FACE RECOGNITION BASED ON IMAGE INTEGRATION AND DCT ANALYSIS

Mohammed Sahib Mahdi Altaei
Department of Computer Science, College of Science, Al-Nahrain University.
e-mail : Mohamedaltaei@yahoo.com.

Abstract
This paper aims at enabling computers to recognize faces without human intervention. This is accomplished by capturing five images for the face under test, the information of five images are concentrated in one integral image by using KL transforms. The wavelet transform was used to diminution the integrated image, DCT was applied on the first band of the wavelet transformed image to extract the recognition features. The recognition features assumed to be driven from the first six DCT coefficients. The recognition was established by searching a match between the features extracted from the test image and library of the face image models. The performance of the designed recognition system gave a promising verification percentage of about 83% in which the contribution weights of the adopted features were unequally.

Keywords: Image integration, KL transforms, wavelet, image integration, and face recognition.

1. Introduction
Recognition is the most interest topic in the visual surveillance applications of computer vision. Face recognition is generally stands for the problem of recognizing an individual among set of different others. Many successful systems have been developed to verify the identity in unconstrained environments which have become increasingly robust, and are able to track individuals for minutes through huge data. These systems typically can provide images of users at low resolution, possibly from single or multiple viewpoints [1].

In face recognition problems, the observation image must be compared to a list of known models. Typically the models are tested and obtained from sets of multiple viewpoint images, these images can be thought of as a set matching problem. The strategy of integrating information from multiple observations may either early or late integration approaches. Early integration approach consists of selecting the best observation from each set using a quality metric measure classification. The late integration based on the common statistical approach that takes the product of the likelihoods of each observation [2].

Many publications focus on literature that deals specifically with recognition from image sets with invariance to pose and illumination. Illumination invariance is perhaps the most significant challenge, since image differences due to the change of illumination may be larger than differences between individuals [3]. Several approaches have been proposed for face recognition based on 2D or 3D images. In general, face recognition technologies include two step approaches [4, 5]; off-line and on-line. An off-line enrollment procedure is established to build a unique template for each registered user. The procedure is based on the acquisition of a predefined set of face images (or video sequence) selected from the input image stream, and the template is built upon a set of features extracted from the image ensemble. An on-line recognition procedure where a set of images is acquired and processed to extract a given set of features. From these features a face description is built to be matched against the test image [6]. The proposed detection and recognition algorithms used the dynamics of the consecutive images of a face or local features of the face, and integrate information into the recognition framework [7, 8], which is the way of recognition in the present work.

In the real word performance of face recognition, the error rate has been reported to be as high as 10% to recognize very controlled images such as those used for passport photographs, while in less control environments the performance degraded even further [9, 10]. We believe that the main reason for this apparent discrepancy between the results reported in the literatures and those observed in the real word is the assumption that most face recognition methods are hard to satisfy in practice. Therefore, we go to decrease the error percentage as small as possible by employing accurate computations with account of the time slot needed to implement the recognition task.

2. Automatic Face Recognition

Automatic face recognition (AFR) occupies the most active research areas because of the vast improvement in the digital electronic fields and urgent need in practical applications. In the last two decades a number of different AFR algorithms have been developed, one of the best studied methods for low dimension representation of faces is the eigenfaces principal component analysis (PCA) approach [11]. This representation was used in [12] for face recognition. The idea behind the eigenfaces representation is to choose a dimensionality reduction linear transformation that maximizes the scatter of all projected samples. In [13], the PCA approach was extended to a nonlinear alternative using kernel functions.

