World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:5, 2007

Walsh-Hadamard Transform for Facial Feature Extraction in Face Recognition

M. Hassan, I. Osman, and M. Yahia

***Abstract*—**This Paper proposes a new facial feature extraction approach, Wash-Hadamard Transform (WHT). This approach is based on correlation between local pixels of the face image. Its primary advantage is the simplicity of its computation.

The paper compares the proposed approach, WHT, which was traditionally used in data compression with two other known approaches: the Principal Component Analysis (PCA) and the Discrete Cosine Transform (DCT) using the face database of Olivetti Research Laboratory (ORL). In spite of its simple computation, the proposed algorithm (WHT) gave very close results to those obtained by the PCA and DCT.

This paper initiates the research into WHT and the family of frequency transforms and examines their suitability for feature extraction in face recognition applications.

***Keywords*—**Face Recognition, Facial Feature Extraction, Principal Component Analysis, and Discrete Cosine Transform, Wash-Hadamard Transform.

I. Introduction

FACE Recognition has gained much attention in recent years and has become one of the most successful applications of image analysis and understanding. A typical application is to identify or verify the person of a given face in still or video images. The important applications of face recognition are in areas of biometrics: computer security and human computer interaction [5]. A general statement of face recognition problems can be formulated as follows [19]:

- Due to variations in pose, illumination and facial expression, the face appearance of an object possesses a complex density (manifold), so large numbers of samples are required to sufficiently represent the complex density.
- Because of the limitations of image acquisition, practical face recognition systems store only a small number of samples per subject, while image sample is high dimension, which leads to dimensionality reduction of face images to extract out the most important features of the human face.
- One can explore the strong visual similarity among face images of a large number of different subjects.

M. Hassan and M. Yahia are with Faculty of Mathematical Sciences, University of Khartoum, Khartoum, Sudan (e-mail: Moh_bash@hotmail.com, meyahia@hotmail.com).

I. Osman is with Collage of Computer Science and Informatiıon Technology, Sudan univcersity of Science and Technology, Sudan (e-mail: izzeldin@acm.org).

Such problems can be solved by using different types of features extraction methods categorized in the following points [20]:

- Holistic methods, which use the whole face region as raw input to the recognition system. One of the most widely used methods in face representation is Principal Component Analysis approach. PCA is a traditional statistical approach applied in face recognition problems. It projects the data images into eignspace that encodes the variation among known face images. Resulting in eigenvectors of the set of faces, which they do not necessarily correspond to isolated features such as eyes, ears, and noses [15].

- Feature-based methods that depend on local features such as the eyes, nose and mouth, which are first extracted and their locations and local statistics are fed into a structural classifier.

A. Traditional Approaches

PCA is a Holistic method. It has some significant limitations. It is not capable of extracting local features like structures in objects. In general, PCA produces global non topographic linear filters [9]. Discrete cosine transforms (DCT) are used to reduce image information redundancy because only a subset of the transform coefficients are necessary to preserve the most important facial features such as hair, eyes and mouth [21].

B. The New Approach

The approach that is presented in this paper focuses on finding out a new model for facial features extraction that depends on local correlation between image pixels. The proposed method is Wash-Hadamard Transform (WHT).

The paper is organized as follows: in section II, an overview of methods that are applied in facial features extraction. Section III discusses the proposed approach for facial features extraction. In section IV, we present the results of experiments applying the proposed algorithm compared with DCT and PCA. In section V, we conclude the paper.

II. Facial Features Extraction Approaches

A. Principal Component Analysis

The Conventional derivation of PCA is in terms of a standardized linear projection which maximizes the variance in the projected space. For a set of observed d-dimensional data vectors $\{x_i\}_1^n$, the M dominant eigenvectors of the sample covariance matrix formulate as follows [6][10][14][15][12]:

$$C = \sum_{\mathrm{i}} (x_i - \mu)^{\mathrm{T}} (\mathrm{x}_i - \mu) \quad (1)$$

where μ is the sample data mean, each v_i is an eigenvector of the Covariance Matrix (C) having associated eigenvalue λ_j:

$$Cv_j = \lambda_j v_j \quad (2)$$

The data projected to eigenface space gives the vector y_i :

$$y_i = V^T (x_i - \mu) \quad (3)$$

where $V = (v_1, v_2, ..., v_m)$.

