Volume 5, Issue 3, March 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Hybridization between Rough Set and FLDA for Face Recognition

M. M. Mohie El-Din[1], Neveen. I. Ghali[2], Ahmed. A. A. G[1], 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 an important _eld of pattern recognition. Rough set theory used to extract the face from an image, by computing ob- ject lower approximation array according to gray level value. Fisher's Linear Discriminant Analysis,(FLDA), was used for feature extraction and dimensionality reduction. In the proposed methods Object sim- ilarity ratio and Euclidean distance were used for classi_cation. Ex- periments on ORL face database showed that the proposed methods gave promising recognition rate and acceptable test time complexity.

Keywords: Face Recognition,Rough Set,PCA ,LDA ,Euclidean Distance.

I. INTRODUCTION

Face recognition considered as 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 such as user authentica- tion, security application,and criminal investigation. There exist several face recognition methods. Turk, M. and Pentland, A. [10] introduced the Eigen- faces method called Principal Component Analysis, (PCA). Projecting the new image in the subspace spanned by the Eigenfaces method. PCA has become one of the most successful methods for data compression, redun- dancy removal, and feature extraction. Yang, J. et al [11] introduced an- other 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 [12] tried to _nd a set of projection vector to solve this problem. A very powerful tool used in the _eld of image analysis is called wavelet transform. The multiresolution decomposition provides a useful image representation for vision algorithms [5]. A lot of researchers integrated the DWT with other algo- rithms to improve recognition rate. Song , L. and Min, L. [9] integrate DWT with 2D-PCA algorithm for feature extraction and got recognition rate 92% on ORL database. Mohie El-din, M.M. et al, [3, 2] proposed a hybrid method of DWT, 2D-PCA and LDA that enhanced recognition rate to 97.5% using ORL database. In addition Rough Set Theory was used in [7] for image retrieval. B. Debotosh et al, [1] integrated rough set to reduce the feature vector generated by the PCA algorithm. Authors in [4] presented a hybrid system consisting of PCA, Rough Set and Neural Network for dimensional- ity reduction and classi_cation in human face recognition. This method was tested on a well known ORL face database and gave 93% average recognition rate.
In this paper three methods are introduced for face recognition problem. This methods integrated Rough Set with FLDA, Object Similarity Ratio and Euclidean Distance, which gave high recognition rate and minimize time con- sumed in testing stage. The rest of this paper is organized as follows: Section 2 describes existing techniques PCA, LDA, Rough Set, Euclidean distance and Object similarity ratio. Section 3 describes the proposed methods. The proposed methods are evaluated in Section 4 by performing tests on a well known ORL face database. Section 5 contains conclusion and future work.

II. FACE RECOGNITION METHODS

This section describes some of existing techniques for face recognition prob- Lem.

2.1 Rough Set preliminaries

Rough set theory depends on the idea that every object is associated with a piece of knowledge indicating relative membership. Knowledge is repre- sented by means of a table, so-called an information system, where rows and columns respectively denote objects and attributes. Consider an information system, S, is given as a pair S=(U,A) where U is a non-empty _nite set of objects, as the universe, and A is a non-empty _nite set of attributes. Let B _ AandX _ U. We can approximate the set X using only the informationcontained in B by constructing the lower and upper approximations of X. If $X \subseteq U$, the sets $\{x \in U : [x]_B \subseteq X\}$ and $\{x \in U : [x]_B \bigcap X \neq \Phi\}$, where [x]B denotes the granule in other words equivalence class of the object x 2 U relative to the equivalence relation IB, are called the B-lower and B- upper approximations of X in U. They are denoted by BX and BX, respec- tively. Figure 1 show the representation of a set X with the upper and lower approximation. Set of dark gray granules represent lower approximation, set of dark gray and light gray granules represent upper approximation.

