Journal of Engineering and Development, Vol. 10, No. 1, March (2006) **ISSN 1813-7822**

High Performance Technique for Face Recognition Based on DCT

Asst. Prof. Dr. Khamis A. Zidan
Computer and Software Engineering Department
Al-Mustansiriya University, Baghdad, Iraq

Abstract

Face recognition is an important and fundamental problem in computer vision, and there have been many attempts to address it. Correlation and high information redundancy in face images result in inefficiencies when such images are used directly for recognition. In this paper, an efficient hybrid approach to face recognition is presented, which combines image compression and neural network (NN) techniques together. The compression is achieved by applying fast discrete cosine transforms (DCTs) 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 outline, eyes and mouth. The compressed transform coefficients are used for back propagation NN classification. A high recognition rate can be achieved by using a very small proportion of transform coefficients. This makes DCT-based face recognition much faster than other approaches.

The proposed system is implemented and tested using gray-scale images contains 40 distinct persons, each person having 10 different images and it gives good performance at high speed.

الخلاصـــة

تبرز أهمية البحث في استخدامه في العديد من التطبيقات الأمنيـة والتـي تشمل التعرف علـى الأشـخاص المخـولين بدخولهم الأماكن المرخص لهم بدخولها فقط، والمراقبـة الأمنيـة، والتحقيقـات الجنائيـة الخاصـة بـالتعرف علـى المشتبه بهم ومراكز المعلومات الحساسة والتي لا يمكن الوصول إلـى بياناتـها إلا من قبل أشـخاص معينـين. وحيث إن وجـه الإنسـان عبارة عن نمط معقد متعدد الأبعاد وعملية تمثيل هذا النمط أو أيجاد صيغة رياضية لتمييز الوجه تعتبر عملية صعبة. لذا يهدف البحث لإيجاد طريقة للتعرف على وجه شخص وتمييزه من بين مجموعـة كبيرة من وجوه الأشـخاص والمخزونـة في قاعدة البيانـات داخل ذاكرة باستخدام الحاسبات. فـي هذا البحث اقترحت طريقـة هجينـه لتمييـز الوجـوه وتتكـون هـذه الطريقة من مراحل: المرحلة الأولى تتضمن تقليص حجم الصورة باستخدام (DCT) ذات خوارزميـة سريعة. المرحلـة الثانية تتضمن إدخال المعلومات الخارجة من المرحلة السابقة إلـى الشبكة العصبية متعددة الطبقات (MLP) عن طريق تدريبها بخوارزمية الانتشار العكسي. حيث يتم جمع الإخـراج النـاتج عن كل جزء من الشبكة العصبية متعددة الطبقات للحصول على الخرج النهائي للشبكة وربط تلك الخواص إلى شفرة التمييز بواسطة شبكة عصبية متعددة الطبقات بحيث يماثل هذا النظام العين البشرية من حيث تركيبية وكفائتة.

تم تنفيذ نظام تمييز الصور وتصنيفها علـى أربعين صنف لصور الوجوه البشرية ذات مستويات التدرج الرمادي وكان أداء النظام سريعا جدا ونسبة التمييز عالية.

1. Introduction

Computerized human face recognition has been an active research area for the last years. There is a wide range of military and civilian applications such as identity authentication, access control, digital libraries, bankcard identification, mug shots searching, and surveillance systems.

A general statement of the problem of face recognition can be formulated as follows: it is used to identify one or more persons from still images or a video image sequence of a scene by comparing input images with faces stored in a database. The solution of the problem involves segmentation of faces from cluttered scenes, extraction of features from face region, identification, and matching [1].

High information redundancy present in face images results in inefficiencies when these images are used directly for recognition, identification and classification. Typically one builds a computational model to transform pixel images into face features, which generally should be robust to variations of illumination, scale and orientation, and then use these features for recognition. Several techniques for facial feature extraction have been proposed. They include methods based on statistical feature, template matching, geometrical features, and NNs.

Statistical features are usually generated by algebraic methods such as principal component analysis (PCA) [2], or the closely related Karhunen-Loeve Transform (KLT) [3], or singular value decomposition. These features are in the form of a set of orthogonal bases such as principal components, or the eigenvectors (referred to as eigenfaces) [4]. Once the eigenvectors are chosen, any image in the gallery (set of training images) can be approximately reconstructed with a linear combination of eigenfaces, and their components will be stored in memory. For an unknown image, its components are calculated by projecting it to the face space (space generated by eigenfaces) and looking for the closest match.

Template matching methods operate by performing direct correlation of image segments [5,6]. Template matching is only effective when the query images have the same scale, orientation, and illumination as the training images. A simple version of template matching is that a test image represented as a two-dimensional array of intensity values is compared using a suitable metric, such as the Euclidean distance, with a single template representing the whole face. There are several other more sophisticated versions of template matching on face recognition. One can use more than one face template from different viewpoints to represent an individual's face. A face from a single viewpoint can also be represented by a set of multiple distinctive smaller templates. The face image of gray levels may also be properly processed before matching

The geometrical approach represents faces in terms of structural measures that include parameters such as ratios of distances, angles, and areas between elementary features such as eyes, nose, mouth or facial templates such as nose width and length, mouth position and chin type [5,7]. Then the features are used to recognize unknown images mainly by matching to the nearest neighbor in the stored database.

