machines classifier using different kernel functions in Yale database

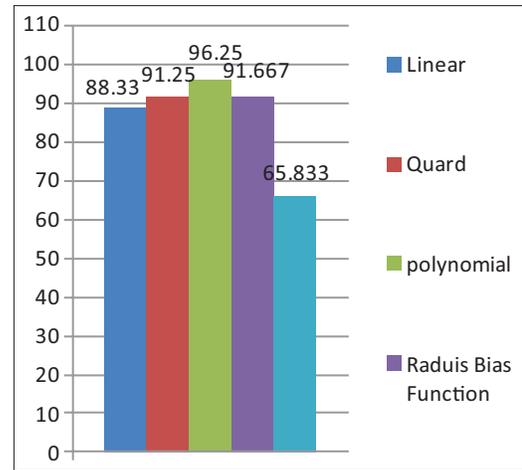


Figure 8: The results of the support vector machines classifier using different kernel function in ORL database

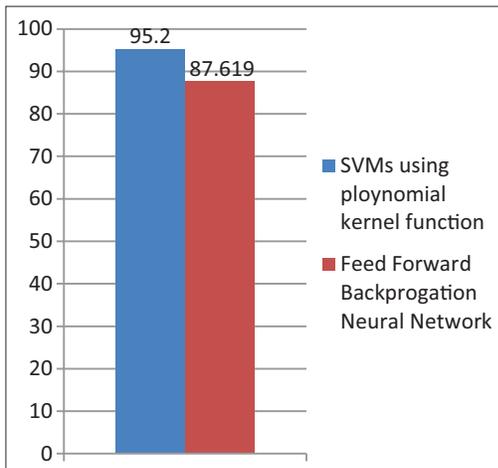


Figure 7: The accuracy results of the support vector machines classifier using polynomial kernel function and artificial neural network

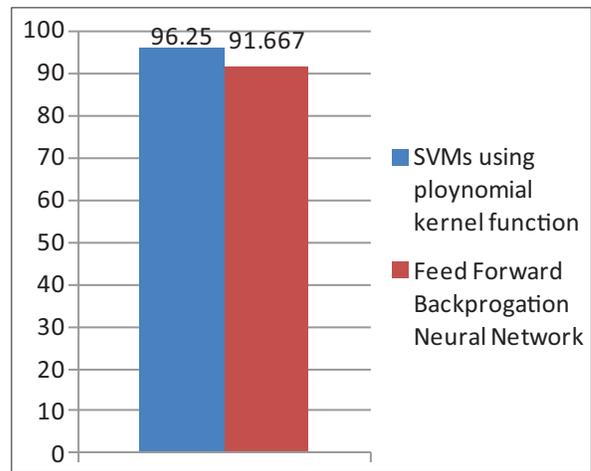


Figure 9: The accuracy results of the support vector machines classifier using polynomial kernel function and artificial neural network

consists of 40 classes with 10 images for each class. This means that database consists of 400 images.

In our experiment, the database is divided into two groups; the first group is for train and the second group is for the test. The training group is applied to learn the system and the training group consists of five images of the first 32 classes. The testing group is applied to calculate recognition rate of the proposed system and it consists of the other six images of the same 32 classes as well as all images of the last eight classes. These three classes are used for test only. This is meant that 40% of the database is applied for training and

60% of the image is applied for the test. Figure 8 describes the results of the SVMs classifier using different kernel function in ORL database.

In the second experiment, FFBPNN is applied as classifier instead of the SVMs. This is done to make a comparison between the results of SVMs and ANNs classifiers as well as to calculate the performance of the system. Figure 9 describes the recognition rate results of the SVMs classifier using polynomial kernel function and feedforward backpropagation neural network classifier.

CONCLUSIONS

Face recognition systems are one of the computer vision applications and it is significant in many of. In this work, face recognition system is designing and implementation. In this system, combines between each of wavelet transform and PCA methods are applied for obtained feature from the image after that SVM is applied to classify these features. The proposed system is tested using two types of databases to guarantee from the power of the system. Comparison between the results

of proposed system and FFBNN of a classifier is implemented. According to the obtained results, the following conclusions can be stated:

- Applying Wavelet-PCA feature extraction algorithms enhanced the recognition rate of the system compared to the using PCA only
- Using SVM classifier based on polynomial kernel function increases the accuracy of recognition rate of the system by 5% compared to feedforward neural network classifier, especially in cases where the features of a person's face are modified with different hairstyles or wearing glasses.

The comparison between SVM and FFBNN classifiers proved that SVM classifier is more accurate than FFBNN classifier in recognition.

REFERENCES

1. S. Gupta, O. P. Sahu, R. Gupta and A. Goel. "A bespoke approach for face-recognition using PCA". *International Journal of Computer Science and Engineering*, Vol. 2, no. 02, pp.155-158, 2010.
2. M. David. "Reversible Integer to Integer Wavelet Transforms for Image Coding". Thesis, University of British Columbia, 2002.
3. J. S. Taneja. "Analysis of *E. coli* Promoters Using Support Vector Machine". Thesis, Thapar Institute of Engineering and Technology, 2006.
4. M. E. Mavroforakis. "Geometric Approach to Statistical Learning Theory Through Support Vector Machines (SVM) with Application to Medical Diagnosis". Thesis, National and Kapodistrian University of Athens, 2008.
5. X. Wu, V. Kumar, J. R. Quinlan and J. Ghosh. "Top 10 algorithms in data mining". *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008.
6. M. Zuhaer and N. Al-Dabagh. "Face recognition using LBP, FLD and SVM with single training sample per person". *International Journal of Scientific and Engineering Research*, vol. 5, no. 5, pp. 180, May 2014.
7. R. O. Duda, P. E. Hart and D. G. Stork. "*Pattern Classification*". 2nd ed. New York: John Wiley and Sons, Inc., 2000.
8. P. N. Bellhumer, J. Hespanha and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
9. F. Samaria and A. Harter. D. Parameterisation of a Stochastic Model for Human Face Identificatio. Conference: Applications of Computer Vision. Proceedings of the Second IEEE Workshop, 1994.