The PCA enjoys the important advantages: model parameters can be computed easily and ease of matrix multiplication required for finding parameters [6]. For PCA an unknown person identified by applying the Euclidian distance formula [16]:

$$D(y_i - y_k) = \sum_{i=1}^{k} (y_i - y_k)^2 \quad (4)$$

where D refers to the Euclidian distance, y_i is unknown sample, y_k represent samples of the training space, and λ_i are eigenvectors.

B. Discrete Cosine Transform (DCT)

The cosine transform, like the Fourier transform, uses sinusoidal basis functions. The difference is that the cosine transform basis functions are not complex; they use only cosine functions and not sine functions [13]. 2D DCT based features are sensitive to changes in the illumination direction [2].The idea of using the transform for facial features extraction is summarized as follows:
the given face image is analyzed on block by block basis given an image block $I(x, y)$, where $x, y = 0,1,..., N_P - 1$, and result is an $N_P \times N_P$ matrix $C(u,v)$ containing 2D DCT coefficients ,The DCT equations are given by formulas (5), (6), (7) [13][2][3][17][1][4] below:

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N_P-1} \sum_{y=0}^{N_P-1} I(x,y) B(x,y,u,v) \quad (5)$$

For $u, v = 1,2,.., N_P - 1$ where

$$\alpha(v) = \begin{cases} \sqrt{1/N_P} & for\ v = 0 \\ \sqrt{2/N_P} & for\ v = 0,1,...N_P \end{cases} \quad (6)$$

And (7) is shown in the last page.

To ensure adequate representation of the image, each block overlaps its horizontally and vertically neighboring blocks by 50%, thus for an image which has N_Y rows and N_X columns, there are N_D blocks found by following formula:

$$N_D = (2(N_Y / N_P) - 1) \times (2(N_X / N_P) - 1) \quad (8)$$

Compared to other transforms, DCT has the advantages of having been implemented in a single integrated circuit because of input independency, packing the most information into the fewest coefficients for most natural images, and minimizing block like appearance [21][22]. An additional advantage of DCT is that most DCT coefficients on real world images turn out to be very small in magnitude [21].

The Network of Networks (NoN) System presented in [15][18] , finds out a method for representation of 2D Blocks of DCT to use in face recognition: let $C \in R^{N \times N}$ be the 1-D vector representation, define $\Phi : R^{N \times N} \to \Re$ as function over the vector C, the block-level coefficient , denoted as λ , is based on taking λ as the sum of the squired of the DCT coefficients c_i in the block, let C be the set of DCT coefficients of an $N \times N$ block arranged in raster scan order, the block-level DCT Coefficient λ founded in five schemes: S1,S2,S3,S4,S5 as shown below:

$$\lambda = \sum_{i=2}^{N^2} c_i^2 \quad (S1)$$

where $i = 2,...N^2$ such that $c_i \in C$.

$$\lambda = \sum_{i=2}^{N^2} |c_i| \quad (S2)$$

$$\lambda = \frac{\sum_{i=2}^{N^2} c_i^2}{N^2 - 1} \quad (S3)$$

$$\lambda = \frac{\sum_{i-2}^{N^2} |c_i|}{N^2 - 1} \quad (S4)$$

$$\lambda = \frac{\sum_{i=2}^{N^2} |c_i - \mu|}{N^2 - 1} \quad (S5)$$

where μ is mean of the DCT Coefficients in the N×N block Computed by:

$$\mu = \frac{\sum_{i=2}^{N^2} c_i}{N^2 - 1} \quad (9)$$

Suppose $x_1, x_2, \ldots x_n$ are computed block-level DCT coefficients for a given image, where n is the number of N×N blocks in the input image, the upper bound (b_i in equation 2.2.6) and lower bound
(a_i in equation2.2.7) can be determined by :

