Face Recognition Using Stationary wavelet transform and Neural Network with Support Vector Machine

Nashwan Alsalam Ali*
Department of Educational and Psychological Sciences, College of Education for Women, University of Baghdad, Baghdad, Iraq.

Abstract
Face recognition is a type of biometric software application that can identify a specific individual in a digital image by analyzing and comparing patterns. It is the process of identifying an individual using their facial features and expressions.
In this paper we proposed a face recognition system using Stationary Wavelet Transform (SWT) with Neural Network, the SWT are applied into five levels for feature facial extraction with probabilistic Neural Network (PNN) , the system produced good results and then we improved the system by using two manner in Neural Network (PNN) and Support Vector Machine(SVM) so we find that the system performance is more better after using SVM where the result shows the performance of the system is better based on the recognition rate measurement.

Keywords: Stationary Wavelet Transform (SWT), Support Vector Machine (SVM).

تمييز الوجوه باستخدام تحويل المويجة المستقرة والشبكة العصبية مع ماكنة الدعم الناقل

نشوان السلام علي *
قسم العلوم التربوية والنفسية، كلية التربية للبنات، جامعة بغداد، بغداد، العراق

الخلاصة:
التعرف على الوجه هو نوع من تطبيقات البرمجيات القياسية التي يمكن من خلالها تحديد شخص معين في صورة رقمية عن طريق تحليل ومقارنة الأنماط. هو عملية تحديد الفرد باستخدام ملامح الوجه والتعبيرات.
في هذا البحث تم اقتراح نظام التعرف على الوجوه باستخدام تحويل المويجات المستقرة مع الشبكات العصبية، تحويل المويجات الثابتة تطبق لخمسة مستويات لاستخراج الوجه مع ميزة الشبكة العصبية الاحتمالية (PNN). ، نتائج النظام جيدة ومن ثم قمنا بتحسين النظام باستخدام ماكنة الدعم الناقل (SVM) لذلك نجد أن أداء النظام هو أفضل بعد استخدام SVM حيث تظهر نتيجة أداء النظام بشكل أفضل اعتمادا على مقياس معدل التمييز.

*Email: nashwan_alsalam60@yahoo.com

1. Introduction

The safety of persons, goods or information is one of the major preoccupations our societies today. Also, the great weakness of the current means of identity verification is clear here: the identity of a person is directly related to what it owns
(a passport, magnetic badge, etc…). However, a badge can be stolen, guessed password or broken by brute strength algorithms: this leads to identity theft [1].
The Stationary Wavelet Transform (SWT) is a wavelet transform algorithm designed to overcome the lack of translation-invariance of the Discrete Wavelet Transform (DWT). Translation-invariance is achieved by removing the downsamplers and upsamplers in the DWT and upsampling the filter coefficients by a factor of $2^{(j-1)}$ in the *j*th level of the algorithm. The SWT is an inherently redundant scheme as the output of each level of SWT contains the same number of samples as the input, so for a decomposition of N levels there is a redundancy of N in the wavelet coefficients [2].
Face recognition is considered to be an important part of the biometrics technique, and meaningful in scientific research. It is the ability to establish a subject's identity based on his facial characteristics. Automatic face recognition has been extensively studied over the past two decades due to its important role in a number of application domains, such as access control, visual surveillance [3].
Many methods have been proposed for face recognition. Fusion of images exploits synergistic integration of images obtained from multiple sensors and by that it can gather data in different forms like appearance and anatomical information of the face, which enriches the system in improving recognition accuracy [4]. As a matter of fact fusion of images has already established its importance in case of image analysis, recognition, and classification. For instance, Aglika Gyaourova et al [5] tried to implemented pixel-based fusion scheme in the wavelet domain, and feature based fusion in the eigen space domain. Although their fusion approach was not able to fully discount illumination effects present in the visible images but they showed substantial improvements in overall recognition performance. They also indicated that IR-based recognition performance degrades seriously when eyeglasses are present in the probe image but not in the gallery image and vice versa. On the other hand for the improvement of the performance of face recognition, when face images are occluded by wearing eye-glasses, Jeong-Seon Park et al [6] first detect the regions occluded by the glasses and generate a natural looking facial image without glasses by recursive error compensation using Principal component analysis (PCA) reconstruction. They proposed a new glasses removal method based on recursive error compensation using PCA reconstruction. George Bebis et al [7] investigated that two different fusion schemes like first one is pixel based and operates in the wavelet domain using Haar transforms, while the second one is feature-based and operates in the eigenspace domain. In both cases, they employ a simple and general framework based on Genetic Algorithms (GAs) to find an optimum fusion strategy. Amit Aran et al [8] demonstrated the spectral band invariant Wave MACH filters which are designed using images of CCD/IR camera fused by Daubechies wavelet transform and implemented in hybrid digital optical correlator architecture to identify multiple targets in a scene . Mrinal Bhowmik et al [9] they have fusion of infrared and CCD camera because the performance of CCD camera is better under good illumination conditions where as IR camera gives a better output under poor illumination or in the night conditions also.