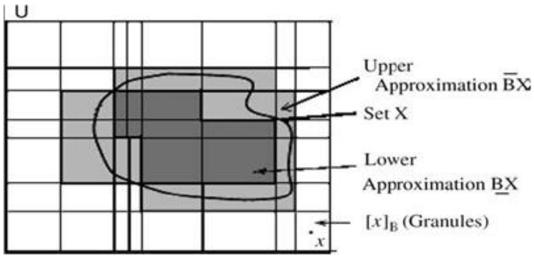

Figure 1: Rough Representation of a Set X

Assume the universe U be an image consisting of a collection of pixels, partition U into a collection of non overlapping windows, each window can be considered as a granule G. Granulation involves decomposition of whole into parts [8]. Let us consider an object-background separation (a two class) problem of an mn gray scaled image with gray level intervals (0,1, ,T,T+1, ...,L-1) with L gray levels; Let B and O represent two properties that char- acterize background and object regions, respectively Object and background can be viewed as two sets with their rough representation with respect to gray level T. Lower approximations of the object OT :

$\underline{O}_T = \{\bigcup_i G_i | P_j > T, \forall j = 1, ..., mn$, and P_j is a pixel belonging to $G_i\}$.

Upper approximation of the object $\overline{O}_T$:

$\overline{O}_T = \{\bigcup_i G_i, \exists j, j = 1, ..., mn, s.t. P_j > T$, where P_j is a pixel in $G_i\}$.

Lower approximations of the background $\underline{B}_T$:

$\underline{B}_T = \{\bigcup_i G_i | P_j \leq T, \forall j = 1, ..., mn$, and P_j is a pixel belonging to $G_i\}$.

Upper approximations of the background $\overline{B}_T$:

$\overline{B}_T = \{\bigcup_i G_i, \exists j, j = 1, ..., mn, s.t. P_j \leq T$, where P_j is a pixel in $G_i\}$.

Thus the roughness of the object at gray level value T is given by:

$$R_{O_T} = 1 - \frac{|\underline{O}_T|}{|\overline{O}_T|} = \frac{|\overline{O}_T| - |\underline{O}_T|}{|\overline{O}_T|}. \quad (1)$$

Also, the roughness of the background at gray level value T is given by:

$$R_{B_T} = 1 - \frac{|\underline{B}_T|}{|\overline{B}_T|} = \frac{|\overline{B}_T| - |\underline{B}_T|}{|\overline{B}_T|}. \quad (2)$$

Where $|\underline{O}_T|, |\overline{O}_T|, |\underline{B}_T| and |\overline{B}_T|$ denotes the cardinality of the lower and upper approximations for object and background respectively.

The mean roughness at a certain threshold T is defined as

$$RM_T = \frac{R_{O_T} + R_{B_T}}{2}. \quad (3)$$

We select T which minimize the mean roughness[8]. Figure 2 show a flow chart of how to select the optimal threshold T.

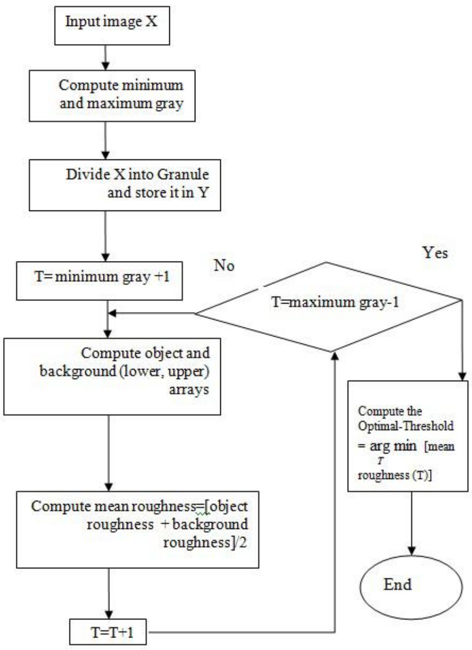

Figure 2: Flow chart of T selection

2.2 Principal Component Analysis

The principal component analysis (PCA) is a statistical model used for face recognition, one of the most successful techniques that have been used in image processing. The purpose of PCA is 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.

$$Y = AX. \tag{4}$$

Where Y is the feature matrix, A is the transformation matrix and X is the original image. PCA can give us data compression, prediction, redundancy removal, feature extraction. The scope of using PCA for face recognition is to express the large one dimensional vector of pixel constructed from the two dimensional facial image into the compact principal components of the feature space. This is called eigenspace projection. Eigenspace can be calcu-lated by calculating the eigenvectors of the covariance matrix derived from dataset of images. In 1991, Eigenfaces method was _rstly introduced by M. Turk and A. Pentland [10]. Assume we have a set of N images X1;X2,…,XN. Originally, each image is a 2-dimentional matrix of size m by n. PCA converts each image to 1-
dimentional column vector of size m _ n as follows.