Another subspace method that aims at representing the face without using class information is nonnegative matrix factorization (NMF) [14]. This algorithm like PCA, represents a face as a linear combination of bases. The bases of PCA are eigenfaces, some of which resemble distorted versions of the enter face, whereas the bases of NMF are localized features that correspond better to the intuitive notations of face parts [14]. An extension of NMF that gives even more localized bases by imposing additional locality constraints is the so called local nonnegative matrix linear discriminant analysis. It is one of the more well-studied methods that aim to find low dimensional representation using the information of how faces are separated to classes. Fisherfaces was then proposed to use linear discriminant analysis (LDA) in a reduced PCA space for facial image retrieval and recognition [15, 16]. Fisherface is two steps dimensionality reduction method. First, the feature dimensionality is reduced using the eigenfaces approach to make the recognition between class scatter matrix nonsingular. After that, the dimension of the new features is reduced further using Fisher's linear discriminant (FLD) optimization criterion to produce the final linear transformation.

Recently, direct LDA algorithms for discriminant feature extraction were proposed [17, 18, 19] in order to prevent the loss of discriminatory information that occur when a PCA step is applied prior to LDA [19]. Such algorithms are usually applied using direct diagonalization methods for finding the linear projections that optimize the discriminant criterion [17, 18, 19]. For solving nonlinear problems, the classic LDA has been generalized to its kernel version namely general discriminant analysis (GDA) [20] or kernel Fisher discriminant analysis (KFDA) [21].

In the present research, many approaches to recognition modeling and matching are possible within current framework. The developed method based on concentrating the energy contained in the image through dual processes of integrating the information and down sampling the target image to produce the best image, this depends on the variability in the observed data. The resulting image is considered to be compared with the available models. The matching criterion is driven from analyzing the DCT coefficients for the down sampled image, and thus outliers and natural variation can be handled properly.

3. Materials and Method

The materials were used in the present work are 30 sets of images, each set achieved either by extracting a sequence of frames from video file for view in an indoor office environment, or by taking a collection of still images at same pose and viewpoint. Each set consist of five independent images (256*256 resolution), in which the background is smooth and bright. Thus, there are 30 models for different sets of images stored in library are used to verify an input set of unknown image.

The adopted recognition method shown in Fig. (1) consists of two phases; training and recognition. The training phase responsible on collecting sample faces to be stored as comparable models in the library. Whereas, the recognition phase responsible on verifying the test face in comparison with the trained models. Both phases are composed of three stages; more details about each stages are explained in the following sections:

