# A Study on the Impact of Wavelet Decomposition on Face Recognition Methods

M. M. Mohie El-Din[1], Neveen. I. Ghali[2], Ahmed. A. A. G[1] and H. A. El Shenbary [1]
[1] Department of Mathematics and Computer Science, Faculty of Science, Al-Azhar University, Cairo, Egypt
[2] Assoc. Prof Computer Science, Faculty of Science, Al-Azhar University, Cairo. Egypt

## ABSTRACT
Face recognition is the important field of pattern recognition. Discrete Wavelet Transform, (DWT), is known as a very powerful tool in the field of image processing, which extracts the feature vector to determine the performance of face recognition system. However, there are different decomposition levels of DWT. In this paper different decomposition levels of DWT integrated with a lot of good feature extraction methods are examined. Experiments on ORL face database showed that three levels of wavelet decomposition gives promising results and the hybrid method 2D-DWT_PCA_SVM gives high recognition rate and less time rather than other used methods.

## Keywords
Face recognition, 2D-DWT, SVM, PCA, FLDA, 2D-PCA

## 1. INTRODUCTION
Face recognition is a very powerful tool for a biometric authentication. It is the ability to identify person based on his facial characteristics. It has been an active research area in the last decades. There are several applications to face recognition in our life such as criminal investigation, security application, and user authentication. Several face recognition methods have been used in the last two decades. Turk, M. and Pentland, A. [1] introduced the Eigenfaces method called Principal Component Analysis, (PCA). Projecting the new image in the subspace spanned by the Eigenfaces method. PCA treats the image as one dimensional vector. PCA has become one of the most successful methods for data compression, redundancy removal, and feature extraction. Yang, J. et al [2] introduced another approach based on face representation and face recognition called Two Dimensional Principal Component Analysis, (2D-PCA). 2D-PCA treats an image as a two dimensional matrix not one dimensional vector as PCA does, which reduce time used to compute eigenvectors of the covariance matrix. Both PCA and 2D-PCA do not make full use of class information. Using between class scatter matrix and within class scatter matrix LDA [3] tried to find a set of projection vector to solve this problem. In addition, Support Vector Machine, (SVM) [4,15] are used in face recognition for classification. SVM is a learning machine algorithm that classifies the data by shaping a set of support vectors.

A very powerful tool used in the field of image analysis is called wavelet transform. The multi resolution decomposition provides a useful image representation for vision algorithms [5]. A lot of researchers integrated the DWT with other algorithms to improve recognition rate. Naresh Babu, N. T. et al, [6] Combined DWT with independent component analysis and improved recognition rate to 96% on MIT face database, 98% on FERRET database and Euclidean Distance was used as classifier in this method. Song , L. and Min, L. [7] integrate DWT with 2D-PCA algorithm for feature extraction and got recognition rate 92% on ORL database. Mohie El-din, M.M. et al, [8] proposed a hybrid method of DWT, 2D-PCA and LDA that enhanced recognition rate to 97.5% using ORL database. In [9] DWT and DCT was used to enhance recognition rate, which gave 96% recognition rate when applied on ORL Database. Zhang, C. et al, [10] used the wavelet packet analysis for image pre-processing, which achieves a good result in edge detection. Chou, Y. et al, [11] combined DWT with sub-pattern PCA which achieve 99.5% recognition rate on applying on ORL Database. Dawoud, N. N. and Samir, B. B, [12] used a multilevel decomposition to get the best wavelet function, results showed that the symelt wavelets are the optimum wavelets for the face classification with four levels when applied on Yale and ORL Database. Wang, W. et al, [13] combined DWT with SVM, results indicated that the proposed method can achieve recognition precision of 96.78% based on 96 persons in Ren-FEdb database. Satone, M. P, and Kharate, G. K, [14] used Euclidian distance measure, city block distance measure and PCA on DWT, results showed that recognition rate was 94.37% when applied on ORL database. In this paper we study the effect of multi level decomposition of the DWT on a lot of hybrid methods as PCA_SVM, 2D-PCA_FLDA_SVM, FLDA_SVM are studied.

The rest of this paper is organized as follows: Section 2 describes existing techniques PCA, LDA, SVM, 2D-PCA and DWT. Section 3 describes different hybrid methods. These methods are compared with each other by performing tests on well known ORL database in Section 4. Section 5 contains conclusion and future work.

## 2. FACE RECOGNITION TECHNIQUES
This section describes some of existing techniques for face recognition as DWT, PCA, 2D-PCA, LDA and SVM literature.