Neural network methods generally operate directly on an image-based representation (i.e. pixel intensity array). This class of methods has been more practical and reliable as compared to geometric feature based methods. The attractiveness of using NN could be due to its nonlinearly. Hence, the feature extraction step may be more efficient than the linear Karhunen-Loeve methods [8].

Its difficult to select a representation that could capture features robustly, most approaches avoid the feature extraction procedure by feeding the pixel images directly to NNs and making use of the ability of NNs as an information processing tool [9]. Nevertheless, Lawrence et. al. [10] applied self-organizing map (SOM) as a feature extractor and then the generated features were exploited as the input of a convolutional NN for recognition, a much similar architecture to neocognitron [11]. Training either the SOM or the convolutional NN is tremendously computationally expensive. Dawwod [12] presented a modified image recognition system based on the neocognitron model for feature extraction and multilayered feed forward to associate the feature with their labeled recognition code.

For the face recognition techniques mentioned above, most approaches have to build a database to store the features from the known faces in order to compare the features extracted from an unknown image with that in the database, while others, like convolutional NN approach, are slowed due to training period [13].

In this paper, high performance system face recognition is presented. In developing view based face recognition that uses machine learning, three main subproblems arise. First, the big dimension of original bitmap face images that is to be fed to the NN which is expensive in terms of the time consuming and space complexity of the network. It can be solved by reducing the input space dimension, while extracting only the most relevant information using DCTs as a way of information packing and exist fast algorithms to compute DCT, which makes it extremely competitive in terms of computational complexity. Redundancy removal to facilitate data processing and image categorization is not a new idea [14,15]. DCTs is 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 outline, eyes and mouth.

Second, the DCT coefficients are fed into a backpropagation NN [16]. Arbitrate neural-networks must be trained to deal with variation in distinguishing faces, which geared towards the use of hybrid system that implements layers of different types of networks. Each of which specialize in a specific task and guarantee a good performance under usual conditions, where lighting, rotation and tilting effects should not affect recognition. Third, the outputs from multiple recognition must be combined into a single decision about the face for classification.

2. Information Packing

2-1 Transform Efficiency

During process an acceptable scheme for compression, is achieved by dividing a bitmap image file M∗M pixels into subimages of N∗N pixels that are processed individually, from left-to-right and top-to-bottom.

In order to discuss the efficiency parameters for N∗N pixel block. The most significant parameters considered for evaluating the transform efficiency are [17]:

i) Correlation reduction: It is the ability of transform to produce uncorrelated coefficients.

ii) Energy packing: It is a measure to the ability of a transform to pack the signal energy into the first few transformed coefficients. Thus, the packing efficiency is defined as the relative amount of energy resident in the first Z∗Z coefficients relative to the total energy resident in all N∗N coefficients,

iii) Computational efficiency: The problem of implementation is considered as principal criteria compared with other efficiency parameters in the considerations related to select a suitable transform for a certain applications. Major efforts are spent toward the development of fast transform techniques. These methods are capable to reduce the number of arithmetic operations required to perform the transform mathematics. All these fast methods which are based on the idea of decomposing the transformation matrix into a product of sparse matrices, lead to divide the large transform calculation into a sequence of much simpler operations.

Although the KLT, is the optimal transform in information packing sense, but the DCTs will be applied instead of KLT to face images. Since KLT is data dependent and obtaining the KLT basis images, in general, is a nontrivial computational task, whereas there exist fast algorithms to compute 2-D DCTs, which makes DCT extremely competitive in terms of computational complexity. Therefore, due to widespread use of DCT, many fast DCT algorithms have been proposed. The most efficient one among these algorithms, in term of the number of arithmetic operations, is the direct approach, which introduced by Cho and Lee [18].

2-2 Fast Algorithm of Discrete Cosine Transform

The cosine transform, like the Fourier transform, uses sinusoidal basis function. The difference is that the cosine transform basis functions are not complex. They use only cosine functions without sine functions and provide a high-energy compaction and increase the computational speed. The 2-D forward DCT for N∗N block pixels is given by:

$$C(u,v)=\frac{2W(u)W(v)}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left(\frac{u\pi(2x+1)}{2N}\right)\cos\left(\frac{v\pi(2y+1)}{2N}\right) \quad \text{……....…...... (1)}$$

and the inverse DCT is:

$$f(x,y) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} W(u)W(v)C(u,v)\cos\left(\frac{u\pi(2x+1)}{2N}\right)\cos\left(\frac{v\pi(2y+1)}{2N}\right) \quad \text{.........} \ \textbf{(2)}$$

where, $\mathbf{W(i)} = \begin{cases} 1/\sqrt{2} & \textbf{if } \ \mathbf{i = 0} \\ 1 & \textbf{otherwise} \end{cases}$

Each block of image data is subjected to a forward DCT, which converts the pixel values into their corresponding frequency coefficients. A block of pixels is submitted to the forward DCT, which returns blocks of frequency coefficients. DCTs are calculated by applying a series of weighting coefficients to the pixel data.