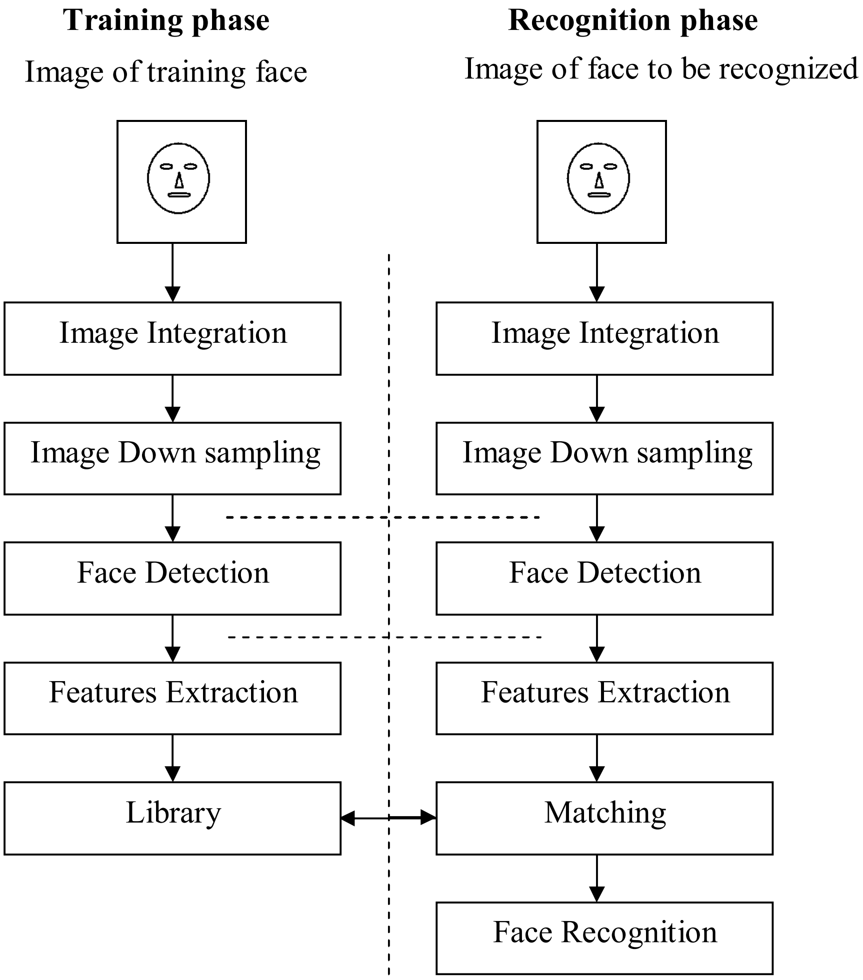
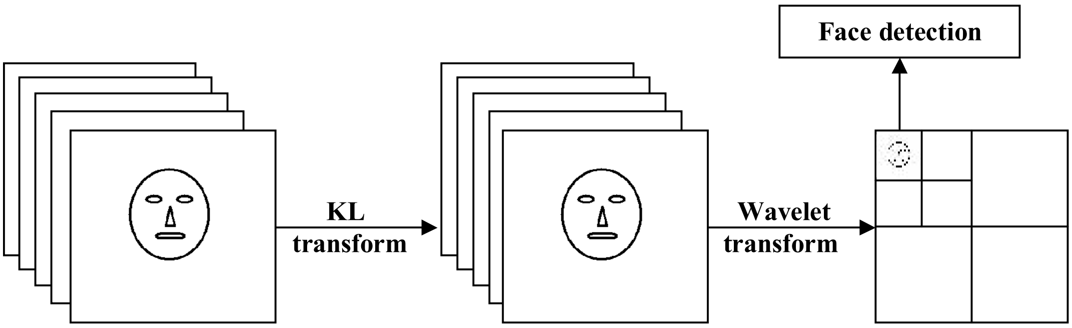


Fig. (1): Sequential steps of the proposed face recognition.

a- Integration and down sampling

This stage aims at preparing the acquired image, it contains a preprocessing steps includes integrating and then down sampling the image. The contrast of acquired image is a measure expresses the amount of chromaticity exist in the image. Since the down sampling may lead to decrease the contrast and then lose some information, it is better to make the loss as small as possible. For this reason Korhunen-Loeve (KL) transforms [22] is used to integrate the image, the newly created image will have characteristics of dense information and best contrast. The five images input to KL transforms yields five outputs of principle components (*PCs*) images as Fig. (2) shows. Many literatures refers to that the first *PC* image contains about 75% of the information carried by the input five images [22], this makes the first *PC* is suitable for purpose of recognition, while other remaining transformed images are ignored. The chosen *PC* image is input to Haar wavelet transform [23] to be down sampled. Down sampling [24] is useful in diminution the size of the image, it is utilized to speed up the computations and enabling to match with huge data. The down sampling was implemented twice, i.e. the image becomes 64*64 resolution, such size embed tiny details and keeps the head cues of the face. The first transformed band is the average image that ready to apply the face detection on.



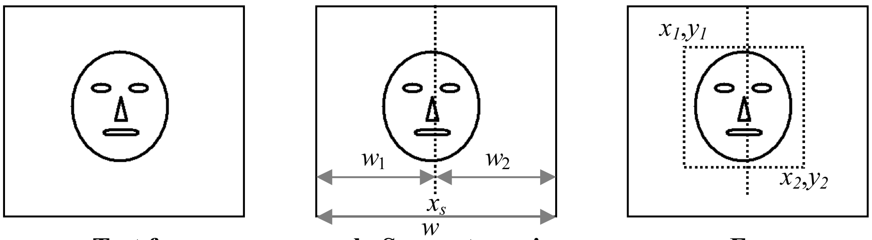
(a) Input set of 5 images (b) 5 resulted PC images (c) Wavelet applied on first PC.