### 2.1 Discrete Wavelet Transform
DWT is considered to be a powerful tool in face recognition. It has been used in a wide range in image analysis and it can captures space-frequency information. Because the face image is considered as two dimensional signal, it is analyzed using two dimensional discrete wavelet transform, (2D-DWT). The decomposition process of the 2D-DWT is illustrated in Figure 1.

In each level, the input signal is filtered along rows and the resulted signal is filtered along column.

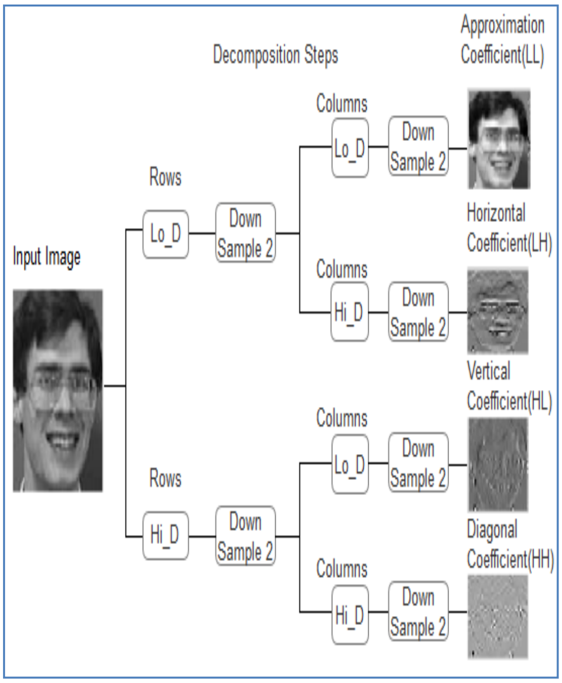


**Figure 1**

## 2.2 Principal Component Analysis

PCA is a statistical model used for face representation prior to its recognition. PCA used to reduce dimensionality of the data space to the smaller intrinsic dimensionality of the feature space, which are needed to describe the data economically. This reduction is realized by the linear transformation

$$Y = AX \quad (1)$$

Where X is the image, A is the transformation matrix and Y is the new image.

Assume we have a set of N images X1, X2 ……XN each image is a 2-dimentional matrix of size m×n. Each image is converted to one dimensional column vector of size mn.

$$X_i = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ . \\ x_{mn} \end{pmatrix} \quad (2)$$

Then we have the images set as follows.

$$X = [X_1, X_2, X_3 X_N] \quad (3)$$

Compute the mean image Xm

$$X_m = \frac{1}{N}\sum_{i=1}^{N} X_i \quad (4)$$

and the covariance matrix is given by the formula

$$C = \frac{1}{N}\sum_{i=1}^{N}(X_i - X_m)(X_i - X_m)^T \quad (5)$$

Then the eigenvalues and eigenvectors of the covariance matrix C are computed and most relevant features components are selected from them. PCA transformation matrix A can be constructed from the eigenvectors corresponding to the k largest eigenvalues of the covariance matrix C.

## 2.3 Two Dimensional PCA

Let Xi be an m×n image, Ak be an n dimensional unitary column, Projecting X on to Ak as shown

$Y_k = X_i A_k, k = 1, 2, 3 \ldots d$ (6)
Yk is a m dimensional vector which is known as feature vector of the image Xi.

Projecting Xi onto a matrix A= (A1, A2 …….Ad), to get a family of projected feature vectors Y= (Y1, Y2 …… Yd) using the training images.

Assume that there are M training samples each of size m×n and the ith training image Xi (i= 1, 2, 3 …. M).

The average image of all samples is $\overline{X}$.

The projection matrix Aopt is obtained by computing the eigenvectors A1, A2 …….Ad of the corresponding largest d eigenvalues of the matrix G.

$$G = \frac{1}{M}\sum_{j=1}^{M}(X_j - \overline{X})^T (X_j - \overline{X}) \quad (7)$$

Assume we have,

$$Y_K = X\,A_k, \quad k = 1, 2, 3 \ldots d \quad (8)$$

Then Y is an m×d matrix Y= (Y1, Y2 …… Yd), which is called the feature matrix or the feature image of the image sample Xi.

## 2.4 Fisher's Linear Discriminant Analysis

**Fisher's Linear Discriminant Analysis, (FLDA), [17], is a** good example for class specific method, since the training set is labeled, it make sense to use this class information to build a more reliable method to reduce the dimension of the feature space.

FLDA maximizes the ratio of the between class scatter matrix and within class scatter matrix and Looks for a linear subspace W(c-1 component) in which the projection of the different classes are best separated.