Fundamental properties in practical implementations are that the forward DCT and inverse DCT equations contain transcendental functions with perfect accuracy, provides a good approximation to the optimum performance of the KLT and fast algorithms. Thus, many fast DCT algorithms have been proposed and simulated for the computation of 2-D DCT. But the most efficient one among these algorithms is a fast 4∗4 DCT for the recursive 2-D DCT [18], which is used in the proposed system.

The algorithm is based on the decomposition of 4∗4 DCT into four 4-point 1-D DCTs. Thus, only 1-D transformations and some additions are required. Thus, $2^m * 2^m$ DCT can be computed using 4∗4 DCT recursively for any integer m. An interesting merit of the recursive DCT algorithm is that its structure is regular and systematic, and only real arithmetic is required. The main idea of the 2-D recursive DCT algorithm is that N∗N DCT can be computed from the next lower DCT matrices for any desired value of $N = 2^m$, $m > 2$. An example of this faster algorithm for m = 2 is shown in the flow graph form in **Fig.(1)**.

Because of the orthogonality of the DCT matrix, the inverse DCT is obtained by inverting the signal flow graph for the forward transform, if the scaling factor is not considered. Hence, the signal flow graph of the inverse DCT algorithm is just the inverse of **Fig.(1)**, in which the 1-D DCT units are replaced by 1-D inverse DCT ones.

2-3 Zig-Zag Process

The output of the forward DCT is set of N∗N coefficients, whose values are uniquely determined by the particular N∗N pixel block [17]. The coefficient with zero frequency in both dimensions "DC coefficient", represents the average intensity of block, that is located in the upper left of the block, and the remaining (N^2-1) coefficients "AC coefficients" represents components of increasing horizontal and vertical frequency as shown in **Fig.(2)**. Because of the typical slow variation of the pixel values from point to point across an image the forward DCT processing step lays the foundation for achieving data compression by concentrating most of the information in the lower frequencies. This concentration of information is further increased by using Zig-Zag process. It is reorders the DCT coefficients in Zig-Zag order which groups non-zero coefficients together and creates longer runs of zero or near zero coefficients. This procedure is based on a simple mapping of the 2-D transform domain on to a 1-D ordering.

The proposed algorithm moves through the block along diagonal paths, selecting what should be the highest value coefficient first, and working toward the values likely to be lowest, as shown in **Fig.(3)**.

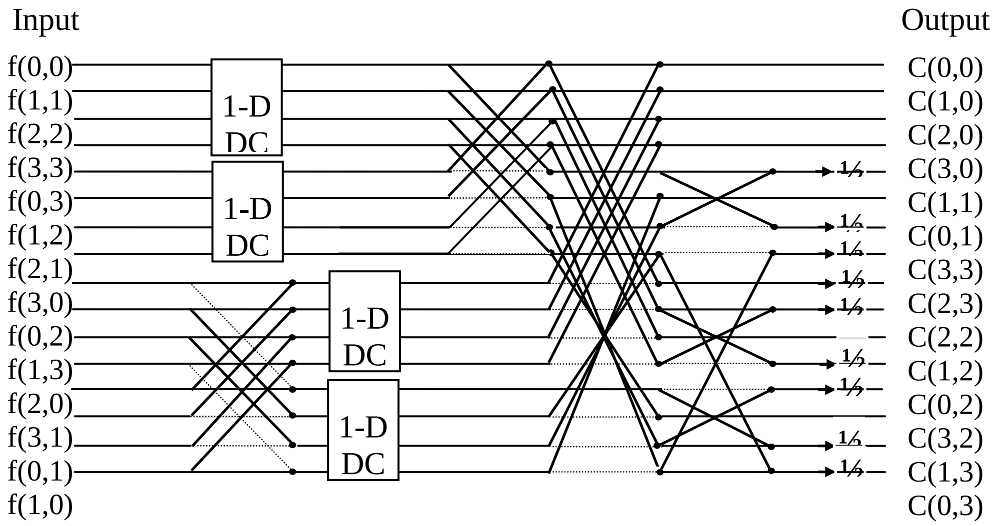


Figure (1) Fast 4∗4 Forward DCT Flow Graph
Doted Lines Represents Transfer Factors -1,
and Solid Lines Represent Unity Transfer Factor
• Represent Adders and → with ½ Represents Multiplication by ½