Fig.(2) : The preprocessing stage applied on the input five images.

b- Face detection

Face detection is a process of determining the area that encloses the face in order to discriminate it from surroundings. The proposed face detection method depends upon the physical constituents of the face. It first determines the position of the vertical symmetry axis between the two right and left halves of the face, and then determine the coordinates of the frame (box) that enclose the face.

The determination of symmetry axis carried out by scanning the image vertically; the image is divided vertically into two parts, the width of the first part (w_1) may varies from one (w_1=1 pixel) up to width of the image (w), while the width of the second part (w_2) takes the remaining of the image width (i.e. $w_2=w-w_1$) as shown in Fig. (3-b). At each time of partitioning variation, a similarity matching is computed between the two image parts. The matching is just a difference between the interfered areas of the two parts after producing a mirror inversion on one of them. The position of the vertical symmetry axis on the width domain (x_s) is equal to the width of the first part (w_1) in which best matching is yield.



a. Test face. b. Symmetry axis. c. Face area.

Fig. (3): The two steps of the face detection.

Whereas the determination of the face coordinates is done by creating a box centered at the mid height of symmetry axis as shown in Fig. (3-c). The area behind the two parts of the box on both sides of the symmetry axis are, also, matched with each other, this matching is repeated; once by resizing the box (e.g. from 8 up to w), and another by shifting the box slightly down and up ward. The coordinates of the box (x_1,y_1 and x_2,y_2) that give best matching are saved as a coordinates of area contains just the face appearing in the image. The following pseudo code explains the detailed setting of the face detection process.

```
Begin
' To find the symmetry axis
Set w←64      ' width of the image
Set h←64      ' height of the image
Set As←4/3    ' set the frame aspect ratio 4/3(height/width)
Let B←8       ' size of the image block
Min1←w*h*127: Min2←Min1
Do for w1←1 to w-1
    w2←w-w1
    w3←min(w1,w2)
    Do for x←1 to w3
        Do for y←0 to h-1
            Ds(w1) ← Ds(w1)+|Img(w1-x,y)- Img(w1+x,y)|/(w3*h)
        Loop  '(y)
    Loop      '(x)
    If Ds(w1)≤Min1 then Min1←Ds(w1): xs←w1
Loop          '(w1)
' To find the box encloses the face
Set ys←h/2: ks←Int(min(xs,w-xs)/B)
Do for k←1 to ks
    Fx(k)←k*B: Fy(k)←As*k*B
    Do for x←1 to Fx(k)
        Do for y←ys-Fy(k) to ys+Fy(k)
            Db(k) ←Db(k)+|img(xs-x,y)-Img(xs+x,y)|/2*k*k
        Loop  '(y)
    Loop      '(x)
    If Db(k)≤Min2 then Min2←Db(k): L←k
Loop          '(k)
x1←xs-Fx(L): y1←ys-Fy(L): x2←xs+Fx(L): y2←ys+Fy(L)
End
```

c- Feature extraction and matching

This stage concerned with extracting and matching the face features. The employed features are concluded from the results of the DCT that applied on the determined face area, the results can analyzed: (1) Quantitatively, by studying the variation amount of the first coefficient (which is called detailed coefficient; *DC*) in the transformed image at same and different faces, and (2) Qualitatively, when the behavior of remaining coefficients (which is called approximated coefficients; *ACs*) are matched for the considered faces. Because of the cosine function is even, the behavior of transformed image (i.e. DCT coefficients) goes with the domain in damping oscillation form as Fig. (4) shows. Because of the first transformed values carry the most of the image information than others; they can utilize to be recognition features (parameters). The analysis shows that the *DC* value describes most of the information contained in the image, few information are distributing along other *AC* values.