$$X_i = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_{mn} \end{pmatrix} \tag{5}$$

The images set is

$$X = [X_1, X_2, ...X_N] \tag{6}$$

then compute the mean image X_m as follows

$$X_m = \frac{1}{N}\sum_{i=1}^{N} X_i. \tag{7}$$

and the covariance matrix is given by the formula

$$C = \frac{1}{N}\sum_{i=1}^{N}(X_i - X_m)(X_i - X_m)^T. \tag{8}$$

Let $W_i = (X_i - X_m)$, be the centered image. Now we want to find the eigenvalues λ_i and eigenvectors e_i of the covariance matrix C.

$$C = WW^T. \tag{9}$$

Where W is a matrix composed of the column vector Wi placed side by side. The size of C is mn _ mn, so image of size 100 _ 100 will create a covariance matrix of size 10000 _ 10000 which will not be practical to solve for the eigenvectors of C directly. A theorem in linear algebra states that the vectors ei and scalars _i can be computed by solving for the eigenvectors and eigenvalues of the N _ N matrix WTW [10]. Let _i , di be the eigenvectors and eigenvalues of WTW, respectively.

That is mean that

$$W^T W d_i = \mu_i d_i \tag{10}$$

By multiplying both sides by W(from left)

$$(WW^T)Wd_i = \mu_i(Wd_i) \tag{11}$$

Which mean that the _rst N □ 1 eigenvectors ei and eigenvalues _i of the covariance matrix C = WWT are given by Wdi and _i , respectively. Wdi needs to be normalized in order to be equal to ei. Since we have a _nite number of image vectors, N and the rank of the covariance matrix can not exceed N □1 (because we subtract the mean vector m). PCA transformation matrix A can be constructed from the eigenvectors corresponding to the k largest eigenvalues of the covariance matrix C.

2.3 Fisher's Linear Discriminant Analysis

PCA do not make full use of class information. Using **between class scat- ter matrix**(S_B) and **within class scatter matrix**(S_W), Fisher's Linear Discriminant Analysis, (FLDA), [12] tried to _nd a set of projection vector to solve this problem. FLDA is a good example for class speci_c 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 SB matrix and SW matrix and Looks for a linear subspace W(c-1 component) in which the projection of the di_erent classes are best separated. Let fYkjk = 1; 2:::Ngbe a set of N samples in d dimensional space. Let li be class label of Yi, li 2 f1; 2:::cg, c is classes number, denote class i sample number by Ni. Then the between class scatter matrix is de_ned as

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \qquad (12)$$

Within class scatter matrix is defined as

$$S_W = \sum_{i=1}^{C} \sum_{k=1}^{N_i} (Y_k - \mu_i)(Y_k - \mu_i)^T \qquad (13)$$

Where the mean of the i_{th} class is

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_j \qquad (14)$$

Total class scatter matrix is

$$S_T = \sum_{i=1}^{N} (Y_i - \mu)(Y_i - \mu)^T \qquad (15)$$

where μ is the global mean of all samples.

$$\mu = \frac{1}{N} \sum_{k=1}^{N} Y_k \qquad (16)$$

FLDA look for Wopt which is the optimal projection matrix for maximizing the discriminant criteria

$$W_{opt} = \frac{|W^T S_B W|}{|W^T S_W W|} \qquad (17)$$

If SW is nonsingular, then Wopt is the matrix with the orthonormal columns which maximizes ratio of the determinant of the SB matrix to the determi- nant of the SW matrix of the projected sample,

$$W_{opt} = arg \max_{W} \frac{|W^T S_B W|}{|W^T S_W W|}. \qquad (18)$$

$$W_{opt} = (W_1, W_2, W_3, ...W_m). \qquad (19)$$

such that Wi, i = 1; 2…m is the set of the generalized eigenvectors of the SB and SW matrices corresponding to the m largest eigenvalues _i; i = (1; 2:::m).