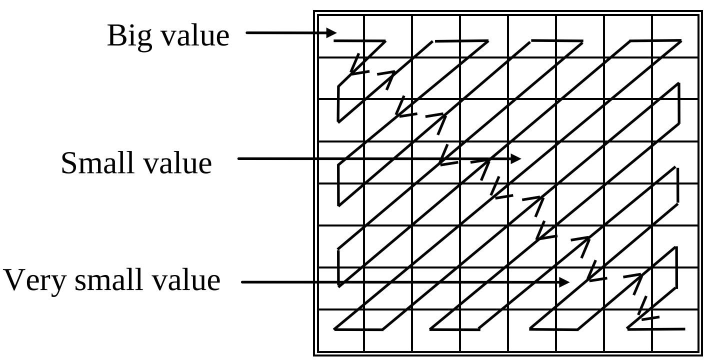
0	1	2					7
8	9						
16							
56							63

Raster order

0	1	5					
2	4						
3							
							63

Zig-Zag order

Figure (2) Raster and Zig-Zag Block Orders for 8∗8 Block Size



(8 ∗ 8) block size

Figure (3) Scan through diagonal path for the Zig-Zag sequence

The aim of the proposed algorithm is to obtain a good performance at high-speed [15]. There are many motivations for this kind of compression system. The basic motivation is the transform coefficients more independent, and most of the data energy is packed into the first few transform coefficients. The high frequency coefficients, which are less important to image content than the low frequency coefficients are discarded. The second motivation is the appearance of new algorithms for DCT that have fast speed, which in turn lead to divided the large transform calculation into a sequence of much simpler operations compared with the conventional algorithms. Therefore, the system will be more efficient and expensive. The hardware oriented motivation for the proposed system, is performed by a massively parallel pipelined processing unit feature of subimage transformation which lead to very fast system and the ability to make the linear mapping. **Figure (4)** shows the main steps for proposed retained DCT coefficients.

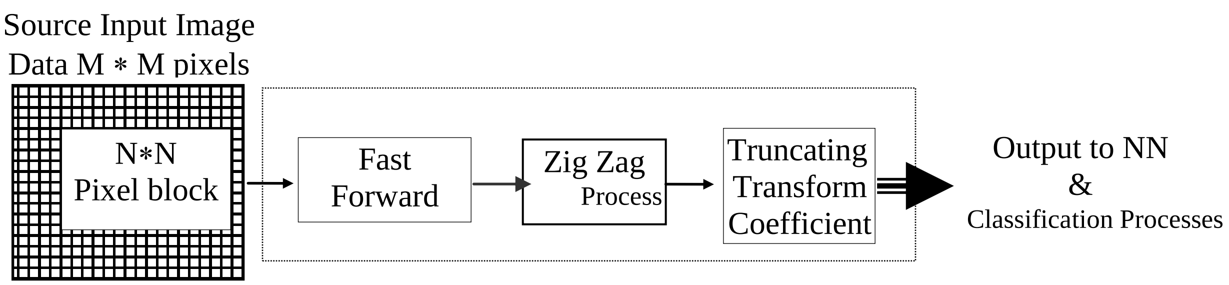


Figure (4) Block Diagram of the Main Steps for Retained DCT Coefficients

2-4 Fidelity Measures

Mostly, compression techniques cause some information losses up to a certain tolerated level. Thus a use of fidelity criteria is required to measure the amount of losses. There are two criteria to denote the fidelity [19]:

2-4-1 Objective Fidelity Criteria

It is required to measure the system quality. Different objective criteria have been considered in the literature, among the most commonly used are:

1. Root-Mean-Square (RMS) Error: Let the original and reconstructed images to be of M∗M array of pixels, Each pixel is an b-bit binary word corresponding to one of the 2^b possible gray level values, and denoted by f(x,y) and g(x,y) respectively, where x,y=0,1,...,M-1. The difference between the original image pixels and the corresponding reconstructed ones are as follows:

$$\mathbf{e\,(x,y) = g\,(x,y) - f\,(x,y)} \quad \ldots\ldots \quad \mathbf{(3)}$$

The RMS error is defined as:

$$E_{RMS} = [\frac{1}{M * M} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} [g(x, y) - f(x, y)]^2]^{1/2} \quad \text{.......(4)}$$

2. Signal-to-Noise Ratio (SNR): Considering the difference between the original and the reconstructed images as a noise, then the mean square value of (SNR) of the reconstructed image is defined as:

$$(SNR)_{ms} = [\sum_{x=0}^{M-1} \sum_{y=0}^{M-1} g^2(x, y) / \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \left(g(x, y) - f(x, y)\right)^2] \quad \text{.......(5)}$$

The $(SNR)_{ms}$ value in (dB) is given by:

$$\textbf{SNR (dB)} = 10 \log_{10} [(SNR)_{ms}] \quad \text{.......(6)}$$

To get an impression that a face can roughly be reconstructed by only few coefficients, several error measures are given to compare the difference between two images. The commonly used error measures are the mean square error (MSE) and signal-to-noise ratio.