2. Stationary Wavelet Transform

The Discrete Wavelet Transform (DWT) is not a time- invariant transform. The way to restore the translation invariance is to average some slightly different DWT, called un-decimated DWT, to define the stationary wavelet transform (SWT). It does so by suppressing the down-sampling step of the decimated algorithm and instead up-sampling the filters by inserting zeros between the filter coefficients. Algorithms in which the filter is up sampled are called "a trous", meaning "with holes". As with the decimated algorithm, the filters are applied first to the rows and then to the columns. In this case, however, although the four images produced (one approximation and three detail images) are at half the resolution of the original; they are the same size as the original image [10].
In summary, the SWT method can be described as follows: At each level, when the high-pass and low-pass filters are applied to the data, the two new sequences have the same length as the original sequences.

To do this, the original data is not decimated. However, the filters at each level are modified by padding them out with zeros. Supposing a function $f(x)$ is projected at each step J on the subset V_J (....□V_3□V_2□V_1□V_0). This projection is defined by the scalar product $c_{i,k}$ of $f(x)$ with the scaling function $\phi(x)$ which is dilated and translated.

$$c_{i,k} =< f(x), \phi_{i,k}(x)> \quad \dots (1)$$

$$\phi_{i,k}(x) = 2^{-j}\phi(2^{-j}x-k) \quad \dots (2)$$

Where $\phi(x)$ is the scaling function, which is a low-pass filter. $c_{i,k}$ is also called a discrete approximation at the resolution 2^j If $\phi(x)$ is the wavelet function, the wavelet coefficients are obtained by

$$\omega_{j,k} =< f(x), 2^{-j}\phi(2^{-j}x-k) > \quad \dots (3)$$

$\omega_{j,k}$ is called the discrete detail signal at the resolution 2^j . As the scaling function $\phi(x)$ has the following property:

$$\frac{1}{2}\emptyset\left(\frac{x}{2}\right) = \sum_n h(n)\emptyset(x-n)$$

$h(n)$ represents the scaling coefficients. $c_{j+1,k}$ can be obtained by direct computation from $c_{i,k}$.

$$c_{j+1,k} = \sum_n h(n-2k)c_{j,n}$$

and

$$\frac{1}{2}\varphi(\frac{x}{2}) = \sum_n g(n)\phi(x-n) \quad \dots (4)$$

t coefficients. The scalar products

$\omega_{j,k} =< f(x), 2^{-(j+1)}\varphi(2^{-(j+1)}x-k) >$
are computed with

$$\omega_{j+1,k} = \sum_n g(n-2k)c_{j,n} \quad \dots (5)$$

Equations (4) and (5) are the multiresolution algorithm of the traditional DWT. In this transform, a downsampling algorithm is used to perform the transformation. That is, one point out of two is kept during transformation. Therefore, the whole length of the function will be reduced by half after the transformation. This process continues until the length of the function becomes one. However, for stationary or redundant transform, instead of downsampling, an upsampling procedure is carried out before performing filter convolution at each scale. The distance between samples is increased by a factor of two from scale to the next. $c_{j+1,k}$ is obtained by