Let {Yk| k=1, 2 …. N} be a set of N samples in d dimensional space. Let li be class label of Yi,

li ϵ {1,2 … c}, c is classes number, denote class i sample number by Ni. Then the between class scatter matrix is defined as

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (9)$$

Within class scatter matrix is defined as

$$S_W = \sum_{i=1}^{c}\sum_{k=1}^{N_i}(Y_k - \mu_i)(Y_k - \mu_i)^T. \quad (10)$$

Total class scatter matrix is

$$S_T = \sum_{i=1}^{N}(Y_k - \mu)(Y_k - \mu)^T. \quad (11)$$

Where the mean of the $i_{th}$ class is

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_j \quad , \qquad (12)$$

μ is the global mean of all samples.

$$\mu = \frac{1}{N}\sum_{k=1}^{N} Y_k \qquad (13)$$

FLDA look for $W_{opt}$ which is the optimal projection matrix for maximizing the discriminant criteria

$$W_{opt} = \frac{|W^T S_B W|}{|W^T S_W W|} \quad (14)$$

If Sw is nonsingular, then Wopt is the matrix with the orthonormal columns which maximizes ratio of the determinant of the between class scatter matrix to the determinant of the within class scatter matrix of the projected sample,

$$W_{opt} = \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|} \quad (15)$$

$$W_{opt} = (W_1, W_2, W_3, \ldots, W_m) \qquad (16)$$

such that Wi, i=1, 2 ….m is the set of the generalized eigenvectors of the between class and within class scatter matrices corresponding to the m largest eigenvalues λi, i=(1, 2…. m).

$$S_B W_i = \lambda_i S_W W_i \; , i = (1,2,3 \ldots \; m) \quad (17)$$

The upper pound of m is c-1 where c is the number of classes in the training sample.

If Sw is singular, PCA is implemented to reduce dimensionality of the training images to solve this problem, and then LDA is implemented to reduce dimensionality to c-1.

Assume that

$$W_{Pca} = \arg\max_W |W^T S_T W| \qquad (18)$$

be the PCA transformation matrix, $W_{fld}$

$$W_{flda} = \arg\max_W \frac{|W^T W_{Pca}^T S_B W_{Pca} W|}{|W^T W_{Pca}^T S_W W_{Pca} W|} \quad (19)$$

Be the FLDA transform matrix, then the optimal projection matrix $W_{opt}$ is given by

$$W_{opt} = W_{Pca} \; W_{flda} \quad (20)$$

## 2.5 Support Vector Machine

SVM is a learning machine algorithm that classifies the data by shaping a set of support vectors. It is a statistical classification method proposed by Cortes and Vapnik [4, 15]. SVM creates a hyperplane between two sets of data for classification[16]. SVM becomes popular because of its success in handwritten digit recognition. SVM finds the maximum separating hyperplane, the one with the maximum distance between the nearest training tuples.

Assume we have two classes and there exist two hyperplanes as shown in Figure 2, Figure 3. Figure 3 with the larger margin should have greater generalization accuracy than Figure 2.

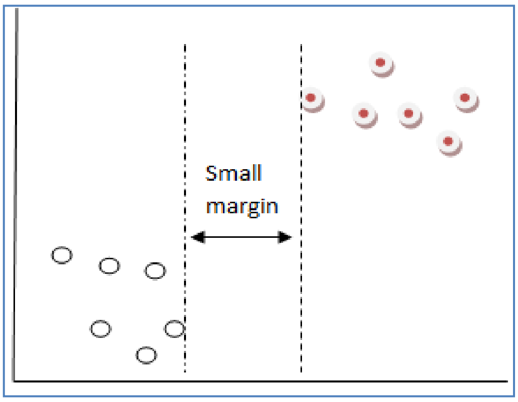


**Figure 2: Small margin**

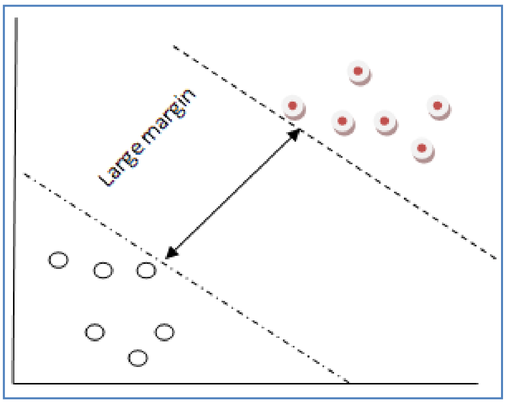


**Figure 3: Large margin**

## 3. EFFECT OF MULTILEVEL OF DWT ON THE PROPOSED METHODS

A multilevel of the discrete wavelet transform was applied on the given face image, the given image is decomposed into 4 subband images: approximation, horizontal, vertical, diagonal coefficients corresponding to the 4 subbands LL, LH, HL, and HH. Three Examples of decomposed images into 1level, 2 levels, and 3 levels are shown in Figure 4, Figure 5 and Figure 6 respectively.