$$S_B W_i = \lambda_i S_W W_i, i = (1, 2, 3...m). \qquad (20)$$

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 FLDA is implemented to reduce dimensionality to c-1. Assume that

$$W_{Pca} = arg \max_{W} |W^T S_T W| \qquad (21)$$

be the PCA transformation matrix,

$$W_{flda} = arg \max_{W} \frac{|W^T W_{Pca}^T S_B W_{Pca} W|}{|W^T W_{Pca}^T S_W W_{Pca} W|} \qquad (22)$$

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

$$W_{opt} = W_{Pca} W_{flda} \qquad (23)$$

2.4 Euclidean Distance

Euclidean distance is used as the classier to identify the image in the train- ing set to which the test image belong. Image classication is performed by comparing the feature vectors (weight matrix) of the images in the training set with the feature vector (weight matrix) of the test image using Euclidean distance,

$$\varepsilon_i = \|\Omega_T - \Omega_i\| \qquad (24)$$

where i is a vector describing the ith face image in the training set.

2.5 Object Similarity Ratio

In order to match between two objects in di_erent images, we compute the object lower approximation array for each images. Then the object similarity ratio[7]is computed as follows:

$$ObjectSimilarityRatio = \frac{N}{T}. \qquad (25)$$

Where N denotes the number of pixels, which are the same in two object lower approximation arrays(Test image, Training image) and T is the size of the object lower approximation array. object similarity ratio equal 1 if the two images are the same, otherwise between [0; 1]. The recognition system retrieve image which have the large object similarity ratio.

III. PROPOSED METHOD

In the proposed methods, we compute the object lower approximation array for all images in the database. Images is divided to two sets (training and testing) . FLDA is used to extract the feature from the object lower approx- imation arrays for training images. Minimum Euclidean Distance is used to compute the distance between the test image and images in the training set. Figure 3, Figure 4 and Figure 5 show a block diagram of the proposed methods.