$$c_{j+1,k} = \sum_l h(l)c_{j,k+2^j l} \quad \dots (6)$$

And the discrete wavelet coefficients

$$\omega_{j+1,k} = \sum_{l} g(l) c_{j,k+2^j l} \qquad ...(7)$$

The redundancy of this transform facilitates the identification of salient features in a signal, especially for recognizing the noises. This is the transform for one-dimensional signal. For a two- dimensional image, we separate the variables *x* and *y* and have the following wavelets.
- Vertical wavelet: $\varphi^1(x,y) = \varphi(x)\ \varphi(y)$
-Horizontal wavelet: $\varphi^2(x,y) = \varphi(x)\ \varphi(y)$
-Diagonal wavelet: $\varphi^3(x,y) = \varphi(x)\ \varphi(y)$
Thus, the detail signal is contained in three sub images [11].

$$\omega_{j+1}^{1}(k_x,k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l=-\infty_y}^{+\infty} g(l_x) h(l_y) c_{j,k+2^j}(l_x,l_y) \qquad ..(8)$$

$$\omega_{j+1}^{2}(k_x,k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l=-\infty_y}^{+\infty} h(l_x) g(l_y) c_{j,k+2^j}(l_x,l_y) \qquad ...(9)$$

$$\omega_{j+1}^{3}(k_x,k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l=-\infty_y}^{+\infty} g(l_x) g(l_y) c_{j,k+2^j}(l_x,l_y) \qquad ...(10)$$

3. Neural Network

A neural network is an information-processing system that has been developed as generalizations of mathematical models matching human cognition. They are composed of a large number of highly-interconnected processing units (neurons) that work together to perform a specific task. A neural network is a massively parallel distributed processor that has a natural prosperity for storing experimental knowledge. It resembles the brain in two respects:

● Knowledge is acquired by the network through a learning process;
● Inter-connected connection strengths known as synaptic weights are used to store the knowledge;

Each neuron has an internal state called it threshold. An activation function (or transfer function) used for classifying vectors. Neural classification generally comprises of four steps:

1. Pre-processing, e.g., atmospheric correction, noise suppression, band rationing, Principal Component Analysis, etc;
2. Training - selection of the particular features which best describe the pattern;
3. Decision-choice of suitable method for comparing image patterns with the target patterns.
4. Assessing the accuracy of the classification [12].

There are many different types of neural networks from relatively simple to very complex just as there are many theories on how biological neural net work and branch out to other paradigm later. A layered feed forward neural network has layers of sub-groups of processing elements. A layer of processing elements makes independent computation or data that it receives and passes the result to another layer. The next layer may in turn make it independent computations and pass on the results to yet another layer. Finally a sub-group of one is more processing elements determine the output from the network. Each processing element makes its computation based upon a weighted sum of its inputs. The first layer is the input layer and the last is the output layer [13].

4. Feature Extraction

The representation of an image to a small number of components carrying enough discriminating information is referred to as “feature extraction “. In this proposed system the face image features are the detail and facial information of this image. The features are extracted by applying SWT five levels for test image and training image, Olivetti Research Laboratory (ORL) used as a face data base, we used 400 images in data base 10 for each person. In SWT low-frequency component, which are useful in recognition are reinforced. The SWT as shown in figure -1 is used to derive the multiresolution facial feature detail in the image.