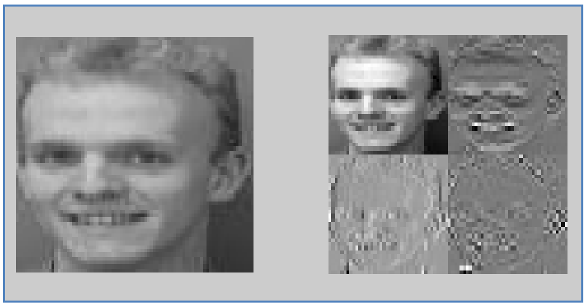

**Figure 4: 1 level 2D-DWT**

**Figure 5: 2 level 2D-DWT**

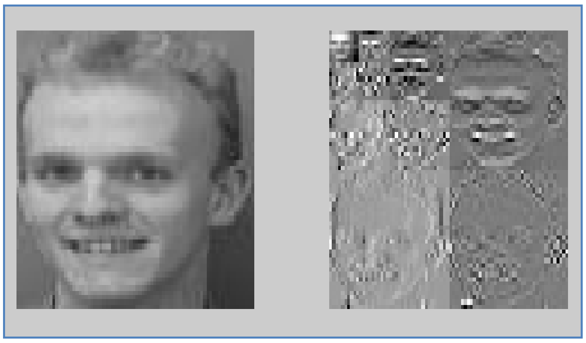

**Figure 6: 3 level 2D-DWT**

Taking in consideration that the LL subband of the 1st level was used as input to the 2nd level decomposition, because LL subband has most of energy of the original image.

Block diagrams of the methods which used in the experiments are shown in Figure 7, Figure 8 and Figure 9. After using DWT different methods were used for feature extraction like PCA, 2D-PCA, FlDA and finally SVM was used for classification.