$$b_i = \beta.\max(x_1, x_2, \ldots x_n) \qquad i = 1,2,\ldots,M \quad (10)$$

$$a_i = \beta.\min(x_1, x_2, \ldots, x_n) \qquad i = 1,2,\ldots,M \quad (11)$$

where M is the number of faces in the gallery and β>=1 is a factor to extend the bounds. Then the last representation is vector $(z_1, z_2, \ldots, z_n)$ that is output of normalization of original blocks in equation 2.2.6 as follows:

$$z_i = \frac{x_i - a_i}{b_i - a_i} \quad (12)$$

C. Wash-Hadamard Transform (WHT)

The WHT differ from the Fourier and Cosine Transforms in the basic functions that are not sinusoids. The WHT is used in data compression [11]. The basic functions are based on square or rectangular waves with peaks of ±1. Here the term rectangular wave refers to any function of this form, where width of the pulse may vary. One primary advantage of the transform is that the computations are very simple. When we project an image onto the basis functions, all we need to do is multiply each pixel by ±1 as seen in WHT equation assuming N×N image [13]:

$$WH(u,v) = \sum_{r=0}^{N-1} \sum_{c=0}^{N-1} I(r,c)(-1)^{\sum_{i-0}^{n-1}[b_i(r)p_i(n) + b_i(c)p_i(v)]} \quad (13)$$

where $N = 2^n$, the exponent on the (-1), and $b_i(r)$ is found by considering r as a binary number and finding the i[th] bit, for example:

n=3 (3 bits, so N=8), and r = 4
So r in binary is 100, giving
$b_2(r) = 1, b_1(r) = 0, b_0(r) = 0$

In addition, $p_i(u)$ is found as follows:

$$p_0(u) = b_{n-1}(u)$$
$$p_1(u) = b_{n-1}(u) + b_{n-2}(u)$$
$$p_2(u) = b_{n-2}(u) + b_{n-3}(u)$$
.
.
.
$$p_{n-1}(u) = b_1(u) + b_0(u)$$

WHT is not classified as a frequency transform, because the basis functions do not exhibit the frequency concept in the manner of sinusoidal functions. However, it considers the number of zero crossings (or sign changes), this is a measure comparable to frequency, and is called sequence [13].

III. WHT FOR FACIAL FEATURE EXTRACTION

A. The Face Database

We have applied the proposed method above (WHT) using human face image database from AT&T Cambridge laboratories in UK created at the Olivetti Research Lab (ORL) prospected in web sites in [7] [8]. The database contains 40 classes of persons, each person is represented by 10 images with different pose. There are 4 females and 36 males, and the size of each image is 48×48 pixels, with 256 gray levels per pixel [17]. Fig. 1 shows samples of the database.

Fig. 1 Face images of ORL database

B. WHT

The basic idea of using WHT in facial feature extraction depends on applying the transform functions described previously in Section II.*C*. Equation (13). The WHT applied in Squared size gallery space images generating M×M blocks from each image, then we applied the five schemas (S1, S2, S3, S4, S5) described in Section II.*B*. But we modified schemas S2, S3 by S2', S3' that is to reduce data redundancy, S2' and S3' defined as follows:

$$\lambda = \sum_{i=2}^{N^2} |c_i| - |\mu| \quad \text{(S2')}$$

$$\lambda = \frac{\sum_{i=2}^{N^2} (c_i - \mu)^2}{N^2 - 1} \quad \text{(S3')}$$

We made use of the Block-level DCT coefficients computed from Equation (12). All blocks of the original image represented in 1-D vector $(z_1, z_2, \ldots z_n)$.

IV. Experimental Results

A. Experiments Setup

To find the block levels of the images, we choose block size 8×8, so the whole number of blocks of each image is 11×11 blocks (depend on N_D formula in Section II.*B*), so the block level coefficients calculated using the five S's Schemas descried in Section II.*B*, a vector of 121 dimension computed from all blocks representing the image, finally, each image would be represented by 5 vectors of coefficients. Fig. 2 below shows the hierarchal process described for facial feature extraction obtained from WHT, and Fig. 3 shows samples of features generated from using DCT.

In addition to that, the results of WHT compared with the result obtained using PCA and DCT methods, using Euclidian distance form for identification. The Length of the data in all experiments is 11x11.