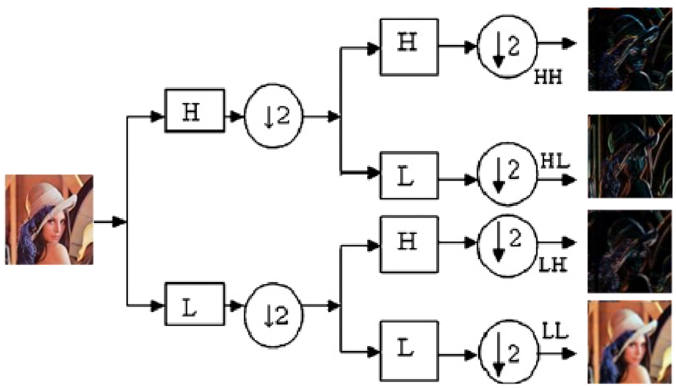

Figure 1- Multiresolution Approach for Image Decomposition

After the five levels of decomposition completed, so we will have (16) sub image but the low frequency sub images are ignored since the most relevant image information has been removed by interactively low pass filtering.

Finally, the normalized energy will be computed for the result (15) sub band images only with high frequency. For each sub band images an energy value will be computed.

The normalized energy was computed on each sub band image and defined as:

$$E = \frac{1}{NM} \sum_{I=1}^{N} \sum_{J=1}^{M} X^2_{i,j} \qquad \ldots(11)$$

Where N and M is the size of rows and columns of the X and N=M

the energy of the low frequency sub band image are generally not considered as face features. All images can be trained using a SWT to extract the features vector. In the tested image the features vector is extracted; finally the matching between the two features vectors can be made in order to classify the tested image by using Neural Network.

When the image is chosen and read for features extraction file, first the size of image is to be converted to (128*128) pixels before applying SWT for this image.

The images in the training set, the first 10 features in the features vectors file represent the first class and it becomes the references for person no. one and the second 10 features in the features vectors file represent the second class and it becomes the references for person no. two and so on until reaching the last 10 features in the features in the features vectors file which represent the last class and it becomes the references for person no. 40.The figure -2 below shows feature extraction steps.