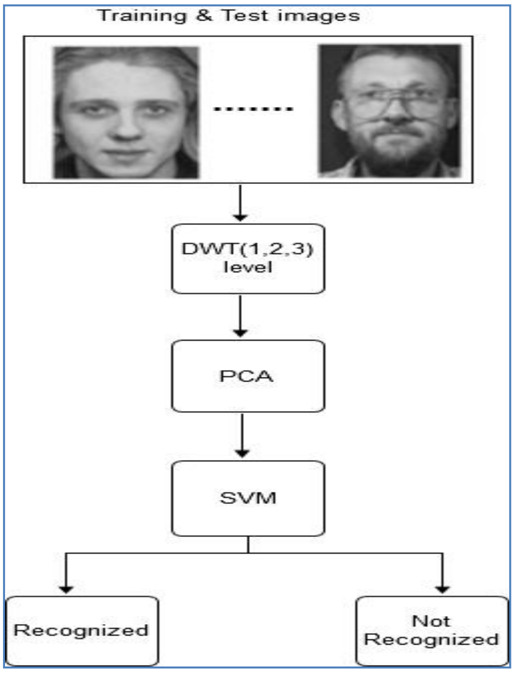


**Figure 7: DWT_PCA_SVM**

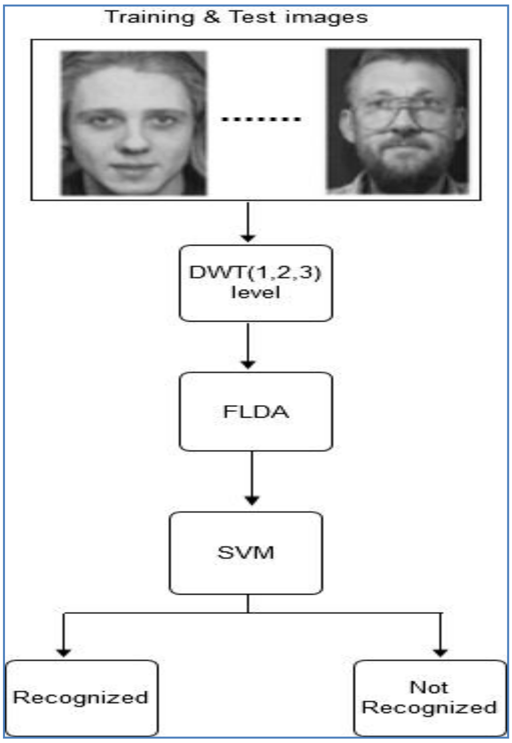


**Figure 8: DWT_FLDA_SVM**

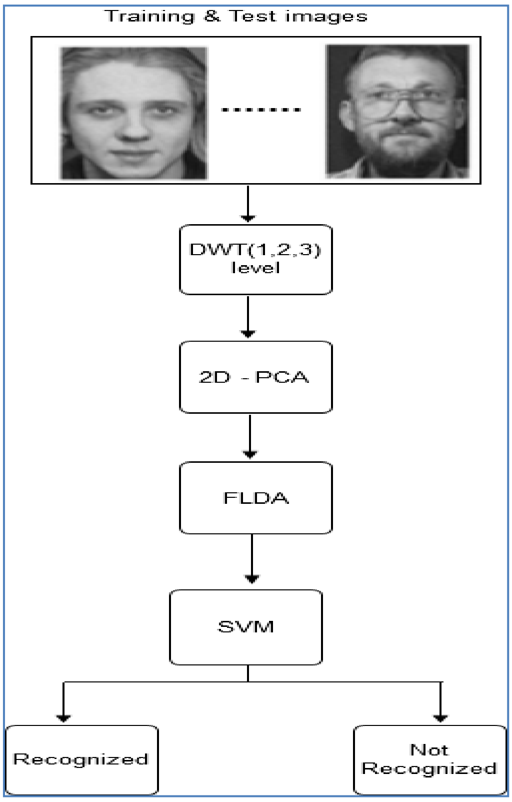


**Figure 9: DWT_2D-PCA_FLDA_SVM**

## 4. EXPERIMENTAL RESULTS

Experiments were carried out on a well known ORL face database [18]. Sample of the ORL images is shown in Figure 10. This database contains 40 classes; each class contains 10 images for each person. The images are taken in different periods, different facial expression and facial details such as smiling, eye open, eye closed, wearing glasses or not, different illumination, total of 400 images. In the experiment some images are used to establish training set and the rest is leaved for the testing stage.

**Figure 10: Sample of ORL Face Database**

Experiments were performed on the ORL database, recognition rate and total time (training + test) complexity are computed for the different used methods and different levels of DWT decomposition were used. Total time complexity and recognition rate are shown in Figure 11, Figure 12 and Figure 13. Figure 14 , Figure 15 comparing the average recognition rate and the average total time for all of the methods, respectively. Figure 11 show that using DWT_ PCA_SVM, level 3 gave high recognition rate rather than level 2 and level 1. Using PCA_SVM only, it is noted that recognition rate is very low (69.61%) and total

time is very high (182 seconds). In Figure 12, using 2 levels of DWT, the recognition rate in case of small sample size is given by (75.83%). It can be noted that 3 levels of DWT gave high recognition rate compared with other levels. Figure 13 shows that 3 levels of DWT on 2D-PCA_FLDA_SVM give high recognition rate rather than 2 levels of DWT in case of high sample size. Note that 3 levels of DWT also gave high recognition rate (79.9%) rather than level 1 (77.25%) and level 2 (66.97%). Figure 14 shows that 3 levels of DWT on PCA_SVM gives high recognition rate (88%) compared with all of the used methods.