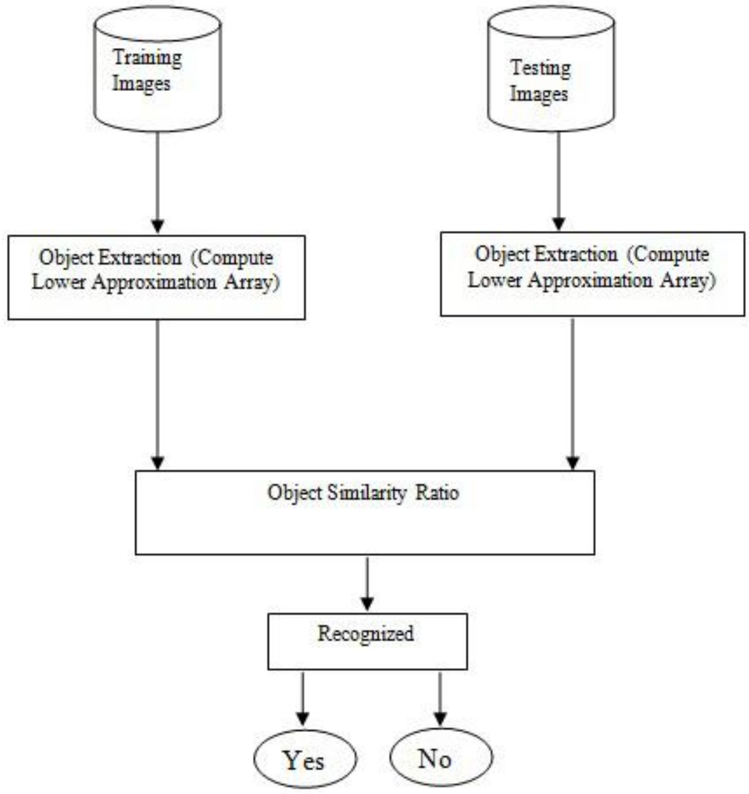

Figure 3: Rough Set and Object Similarity Ratio Method

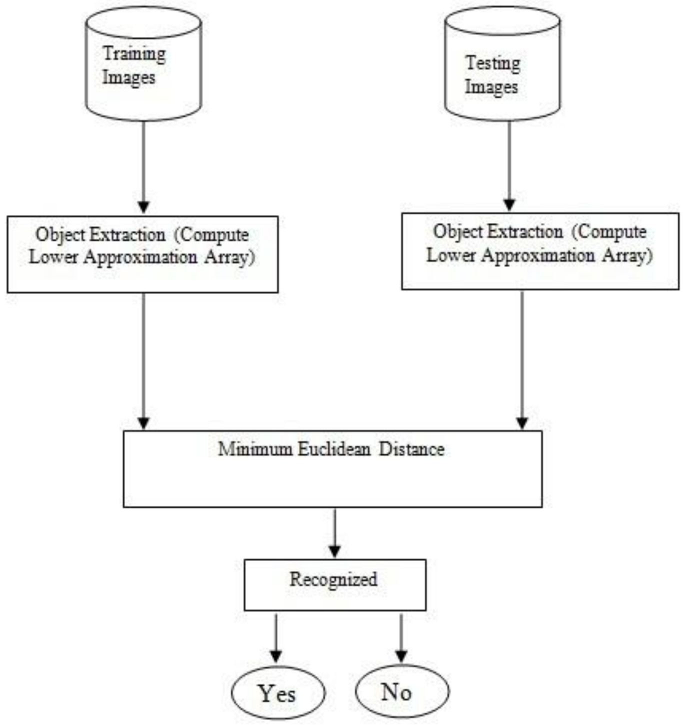

Figure 4: Rough Set and Euclidean Distance Method

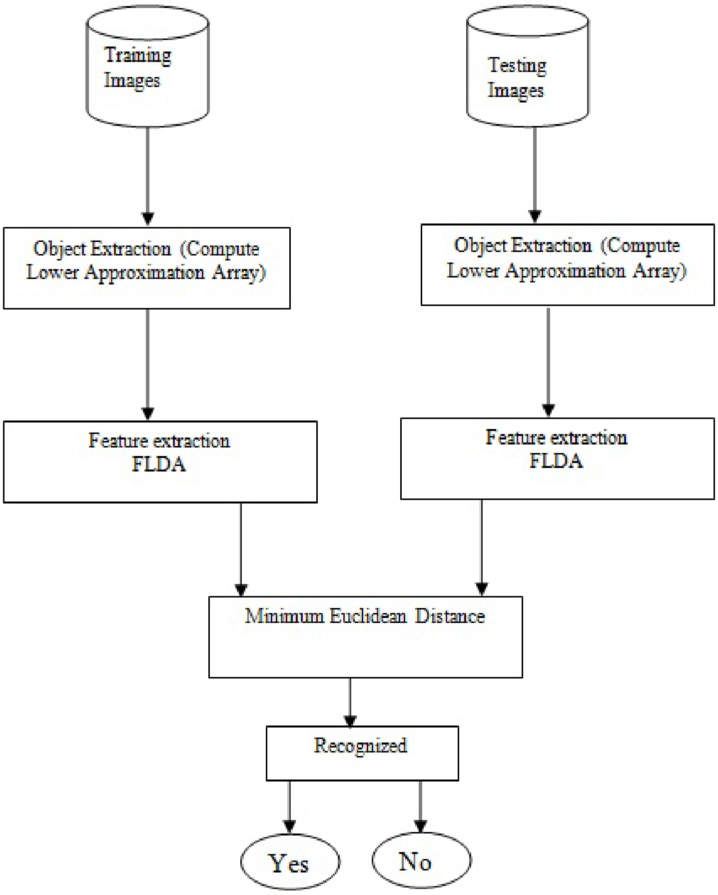

Figure 5: Rough Set and FLDA Method

IV. EXPERIMENTAL RESULTS

4.1 Face Database

Experiments were carried out on a well known ORL face database [6]. Sample of the ORL images is shown in Figure 6. This database contains 40 classes; each class contains 10 images for each person. The images are taken in di_erent periods, di_erent facial expression and facial details such as smiling, eye open, eye closed, wearing glasses or not, di_erent 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 6: Sample of ORL Face Database

4.2 Results

In order to compute the object lower approximation array for all images in the given ORL face database(400 images), it takes 414 seconds. Then object similarity ratio and Euclidean distance were used to classify the test images. To reduce time, FLDA was used for dimensionality reduction. Table 1 show the recognition rate of the proposed methods. When Rough set integrated with FLDA, recognition rate between 90%, 97.5%, but when we use Rough set Object similarity ratio method recognition rate was between 90.5% and 100%. Table 2 show a comparison of testing time complexity for the di_erent methods. It is clear that integrating FLDA and rough set gave signi_cant improvement in time complexity.