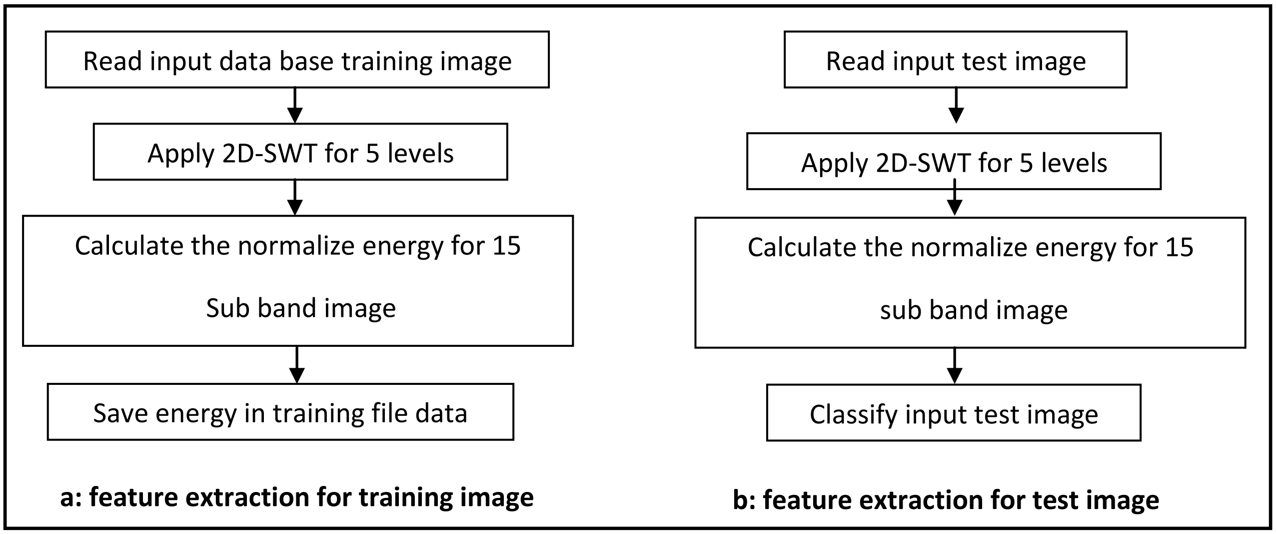

Figure 2 - Block Diagram for Feature Extraction Steps

5. Support Vector Machines

Support vector machines are learning machines that classify data by shaping a set of support vectors. SVMs provide a generic mechanism to robust the surface of the hyper plane to the data through. Another benefit of SVMs is the low expected probability of generalization errors. Moreover, once the data is classified into two classes, an appropriate optimizing algorithm can be used if needed for feature identification, depending on the application. SVM creates a hyper-plane between two sets of data for classification; in our work, we separate the data into two classes: face belongs to the train database and face doesn't belong to the train database. (XT•W– b) > 0, are labeled as +1 and those that fall on the other area, (XT•W–b) < 0, are labeled as -1.

We seek the linear classifier that separates the data with the lowest generalization error. Intuitively, this classifier is a hyper plane that maximizes the margin error, which is the sum of the distances between the hyper plane and positive and negative examples closest to this hyper plane. We consider the example in figure -3-a) where there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin shown in figure -3-b. This classifier is termed the optimal separating hyper-plane (OSH) [1].

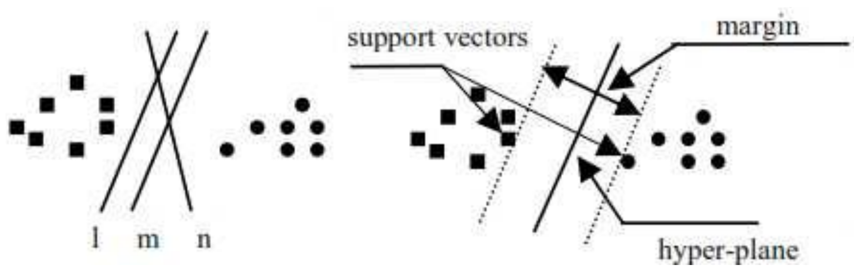

Figure 3 - Linear Classifier; (3-a) Arbitrary hyper-planes: l, m, n; (3-b) Optimal hyper-plane.

6. Proposed System

The following algorithm describes the main steps of face recognition system based on SWT and PNN with SVM.

Algorithm.

1. Read the test image.
2. Perform preprocessing to resize the test image to (92*112) as the same of data base.
3. Apply 5 levels 2D-SWT to test image to obtain 15 sub bands of coefficients, in each level (LLi, LHi, HLi and HHi) where i represent the number of the level.
4. Calculate the energy just for (LH, HL and HH) with discarding the LL in each level, this known as (feature extraction) and obtain features extraction for 15 sub bands.
5. Design the Neural network including the flowing steps:
    - Load features vector for all training data base.
    - Enter feature vectors for test image to the neural network.
    - Multiply the features vector of test image with all the data base features vectors.
    - Estimate the summation of each person (10 images for each) in summation layer.
    - Estimate the summation for all data base.
    - Estimate the probability of each person using the equation:
      P= summation of each person/ summation of all data base.
    - Specify the largest p of this class to the test image.
6. If not find face calculates SVM.

The support vector machine (SVM) has been computed between the new (testing) image feature vector and feature vector of the training image set to classify the class of the input test image.

1. Calculate threshold
2. If SVM > threshold go to step 7 Else
3. Specify the test image to the class of image (i) which has the minimum value of SVM

7. End.

Figure -4 demonstrates the complete system using the PNN and SVM together.