B. Experiments

In accordance with the specifications, we conducted three different experiments .In these experiments:

Number of subjects training set is 40
Size of training set is 40, one sample per subject.
Size of training set is 80, two samples per subject.
Size of training set is 120, three samples per subject.
Size probe test data 280, seven samples per subject.
Training samples different from probe samples.

1. Experiment One

Recognition rates were computed using DCT and WHT, and the results were compared. Table I shows the results and comparison.

2. Experiment Two

Three methods compared using one representation for facial features for each method (S1 formula selected for representation). Table II below shows the results of recognition rates.

3. Experiment Three

We combined DCT with PCA, then WHT with PCA. For more details we applied DCT then we applied PCA on the result, and we applied WHT then PCA on the result of WHT representation. Recognition rate of the experiment is presented in Table III.

TABLE I
RECOGNITION RATES USING 5S'S

Size of training set (Samples per subject)	Recognition Rates		
Method	1	2	3
WHT	52.43%	62.93%	66.57%
DCT	54.00%	65.36%	68.93%

TABLE II
RECOGNITION RATES USING S1

Size of training set (Samples per subject)	Recognition Rates		
Method	1	2	3
PCA	56.07%	64.64%	70.36%
DCT	53.21%	62.86%	65.36%
WHT	55.00%	63.93%	66.79%

TABLE III
RECOGNITION RATE FOR COMBINATION

Method	Recognition Rates for training set size of 3 samples per subject
DCT+PCA	65.36%
WHT+PCA	66.79%

The results show that the WHT is comparable to the well known face recognition feature extraction methods, PCA and DCT. The DCT and WHT reach their results by using the correlation of local pixels of a single image ,while PCA correlates pixels from the whole training set .The WHT results are close to PCA and DCT results .hence the WHT and DCT are both easier to compute than PCA.

V. Conclusion

In this paper, we applied DCT and WHT on ORL face database using one to three training samples for facial feature extraction. DCT and WHT are traditionally applied in data segmentation and data compression. The results of both algorithms were compared with the result of the traditional facial feature extraction algorithm (PCA). The recognition rates using both algorithms were close to PCA recognition rates, when we use Euclidian distance formula for recognition.

The results show that there is no noticeable improvement in the results, when we project DCT or WHT data space to eigenface using PCA. The PCA correlates pixels in the whole training samples, while DCT and WHT correlate pixels in the face image individually.

This paper initiates the research into WHT and the family of frequency transforms and examines their suitability for facial feature extraction in face recognition applications.

References

[1] Almas M. , Younus M. and Basit A., "Anew Approach to Face Recognition Using Dual Dimension Reduction", International Journal of Signal Processing Volume 2 Number 1,ISBN 1304-4494, 2005.
[2] Conrad S., Kuldip K., "Fast Features for Face Authentication Under Illumination Direction Changes", Pattern Recognition Letters 2, 2409-2419, 2003.
[3] Eicheler, S., Muller, S. and Rigoll, G., "Recognition on JPEG Compressed Face Images Based on Statistical Methods", Image Vision Compute. 18(4), 279-287 ,2000.
[4] Hazem K.,Rainer S., "Local Appearance Based Face Recognition Using Discrete Cosine Transform", 13th European Signal Processing Conference (EUSIPCO), Antalia, 2005.