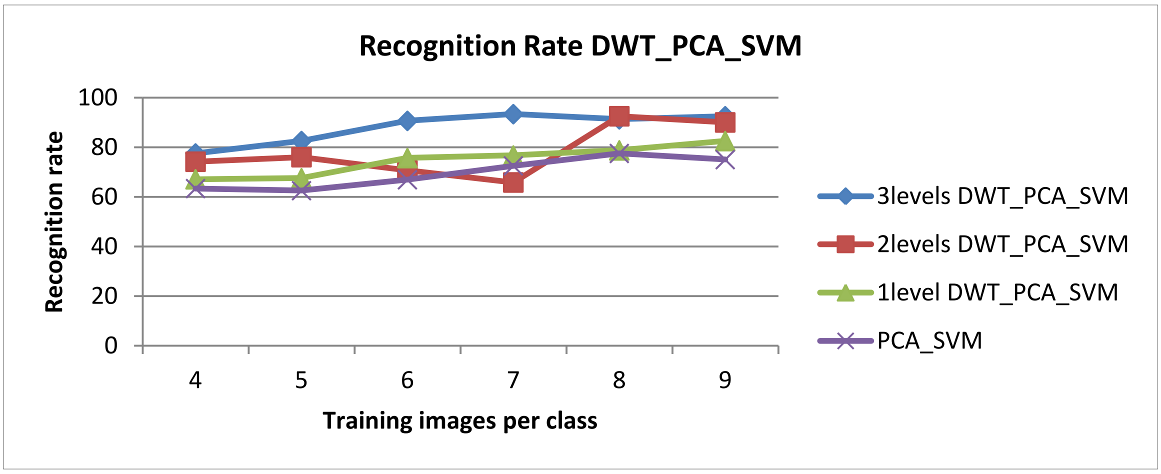


**Figure 11: DWT_PCA_SVM Recognition Rate**

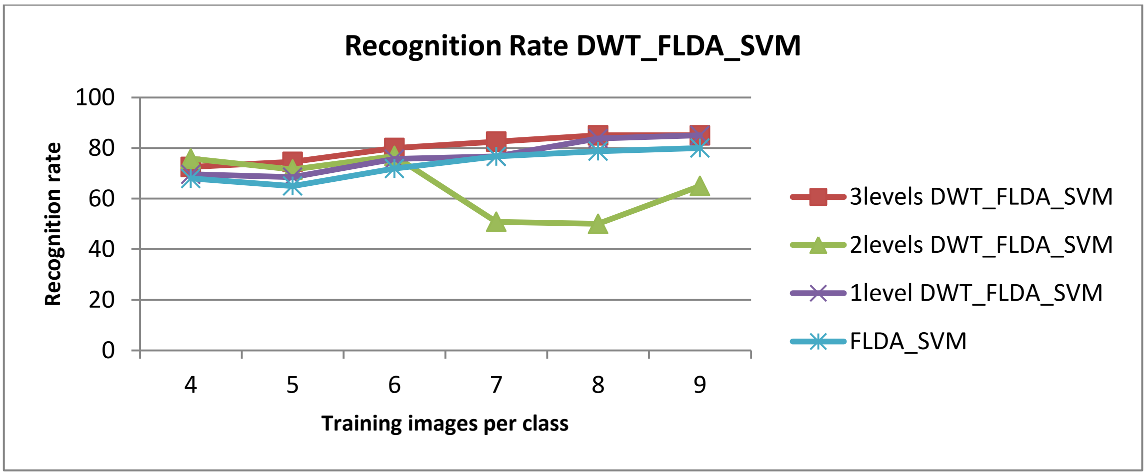


**Figure 12: DWT_FLDA_SVM Recognition Rate**

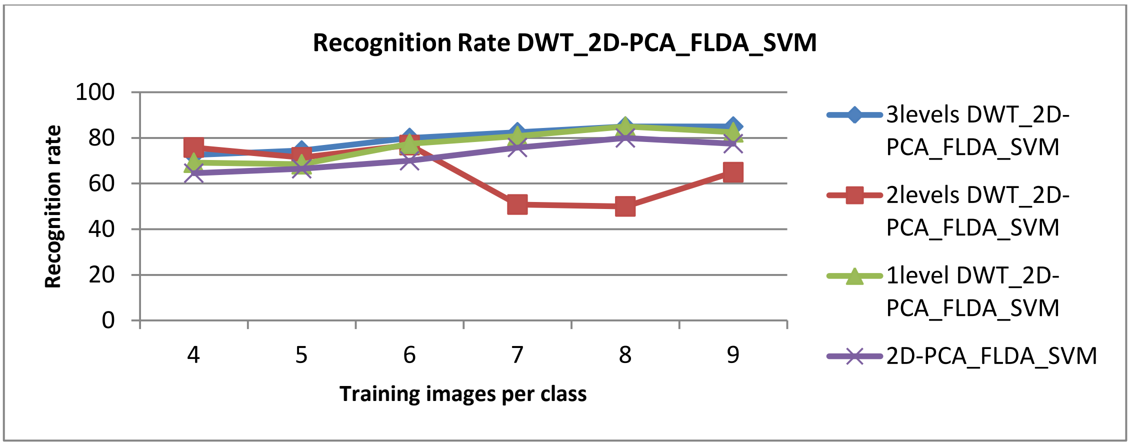


**Figure 13: DWT_2D-PCA_FLDA_SVM Recognition Rate**

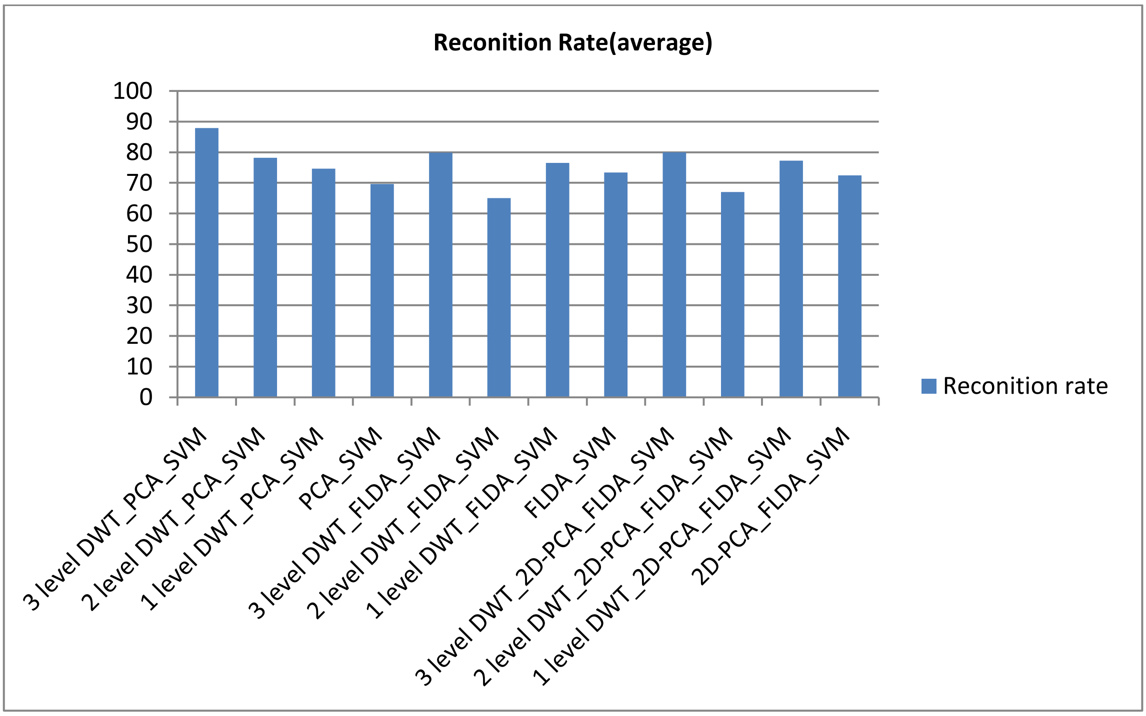


**Figure 14: average recognition rate for all methods**

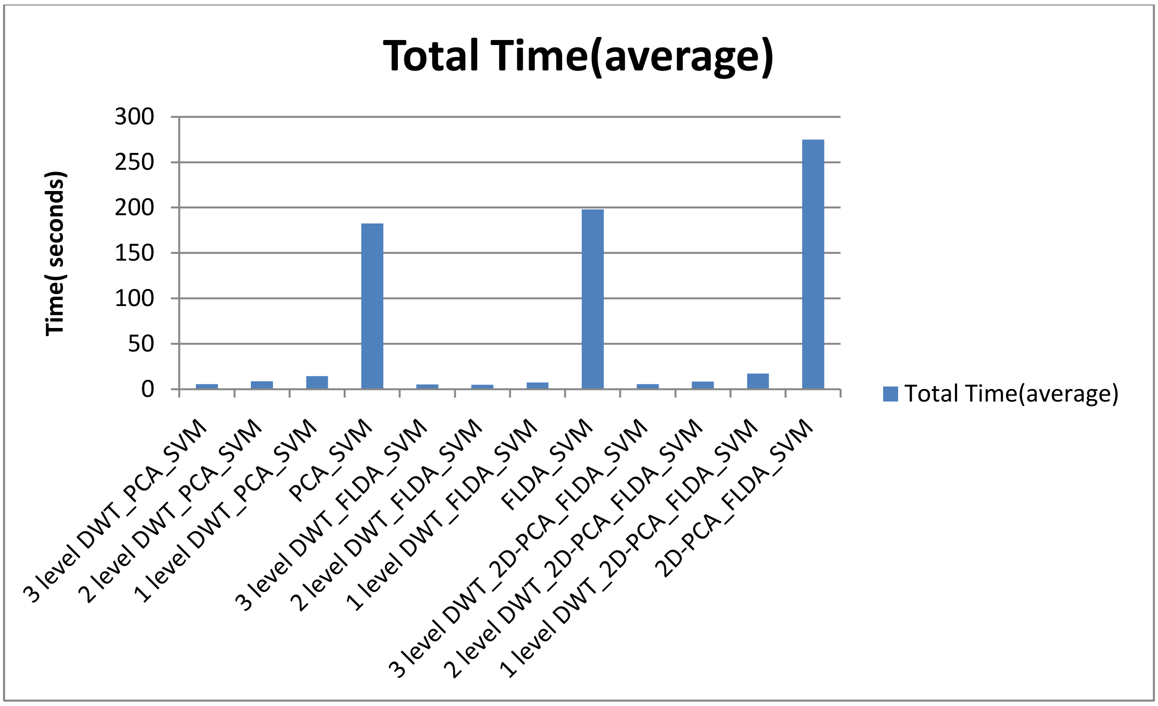


**Figure 15: Comparing average total time for all methods**

## 5. CONCLUSION

In this paper different methods were examined using different DWT levels. Results show that when PCA_SVM or FLDA_SVM or 2D-PCA_FLDA_SVM was used without using any level of wavelet decomposition, the total time was very high (182.5 seconds) and the recognition rate is low (69.61%).

On applying DWT, it is noted that the total time is diminished with 97%, in addition to an increase in the recognition rate with 26%. It can be depicted that 3 levels of DWT is the best number of decomposition levels and 3 levels DWT_ PCA_ SVM is the best method in all of the used techniques when applied on ORL Database.

Future Work in connection to this paper will test these methods using different databases and try to enhance recognition rate in case of small training sample size.

## 6. REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for Recognition." Journal of Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991.

[2] J. Yang, D. Zhang, A. F. Frangi, and J.-yu Yang, "Two-dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131-137, Jan. 2004.