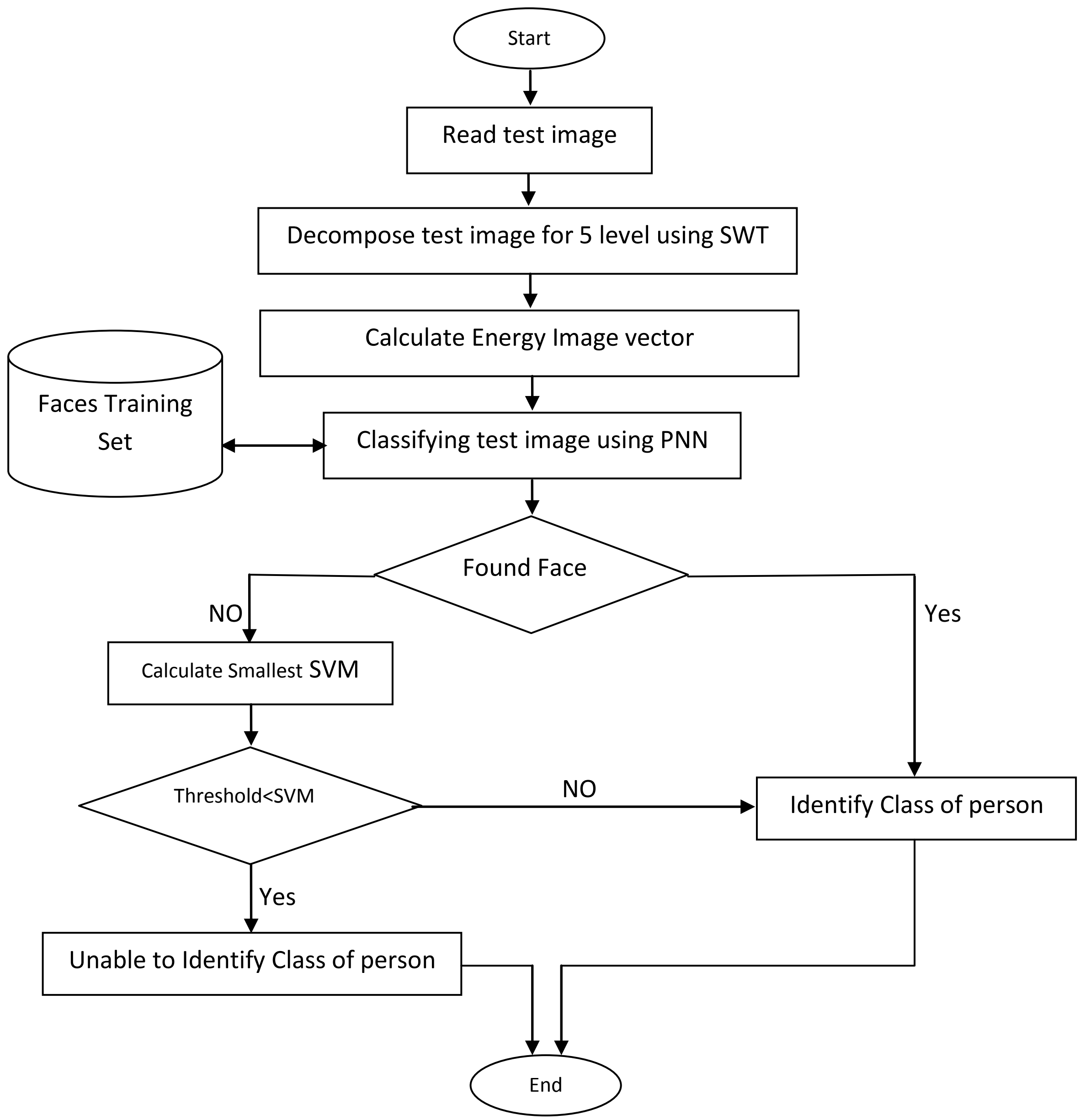

Figure 4 - Flowchart of Complete Proposed Recognition System

7. Experimental result

In this section we implemented and applied the proposed system using face database Olivetti Research Laboratory (ORL). Our database of faces has ten different images for each for 40 distinct subjects. For some subjects, the image were taken at different times, varying the lighting, facial expressions (open/ closed eyes, smiling/ not smiling) and facial details (glasses / on glasses). All the images were taken against a dark homogeneous background with the subjects in an upright,

frontal position with tolerance for some side movement. The size of each image is 92*112 pixels with 256 grey levels per pixel. The ORL database was illustrated in figure -5 which shows a sample from this data base and can be downloadable from the AT&T Laboratories, Cambridge, U.K., “The ORL Face Database”, and Available at [14].