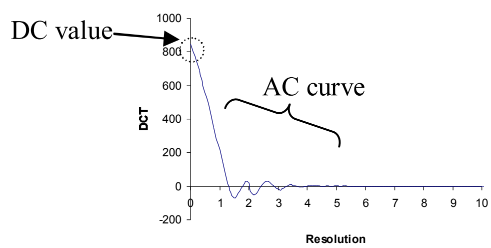


Fig.(4): Behavior of first few values in the transformed image (zigzag arranged).

Therefore, it can be assigned 70% from the weight of the recognition decision to the difference between the *DC* values of the target image and the models in the library, while 30% is assigned to the mean squired difference (MSD) between the *AC* curves of the target image and that of the models in the library. Therefore, the similarity measure between the test face and its correspondent model in the lookup table will be indicated the recognition percentage as follows:

$$S_{ki} = 0.7 \times \left(DC_{kj} - DC_{ij}\right)^2\Big|_{j=0} + \frac{0.3}{5}\sum_{j=1}^{5}\left(AC_{kj} - AC_{ij}\right)^2 \quad \text{.....................(1)}$$

Where, S_{ki} is the similarity measure between the k[th] test face and i[th] model, DC_{kj} and AC_{kj} are recognition parameters of the test face, while DC_{ij} and AC_{ij} are the recognition parameters of the i[th] face model in the library, see Fig.(5). As a result, the recognition is decided according to quantitative and qualitative comparison that accumulated to make the identification seems to be more accurate. Later, the analysis of S_{ki} leads to decide whether the face is belonging to any model in the library.

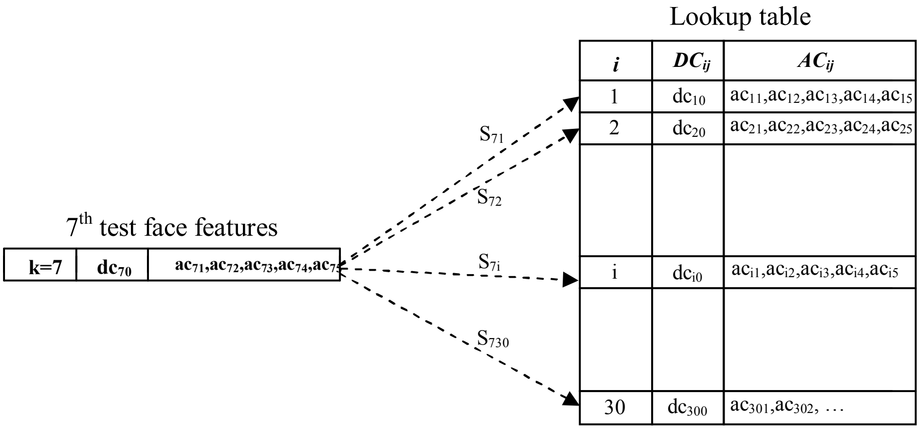
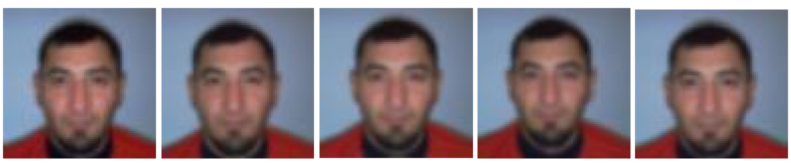


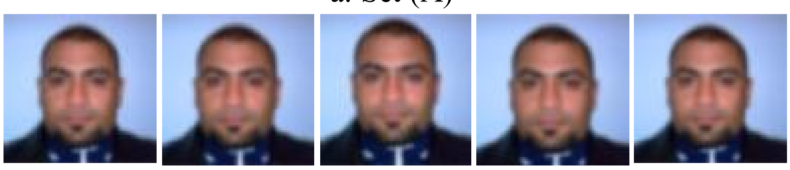
Fig.(5): Similarity measure computation between the k[th] face and i[th] model.