Table 1: Comparison of Recognition Rates on ORL Database

no of training images	Rough Set _ Object Similarity Ratio	Rough Set _Euclidean Distance	Rough Set _ FLDA	PCA_Rough set_Neural network [4]
5	90.5	91	90	...
6	98	94.37	94.37	...
7	99.17	95.8	95	...
8	100	95	96.25	...
9	100	92.5	97.5	...
average	97.5	93.73	94.6	93

Table 2: Comparison of Testing Time Complexity(Seconds) on ORL Database

no of training images	Rough Set _ Object Similarity Ratio	Rough Set _Euclidean Distance	Rough Set _ FLDA
5	8.35	4.83	2.57
6	7.64	4.52	2.35
7	6.61	3.8	2
8	5	2.86	1.5
9	2.91	1.6	1.22

V. CONCLUSION

In this paper di_erent methods were examined. Results show that when Rough set Object similarity ratio method was used, it gave high recognition rate between 90.5% and 100%. Note that time complexity was very high in this method compared with the other methods between 2.91 and 8.35 seconds. When Euclidean distance was used with Rough set, recognition rate was between 91% and 95.8%, but in this method test time complexity was diminished with 41%. Rough set FLDA gave recognition rate between 90% and 97.5%. Also, test time complexity was diminished with 66% compared with Rough set Object similarity ratio method. Future Work in connection to this paper will test these methods using di_erent databases, try to enhance recognition rate in case of small training sample size and try to decrease time consumed in object lower approximation array computation.

REFERENCES

[1] B. Debotosh, B. D. Kumar, N. Mita, and M. Kundu, Reduction of feature vectors using rough set theory for human face recognition, International Conference on Cognitive System (ICCS) pp. 1-12 (2005).

[2] M. M. Mohie El-din, N. I. Ghali, A. G. Ahmed, and H. A. El Shenbary, A study on the impact of wavelet decomposition on face recognition meth- ods, International Journal of Computer Applications (IJCA) vol. 87, no. 3, pp. 14-21 (Feb. 2014).

[3] 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).

[4] Y. F. Hassan and N. Habeb, Hybrid system of pca, rough sets and neu- ral networks for dimensionality reduction and classi_cation in human face recognition, International Journal of Intelligent Information Pro- cessing(IJIIP) V. 3. no. 1. (mar. 2012).

[5] S. G. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, University of Pennsylvania, Department of Com- puter & Information Science pp.1-30 (1987).

[6] F. Samaria and A. Harter, Parameterisation of a stochastic model for human face identi_cation, Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL. http://www.face- rec.org/databases. (Dec. 1994).

[7] G. A. Sharawey, N. I. Ghali, and W. A. Ghoneim, Object-based image retrieval system based on rough set theory, International Journal of Com- puter Science and Network Security(IJCSNS) VOL.9 No.1, pp.160- 165 (Jan. 2009).

[8] G. A. Sharawey, M. A. Gomaa, N. I. Ghali, and W. A. Ghoneim, Object- oriented image segmentation in presence of uncertainty, Advances in Fuzzy sets and Systems Vol. 3(3), pp.367-377 (2008). [9] 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.).

[10] M. Turk and A. Pentland, Eigenfaces for recognition, Journal of Cogni- tive Neuroscience vol.3, no.1, pp. 71-86 (1991).

[11] J. Yang, D. Zhang, A. F. Frangi, and Jing yu Yang, Two-dimensional pca: A new approach to appearance-based face representation and recog- nition, IEEE Transactions on Pattern Analysis and Machine Intelligence VOL. 26, no. 1, pp. 131-137 (Jan. 2004).

[12] H. Yu and J. Yang, A direct lda algorithm for high-dimensional data with application to face recognition, Journal of Pattern Recognition vol. 34, pp. 2067-2070 (Jan. 2001).