Figure 5 - Sample from ORL Database (400 images: 40 people, 10 images each)

The system recognition rate was computed according to the formula below

$$\text{Recognition rate} = \frac{\text{No. of correct test}}{\text{Total No. of test}} \text{ X } 100\% \qquad \ldots 12$$

Our system has a recognition rate equal to 96% using normal image test, 81% with test image added Salt&pepper noise, 77% with test image added Gaussian noise, 73% processed with median filter and 66% Gaussian noise and Wiener filter, these results without using SVM. When we make a combination process between probabilistic neural network (PNN) and Support Vector Machine (SVM) the system performance rate will be increased. The system recognition rate will be equal to 99% using normal image test, 92% with test image added Salt & pepper noise, 89% with test image added Gaussian Noise, 82% with Median filter and 78% with Gaussian noise and Wiener filter. Figure -6 presents one example when using these types of noises and filters.

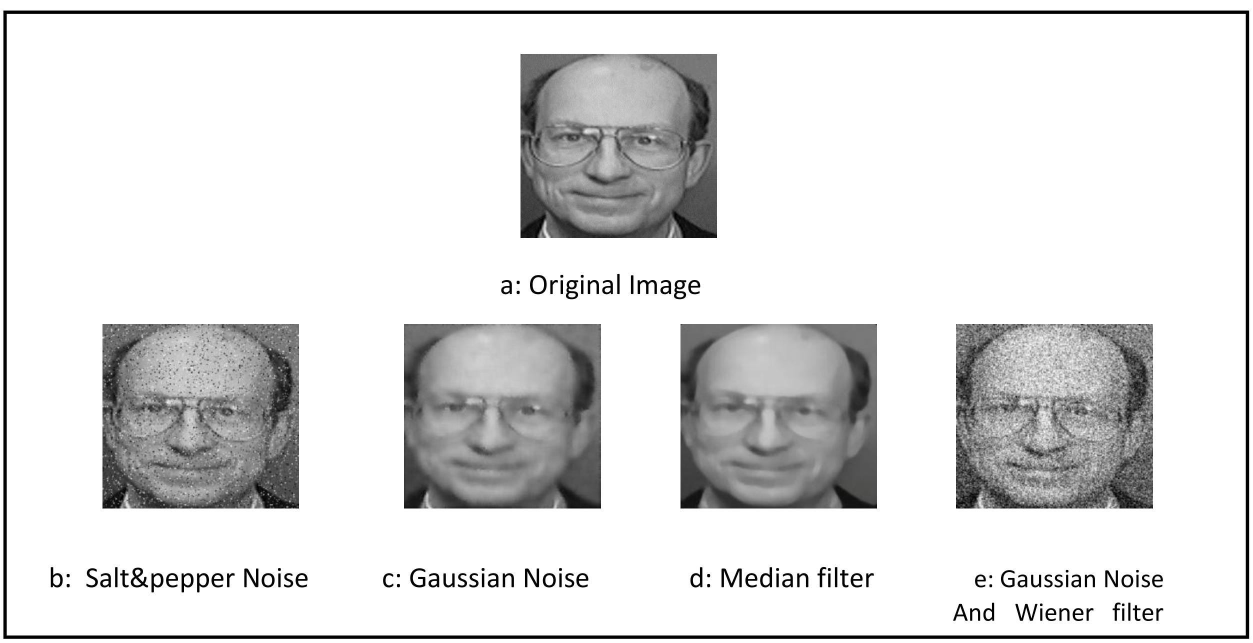

Figure 6 - Noise type and filters

Table -1 explains the difference in results for recognition rate before using SVM and after using it.

Table 1- Results without using SVM and with using SVM

Test image	Without SVM	With SVM
Normal image	96%	99%
Add Salt&pepper	81%	92%
Add Gaussian noise	77%	89%
Add median filter	73%	82%
Add Gaussian noise and Wiener filter	66%	78%

Figure -7 and figure -8 show the difference in distribution of recognition rate before and after using SVM.