[3] H. Yu, J. Yang, and A. As, "A Direct LDA Algorithm for High-dimensional Data with Application to Face Recognition," Journal of Pattern Recognition, vol. 34, pp. 2067-2070, 2001.

[4] S. R. Gunn, "Support Vector Machines for Classification and Regression," University of Southampton, pp. 168-194, May. 1998.

[5] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation," University of Pennsylvania, Department of Computer & Information Science,pp.1-30, 1987.

[6] Naresh. Babu. N. T, Annis. Fathima. A, and V. Vaidehi3, "An Efficient Face Recognition System Using DWT-ICA Features," IEEE International Conference on Digital Image Computing: Techniques and Applications, PP. 146-151, 2011.

[7] L. Song and L. Min, "Face Recognition Based on 2DPCA and DWT," IEEE Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, no. 4, pp. 1459-1462, July. 2011.

[8] M. M. Mohie El-din, M. Y. El Nahas, and H. A. El Shenbary, "Hybrid Framework for Robust Multimodal Face Recognition," International Journal of Computer Science Issues (IJCSI), vol. 10, no. 2, pp. 471-476, 2013.

[9] M. Wang, H. Jiang, and Y. Li, "Face Recognition based on DWT/DCT and SVM," IEEE International Conference on Computer Application and System Modeling (ICCASM), vol. 3, pp. 507-510, 2010.

[10] C. Zhang , Y. Hu , T. Zhang , H. An, W. Xu, "The Application of Wavelet in Face Image Pre-Processing," *IEEE International Conference on* Bioinformatics and Biomedical Engineering (iCBBE),vol.4, pp. 1-4, 2010.