2-4-2 Subjective Fidelity Criteria

The quality of the reconstructed image can be measured by a quantitative distortion measure (objective criteria). However, the most reliable judge of the image quality is the human eye, which does not necessarily agree with quantitative measures. When the output images are to be viewed by people, it is more appropriate to use a subjective fidelity criterion corresponding to how good the images look to human observers. The human visual system (HVS) has peculiar characteristics so that two images having the same amount of quantitative measures may appear to have drastically different visual qualities. Hence a human vision can be exploited to enhance the subjective quality of an image [19].

2-5 Effect of Block Size (Subimage Size Selection)

Transformation takes place over blocks rather than the whole image for two reasons. Firstly, the transform of small blocks is much easier and faster to processing and storing requirements than it is applied on the complete image. Secondly, the correlation between pixels is less, on average, between distant pixels than between close pixels and the law of diminishing returns takes a hand [20].

Despite that the more compression is achieved by choosing a large block size (increase the discarding of coefficients of zero values), the objective performance should be improved with increasing the block size even for smaller size say 8∗8, does not significantly increase the error. But the image compression here involves processing blocks of small size to avoid some overflow problems with a finite word machine and any distortion due to discarding gets distributed over the entire image blocks not only the complete image, during the inverse transformation, that is less objectionable.

In general, the subimage size should be an integer power of 2, in order to simplify the computation of the subimage transform. If the image dimensions do not divide by the subimage size, the image may be zero-padded to the next multiple of that.

2-6 Simulation Results of DCT Process

The advantage of DCT is that most transform coefficients on real world images turn out to be very small in magnitude. Truncating these small coefficients from the representation introduces only small errors in the reconstructed image. For some image faces, most of the information exists in the coefficients with small u and v (the upper-left corner). This characteristic simplifies optimum coefficient selection for image recognition applications. **Figure (5)** shows the effect of increasing the number of coefficients on reconstructed images.

Figure (6) and **(7)** illustrates graphically the impact of subimage size on reconstruction error. The data plotted were obtained by dividing the original image into four various types of subimages and then reconstructing the image using only 1/16 of the resulting coefficients.

The SNR is utilized as a measure of the degree of closeness between the original and its reconstructed companion images, although subjective performance is exercisable by visual inspection of image examples presented. The SNR can be improved and minimize the effect of blocking artifact when the boundaries between subimages become visible by increased the block size. It is noted by observation of the simulated results that the best choice on the dimension of the block size is 16∗16 pixel block, so that the reconstruction error reaches the optimum. It has been found that the convergence time required for the implementation of the compression system task of 16∗16 pixel block size is about 0.0003 second.

(a) **(b)** **(c)**

Figure (5) Effect of Increasing the Number of Coefficients on Reconstructed Images
(a) A 256*256 8-bit Original Face Image (b) and (c) The Reconstructed Images using only 1000 and 12500 of the Resulting DCT Coefficients of Original Image Respectively

(a) (b) (c) (d)

Figure (6) Illustration of the Effect of the Size of Subimages on Reconstructed Images
(a), (b), (c), and (d) the Reconstructed Images by Dividing Original Image into Subimages of Size 4*4, 8*8, 16*16, 32*32, Respectively and then Retaining 1/16 of the DCT Coefficients