[5] Kresimir,D.,Mislav,G.,Panos,L,"Appearnce-based Statistical Methods for Face Recognition",47 th International Symposium ELMAR-2005,Zedar,Croatia, June 2005.
[6] Musa, M., "Towards Finding Optimal Mixture of Subspaces for Data Classification", PhD Thesis ,The Middle East Technical University, Turkey,Oct,2003.
[7] O.R.L(ORL), www.mambo.ucsc.edu/psl/olivetti.html.
[8] O.R.L(ORL),www.uk.research.att.com/facedatabase.html .
[9] Penio S. and Joseph J , "Local Feature Analysis: A general Statistical Theory for Object Representation", Network: Computation in Neural Systems 7(3),p 477-500, 1996.
[10] Picard R. and Kabir,T.,"Finding Similar Patterns in large image databases", M.I.T,Technical Report No. 205, IEEE ICASSP ,Minneapolis, MN, Vol. V.,PP.161-164, Apr. 1993.
[11] Rauf K.,Vladimir A., and Leonid P. ,"Face Recognition Algorithm on the Basis of truncated Walsh-Hadamard Transform and Synthetic Discriminant Functions", Proc. 6th International Conf. on Face Recognition and Gesture Recognition ,Seoul, Korea, 2004.
[12] Ross, B.,Kai, S,Bruce A.,and Geof, H.,"A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant subspaces for Face Recognition", IEEE Society Conference on Computer Vision and Pattern Recognition, Vol. 1, p535 ,2001.
[13] Scott E., "Computer Vision and Image Processing ",Prentice Hall PTR, Prentice Hall,Inc,Upper Saddle River,NJ 07458,ISBN :0-13-264-599-8,USA,1999.
[14] Shaokang, S.,Brain, C. and Sai, S. ,"Face Recognition With APCA in Variant Illuminations", Workshop on signal processing and its applications, Brisbane, Dec 2002.
[15] Turk,M., and Alex, P., "Face Recognition using Eignfaces", Proc. IEEE Conf. on Computer Vision and pattern Recognition, pp.586-591, 1991.
[16] Turk M. and Pentland A., "Eigenfaces for Face Detection/Recognition" , Journal of Cognitive Neuroscience ,Vol. 3, No. 1 , pp. 71-86, 1991.
[17] Willie L. Scott, II, "Block-Level Discrete Cosine Transform Coefficients for Autonomic Face Recognition", Ph.D Thesis Submitted to Graduate Faculty of Louisiana State U. and Agricultural and Mechanical Collage , USA, May 2003 .
[18] Wellie L. Scott,II, "The NoN Approach to Autonomic Face Recognition", print arXiv:CS/0603042,2006cs, 2006.
[19] Zhao,S.,Chellappa,R.,"Multiple exampler Discriminant Analysis for Face Recognition" , In proc. of Intl. Conf. on Pattern Recognition, August 2004.
[20] Zhao,S.,Chellappa, R.,Phillips, J.and Rosenfeld, A., "Face Recognition: A literature Survey", ACM Computing Surveys ,Vol 35, No.4, pp.399-458, Dec/2003.
[21] Zhengju P., Hamid B., "High speed Face Recognition based on Discrete Cosine Transforms and Neural Networks", IEEE Trans. on Pattern Analysis & Machine Intelligence, 1999.
[22] Zhengju P., Rod A., Hamid B., "Dimensionality Reduction of Face Images Using Discrete Cosine Transforms for Recognition", IEEE Conf. on Computer Vision and A Pattern Recognition (CVPR2000), South Carolina, USA, 2000.

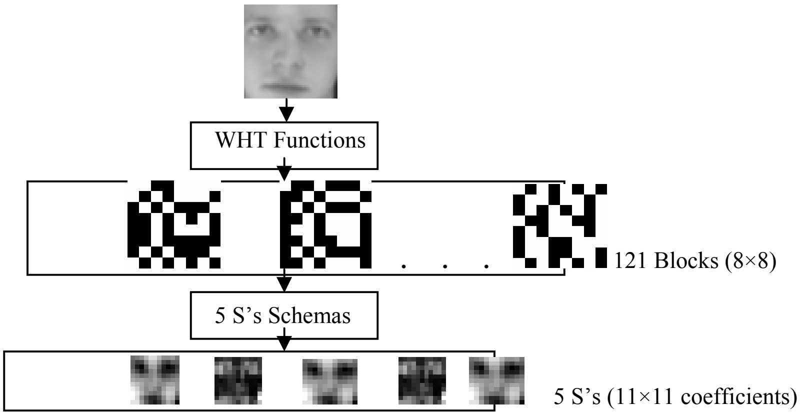

Fig. 2 Hierarichal Process of Finding Face Feature using WHT

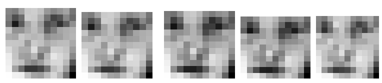

Fig. 3 5S's 11×11 DCT coefficients

$$B(x, y, u, v) = \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (7)$$