[11] Y. Chou, S. Huang, S. Wu, and J. Yang, "DWT and Sub-pattern PCA for Face Recognition Based on Fuzzy Data Fusion," *IEEE International Conference on Intelligent Computation and Bio-Medical Instrumentation (ICBMI)*, pp. 296-299, 2011.

[12] N. N. Dawoud and B. B. Samir, "Best Wavelet Function for Face Recognition Using Multi-Level Decomposition," IEEE International Conference on Research and Innovation in Information Systems (ICRIIS), pp.1-6, 2011.

[13] W. Wang, X. Sun, S. Karungaru, and K. Terada, "Face Recognition Algorithm Using Wavelet Decomposition and Support Vector Machines," *IEEE International Symposium on Optomechatronic Technologies (ISOT)*, pp. 1-6, Oct. 2012.

[14] M. P. Satone and G. K. Kharate, "Face Recognition Based on PCA on Wavelet Subband," *2012 IEEE Students Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1-4, Mar. 2012.

[15] C. C. and V.Vapnik, "Support Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[16] F. Bellakhdhar and K. Loukil, "Face Recognition Approach Using Gabor Wavelets, PCA and SVM," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 2, pp. 201-207, 2013.

[17] R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis".New York: Wiley, pp. 114-118, 1973.

[18] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification" 2nd IEEE Workshop on Applications of Computer Vision December 1994, Sarasota (Florida). "http://www.face-rec.org/databases".