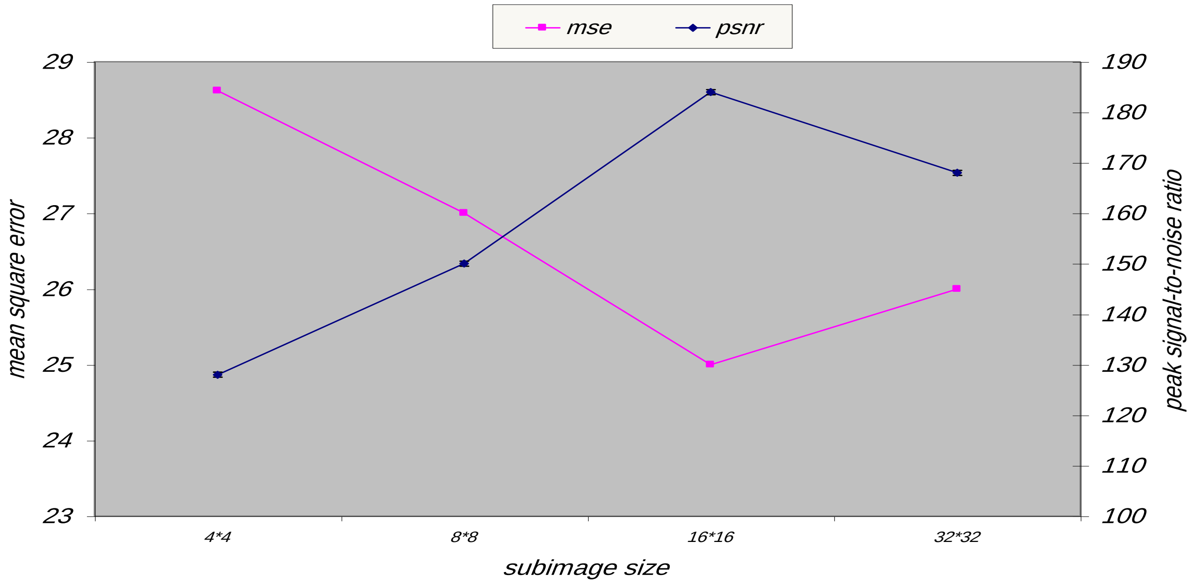


Figure (7) Reconstruction Error versus Subimage Size

3. Features of Neural Network

NN represents massively parallel-distributed processing capability with the potential for improving performance through dynamical learning. This approach is introduced in the recognition field by defining the recognition problems as a mapping between input space and an output space [21].

The characteristics and properties of NN with specific reference to NN in the face recognition field are:

1. Parallel Distributed Processing: NN has a highly parallel structure, which leads to use it in practical application.

2. Hardware Implementation: This is closely related to the preceding point. Networks can not only be implemented in parallel but it can be modified to dedicate on VLSI hardware implementation. This brings additional speed and increases the scale of network.

NN employ an algorithm to determine the similarity of the unique global features of live versus enrolled or reference faces, using as much of the facial modify the weight it gives to certain facial features. This method, theoretically, leads to an increased ability to identify faces in difficult conditions. As with all primary technologies, NN facial recognition can do 1-1 or 1-many. Therefore it is perhaps the most widely utilized face recognition technology [22].

Recognition systems are mainly compared on the basis of their recognition rate, training time and recognition time [23]. The recognition rate is the most important index of a recognition system. The smaller the training time, the more resources will be available to exploit other techniques to improve the performance, for example, instead of one multi-layer perception (MLP) one can apply ensemble or bootstrapping techniques to reduce the generalization error, or self-renovate the system after misclassification. The requirement of real time applications is the short recognition time. For example, one could not ask everybody to wait minutes to pass an access control system.

4. A Proposed Face Recognition System Description

The system used for face recognition is a combination of a high-level block diagram as shown in **Fig.(8)**, which is breakdown to various subsystems. The description for the practical implementation of the proposed system is given in below. The main idea of the approach is to apply the DCT to reduce the information redundancy and to use the packed information for classification. For a face image, the system first computes the DCT coefficients of the image or its subimages, then selects only a limited number of the coefficients and feeds them as input into a classifier. DCT computation and subimage divisions are performed in the manner described in the preceding sections.

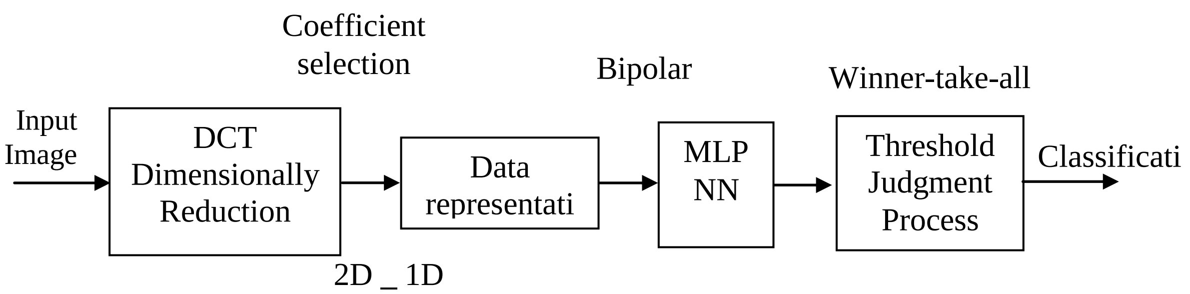


Figure (8) Schematic Diagram of the Main Steps of the Proposed Face Recognition System

4-1 Coefficient Selection

The coefficient allocation method is usually fixed in the system after determination and applied to all images/subimages. The location of the transform coefficients retained for each image/subimage remains unchanged from one image to another, similar to the zonal mask allocation method in image compression literature [24]. The threshold mask allocation method was examined experimentally, which selects the transform coefficients on the basis of magnitude instead of location. Although the transform coefficients of largest magnitude make the most significant contribution to the reconstructed image quality, their locations are not easy to record since they vary from one image to another. Therefore, threshold mask allocation requires the storage of one set of coefficient locations per image. As a result, for the same total number of coefficients stored, threshold masks allocation results in much poorer recognition performance than the method (approximately 27% worse in the experiments). The scanning strategy for converting the 2-D DCT array into a 1-D array (for feeding into the MLP) can affect the classification rate. These scanning strategies are all based on the fact that most of the image information is concentrated in the upper-left corner of the DCT.

4-2 Data Representation and Classification

After the number and locations of transform coefficients are selected, the selected coefficients are arranged in 1-D format and are fed into a classifier for recognition. The classifier used in the system is a feed forward NN. A quick backpropagation (BP) algorithm is used as the training algorithm. Learning is improved by representing the input and output in bipolar form.