4. Experiment Results and Analysis

The library was established to be contains 30 training models. The KL is used to transforms each set (five low resolution images) at same pose and acceptable illumination to be one model. Fig. (6) displays two sample sets, these sets belong to faces contributed in the models of library. In order to yield an integrated image, each set inputs the KL transform. It is noticed the first resulted PC shown in Fig. (7) is integrated well since it appears more contrast and can resolved clearly in comparison with the input ones present in Fig. (6). Different amounts of illumination in the same set of images were found not very interest because the resulted image was compensated well by means KL transform. Also, it seen the down sampled image shown in Fig. 8 was losing tiny details and saving distinct geometrical ones, the geometrical features are position and size of the major face cues. Whereas Fig. (9) demonstrates face detection results, just the face is looked framed and ready for matching and recognition. Fig. (10) shows the behavior of first six DCT coefficients that zigzag arranged for the given two test faces.



a. Set (A)

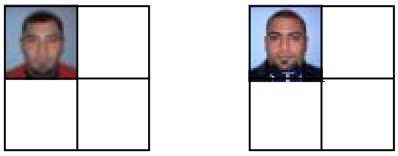


b. Set (B)

Fig.(6): Two sample sets of low resolution images.

a. PC1 of set (A). b. PC1 of set (B).

Fig.(7): Integrated image resulted from KL transform.



a. Face (A). b. Face (B).

Fig.(8): The down sampled image applied on the integrated image.

a. Face (A). b. Face (B).

Fig.(9): Detecting the face existing in the down sampled image.

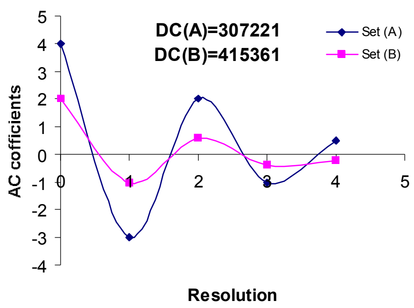


Fig.(10): The DC and ACs coefficients of the cosine transformed image.

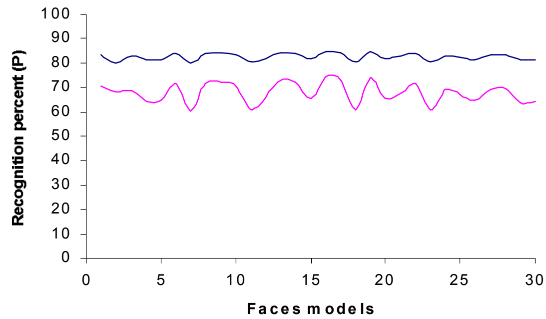


Fig.(11): Behavior of the adopted features versus weights.

The use of unequal contribution weights for the adopted recognition parameters shows the estimated recognition score by just *DC* coefficient was about 74%, while the use of *ACs* coefficients only gave a recognition score of about 26%. The coalition work of both *DC* and *ACs* make the recognition percentage rise up to 83% when random set of image was input.

The analytical study for the behavior of correct recognition decision shows the change in contribution weights was greatly affecting the recognition decision. Fig. (11) shows the behavior of the recognition percentage versus weights. The upper behavior represents the recognition of adopted weights (70% for *DC* and 30% for *ACs*) which shows higher recognition score, less fluctuation, and more stability in comparison with the lower behavior of equally weights (50% for *DC* and 50% for *ACs*). This refers to properly determination of the proposed weights. Fig.(12) presents monotonous results of recognition accuracy that may giving an impress about the stability of the correct recognition for employed member faces. It seen that the recognition score increases by increasing the codebook size, in turn the stability of recognition score increases with increasing the codebook, where one can note the least fluctuations shown in Fig.(12) occur at high codebook size.

Practically, one can notes the features that extracted from same two identified faces (belong to different sets) have no exact matching with each other, or with that of the library models, but they don't go far away from "at least" one model available in the library. It is found that the similarity measure that less than 62% leads to different faces, other wise leads to similar faces. That is because the similar head cues appear in same and different faces.