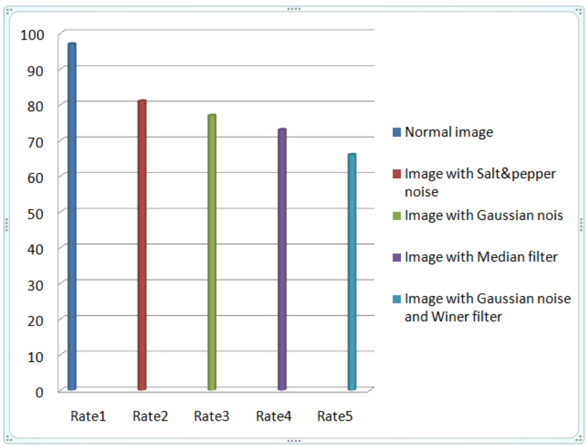

Figure 7- Recognition rate before using SVM

Figure 8- Recognition rate after using SVM

8. Conclusion

In this paper a neural Network and SWT are combined for presenting a face recognition system, this system performs human face recognition at a very high degree accuracy, which uses the SWT to derive the frequency facile features. This transforms is an exactly shift-invariant transform.

The results show that this system approach has a higher classification accuracy and the robustness to passion noise and some damaged images. It is also invariant to illumination and facile expression variations.

In our experiment, proposed SWT that applied for 5 levels decomposition combined with Neural network classifier shows very good performance which can achieve 96% maximum correct recognition rate on ORL data set, better result achieved when SVM is used it nearby %99 with normal image test.

References

1. Bellakhdhar, F., Loukil, K. and ABID, M . **2013**. Face recognition approach using Gabor Wavelets, PCA and SVM, *IJCSI International J. Computer Science* 10(3).
2. http://en.wikipedia.org/wiki/Stationary_wavelet_transform
3. Anil, K. J., Brendan, K. and Unsang, P. **2011**. Face Recognition: Some Challenges in Forensics, *IEEE International Conference Automatic Face & Gesture Recognition and Workshops*, pp: 726-733.

4. Pavlidis, I. and Symosek, P. **2000** .The imaging issue in an automatic face/disguise detection system, *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp:15–24.

5. Gyaourova, A. , Bebis, G. and Pavlidis, I. **2004**. Fusion of infrared and visible images for face recognition. Lecture Notes in Computer Science, Springer, Berlin.

6. Park, J. S. , Oh, Y. H. , Ahn, S. C. and Lee, S. W. **2005**. Glasses Removal from Facial Image Using Recursive Error Compensation, *IEEE Transaction on PAMI*, Vol. 27.

7. Bebis, G., Gyaourova, A., Singh, S. and Pavlidis, I. **2006**. Face Recognition by Fusing Thermal Infrared and Visible Imagery, Image and Vision Computing, 24(7), pp: 727-742.

8. Aran, A. , Munshi, S. , Beri, V. K. and Gupta, A. K. **2008**. Spectral Band Invariant Wavelet-Modified MACH filter , *Optics and Laser in Engineering*, 46, pp: 656-665.

9. Bhowmik, M., Kanti, B. D., Nasipuri, M. and Mahantapas . **2009**. Fusion of Daubechies Wavelet Coefficients for Human Face Recognition.

10. Gurjar, A. A. ,Swapna, M. P. and Kasturiwal , S. B. ,**2012**. Fusion Technique for Multi-focused Images using Stationary Wavelet Packet Transform, *International J. Scientific & Engineering Research* , 3(3).

11. Alwan, M.G. **2012**. Color Image Denoising Using Stationary Wavelet Transform and Adaptive Wiener Filte, *Al-Khwarizmi Engineering J.*, 8(1), PP: 18 -26.

12. Hossein, S. , Davood, S. and Hamid, D. **2008**. Face Detection using Gabor Wavelets and Neural Networks, World Academy of Science, Engineering and Technology.

13. Revathy, N. and Guhan, T. **2012**. Face recognition system using back propagation artificial neural networks, *International J. Advanced Engineering Technology*, IJAET/3(1).

14. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.