The number of output nodes depends on the number of classes that are recognized. Since the recognition system is specified for 40 classes, therefore the total number of nodes for network structure is 41 nodes. The appended node for rejection unknown faces that does not belong to the 40 class. When it uses the coded output nodes of NN, the rejected images gives a specified code not similar to the coded one of any class of 40 classes that is known.

The disadvantage of NN is the probability of wrong classification for all the unknown images in the known classes. Therefore, to solve this problem the network is trained on the larger number of these rejected faces [16]. A large number of nonface or wrong (rejected) face images are needed to train the face recognizer, because the variety of nonface images is much greater than the variety of face images.

BP theory suggests that the training procedure can be speeded up if all input and output vectors are in the same range. However, DCT coefficients in different locations usually have different orders of magnitude. If they are converted using normalizing factor for all locations, some of them could be very small. This irregularity could make the training of the NN very hard.

5. The Database

The images are BMP standard format with the size of 256*256 pixels. Such image size compared to other sizes seems to provide a reasonable resolution. On the other side, another assumption is made that these images are 256 gray-scale levels [8-bit (2^8=256)]. According to the BMP image format digital image header and palette represent the first 1078 samples of the original BMP image (header (54) plus the palette (1024) byte).

The image databases were collected at featured differences in lighting conditions, facial expressions, and background:

1. The first database contains 40 distinct persons, each person having 10 different images, taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses) as shown in **Fig.(9)**. All the images are taken against a white homogeneous background and the persons are in up-right, frontal position (with tolerance for some side movement). Which obtained using a digital camera coupled to a 1000MHz PIII personal computer [16].

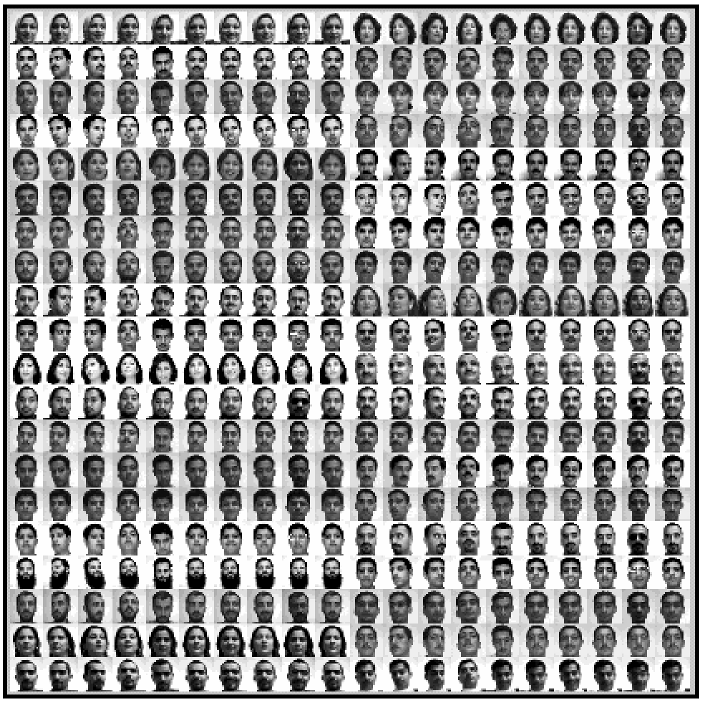


Figure (9) Example of Database Face Files

2. The second database used the Olivetti Research Ltd. (ORL) face database [10]. The database contains 400 images compiled of 40 subjects with 10 images each. The images were taken against a dark homogeneous background with varying lighting conditions. The faces are in a frontal upright position and show a range of expressions. Side movement and head tilt were tolerated to a limited extent only.

6. Simulations Results

6-1 Experimental Setup

To evaluate the recognition rate of any NN recognition system, the accuracy of the system can be calculated as:

$$\textbf{Recognition rate} = \frac{\textbf{Number of correctly classified patterns}}{\textbf{Total number of paterrns}}$$

In the following experiments, the weights and biases of the MLP are initialized to random values in [-0.05; 0.05]. The maximum number of training epochs is 3000 (stopping condition).

To reduce the influence of the presentation order of training samples, for every training loop, the training samples were shuffled once randomly. To allow comparisons, the same training and test set size are used, i.e., the first 5 images for each subject are the training images and the remaining 5 images are used for testing. Hence there are 200 training images and 200 test images in total and no overlap exists between the training and test images. Due to the small size of the available data, a validation set was not used and the best-so-far recognition rate on testing images is reported as the testing recognition rate.

To evaluate the effects of adding noise to a recognized image on the system performances **Fig.(10)** displays samples of recognized two images of one subject from the database. To evaluate the effects of different facial expressions on the system performances **Fig.(11)** displays samples of recognized two corrupted images of one subject from the database.