Beside the accuracy, the performance of the proposed engine is very important in such research work. The estimated time needed to make face recognition was about 12 *s* at which the test face was compared with 30 face models in the library. The slot time show an increasing behavior with increasing number of models in the library. Fig.(13) pictures how long the time needed to make a recognition decision with varying the code book size.

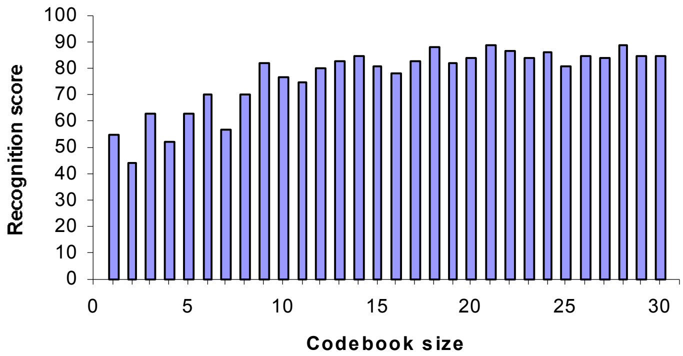


Fig.(12):The recognition decision percentage for 30 tested face.

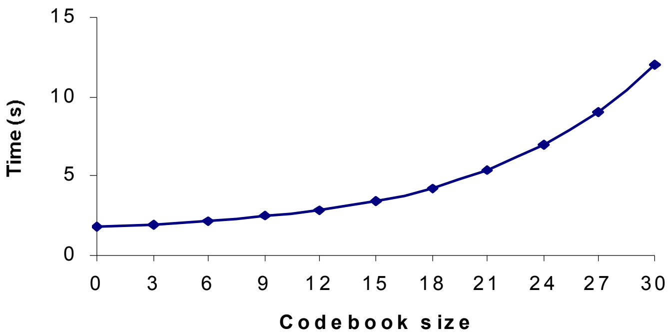


Fig.(13): The behavior of slot time versus number of models.

It should be mentioned the computation of prediction refers to that the slot time was far away from the real time when the target image is compared with a huge data (e.g. 1000 models), in which the recognition decision be late for 35 minutes. This is because of increasing the accuracy lead to make the used computations to be lengthy. This is done at assumption that the library contains an image not the recognition parameters (the existence of the images in the library instead of recognition parameters is intended in some applications). Whereas the recognition decision and matching were established during less time for same huge data when the recognition parameters are stored in the library rather than the images them selves. In this case the spent time needed to complete the recognition task was about 29 *s*, at which the used processor was 650 MHz. This percent is increasable whenever the efficiency of the processor performance improves, or using parallel task processor, both suggestions make the recognition approached to be decided in the real time. In general, the recognition performance was acceptable, which may tell that the proposed method was efficient to match faces, also the adopted DCT coefficients were succeeding recognition parameters. This encouraging result provides the chance to improve the proposed method and techniques to be utilized in the practical application.

5. References

[1] Darrell T., Demirdjian D., Checka N., and Felzenswalb P. *Plan-view trajectory estimation with dense stereo background models*. In proceedings of the international conference on computer vision, Vancouver, BC, July 2001, pp. 101–114.

[2] Adini Y., Moses Y., and Ullman S. *Face recognition: The problem of compensating for changes in illumination direction*, IEEE transactions on pattern analysis and machine intelligence, Vol. 19, No. 7, July 1997, pp. 721–732.

[3] Brunelli R. and Poggio T. *Face recognition: features vs. templates*, IEEE transactions on pattern analysis and machine intelligence, Vol. 15, No. 10, 1993, pp.1042-1052.

[4] Moghaddam B., Jebara T., and Pentland A. *Bayesian face recognition*, Pattern recognition, Vol. 33, 2000, pp.1771-1782.

[5] Belhumeur P. N., Hespanha J. P., and Kriegman D. J. *Eigenfaces vs. Fisherfaces: Recognition Using Class Speci_c Linear Projection*, IEEE transactions on pattern analysis and machine intelligence, Vol. 19, No. 7, July 1997, pp.711-720.

[6] Penev P. S. and Atick J. J.. *Local feature analysis: A general statistical theory for object representation*, Network: computation in neural systems, Vol. 7, No. 3, 1996, pp.477-500.

[7] Moghaddam B., Wahid W., and Pentland A. *Beyond eigenfaces-probabilistic matching for face recognition*, In Proceeding of IEEE conference on automatic face and gesture recognition, 1998, pp. 30–35.

[8] Belhumeur P. N., Hespanha J. P., and Kriegman D. J. *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*, IEEE transactions on pattern analysis and machine intelligence, Vol. 19, No. 7, July 1997, pp. 711–720.

[9] Kepenekci B. *Face Recognition Using Gabor Wavelet Transform*, PhD thesis, The Middle East Technical University, 2001.

[10] Globe B. *Face recognition fails in Boston airport,* Boston Globe, July, 2002.

[11] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, Jan 1990, pp.103–108.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol.3, No. 1, 1991, pp. 71–86.

[13] B Sch¨olkopf, A Smola, and KR Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Technical Report* No. 44, December 1996.

[14] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol.401, No.6755, October 1999, pp. 788–791.

[15] Hua Yu and Jie Yang. A direct LDA algorithm for high-dimensional data– with application to face recognition. *Pattern Recognition*, Vol.34, No.10, 2001, pp. 2067–2070.

[16] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*, Vol.14, No.1, Jan 2003, pp. 195–200.

[17] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Comput.*, Vol.12, No.10, 2000, pp. 2385–2404.

[18] K.-R. Muller, S. Mika, G. R¨atsch, K. Tsuda, and B. Sch¨olkopf. An introduction to kernelbased learning algorithms. *IEEE Trans. on Neural Networks*, Vol.12, No.2, Mar 2001, pp. 181–201.

[19] J Lu and KN Plataniotis. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. on Neural Networks*, pages, January 2003, pp.117–126.

[20] S. Mika, G. R¨atsch, J. Weston, B. Sch¨olkopf, and K.R. Mullers. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, Aug 1999, pp. 41–48.

[21] S. Mika, G. R¨atsch, J. Weston, B. Sch¨olkopf, A. Smola, and K. M¨uller. Invariant feature extraction and

classification in kernel spaces. *Advances in Neural Information Processing Systems 12*, 2000, pp. 526–532.

[22] Lee J. *Optimized quad tree for Korhunen-Loeve transform in multispectral image coding*, IEEE transaction on image processing, Vol.8, April, 1999.

[23] Buccigrossi R. W. and Simoncelli E. P. *Image comperission via joint statistical characteristics in the wavelet domain*, IEEE transaction on image processing, Vol. 8, No. 12, 1999, pp. 1688-1700.

[24] Chiptrasert B. and Rao R. K. *Discrete cosine transform filtering*, Signal processing, Vol. 19, No. 3, 1990, pp.233-245.

الخلاصة

يهدف هذا البحث الى جعل الحواسيب قادرة على تمييز الوجوه بدون التدخل البشري. وقد تم ذلك بأخذ خمسة صور للوجه الأختباري، ومن ثم مكاملة المعلومات في الصور الخمسة لتكون مركزة في صورة واحدة بأستخدام تحويل KL. بعد ذلك طّبق التحويل المويجي لتقليص حجم الصورة المكاملة وتحويل الجيب تمام المتقطع لأستخلاص الصفات التي تستخدم في مهمة التمييز.تم التمييز من خلال البحث عن تطابق للصفات المستخلصة بين الصورة المتحولة ونماذج الصور في المكتبة. أعطت منظومة التمييز نسبة تمييز واعدة تتراوح بحدود 83% عندما تكون حظوظ مشاركة الصفات في القرار غير متساوية.