(a) **(b)**

Figure (10) Test Images with and without Noise
(a) Test Image without Noise (b) Test Image with Grain Noise
(SNR = 28 dB)

(a) **(b)** **(c)**

Figure (11) Sample Corrupted Faces used in the Experiment under Facial Expression Changes
(a) Training Image (b) Testing Image of Smiling Expression
(c) Testing Image of Surprised Expression

6-2 Effect of Hidden Neurons and Hidden Layers Variation

The simulation results for variation on the number of nodes for single hidden layer are shown in **Table (1)**. To reduce the classification error and exclude the mistaken acceptance of unknown faces, the threshold for the output varies to reduce the error.

Table (1) Accuracy of Face Recognition System with Varying Number of Nodes for the Single Hidden Layer

No. of nodes for single hidden layer	65	75	85
Recognition rate	90.8%	92%	91.2%

Hidden layers (hidden neurons) play a critical role in the operation of BP learning because they act as a feature detectors. As the learning process progresses, the hidden neurons in these layers begin to gradually discover the salient features that characterize the training data.

For most applications, a single hidden layer is sufficient. Sometimes, difficult learning tasks can be simplified by increasing the number of internal layers. So, for complex mapping, two hidden layers may give better generalization and may make training easier than a single hidden layer. The accuracy of the system as the number of nodes varies for the second hidden layer is shown in **Table (2)**. The size of the first hidden layer is fixed on 75 nodes. Referring to these results it is clearly noticed that best performance is obtained when using the network with two hidden layers (75 and 45 nodes respectively). **Table (3)** shows the accuracy of the system as the number of hidden layers in the network is varied.

Table (2) Accuracy of Face Recognition System with Varying Number of Nodes for the Second Hidden Layer in the Network Subnets Number of Nodes in First Hidden Layer=75

No. of nodes for second hidden layer	35	45	55
Recognition rate	91.5%	93.7%	92.6%

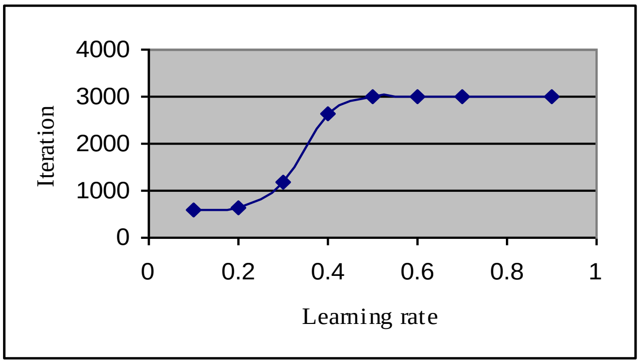
Table (3) Accuracy of Face Recognition System with Varying Number of Hidden Layers in the Network Subnets

No. of hidden layers	One hidden layer	Two hidden layers
Recognition rate	92%	93.7%

6-3 Effect of Learning Rate

It is desirable to use a small learning rate to avoid a major disruption of the direction of learning when an unusual pair of training patterns is presented. However, it is also preferable to maintain training at a fairly rapid pace as long as the training data are relatively similar.

When the learning rate is low, a network will adjust its weights gradually but in this case convergence may be slow. While with a high learning rate, the network can make a drastic change that is not desirable in a nearly trained network, but this is not a problem when starting from random weights. **Figure (12)** shows the effect of learning rate on maximum iteration needed and minimum error obtained.



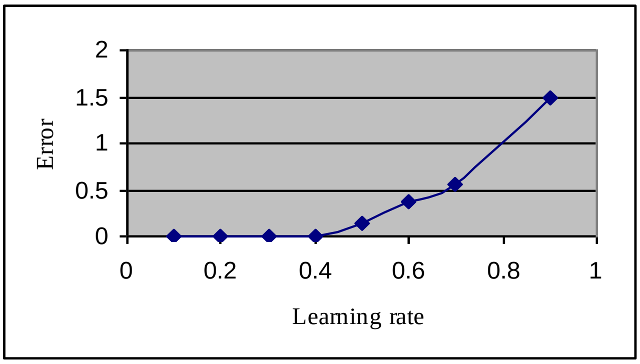


Figure (12) Effect of Learning Rate on Maximum Iteration Needed and Minimum Error Obtained

From the recognition performance on the testing images reconstructed from different numbers of DCT coefficients, it is demonstrated that the recognition rate decreases when more DCT coefficients are retained. The reader may at first imagine this effect to be caused by having too many or too few hidden neurons. However, the number of hidden neurons is sufficiently large that in all cases the training set can be fully learnt.

Furthermore, reducing the number of hidden neurons reduces the performance too. This confirms the idea that more specific pixel information introduced by using more DCT coefficients could decrease the recognition rate since hair, face outline; eyes and mouth have been determined to be the most important facial features for perceiving and remembering faces.

6-4 Comparison Based on ORL Database of Different Recognition Approaches

The ORL database has been used to test several face recognition approaches [11,25]. It is difficult to compare the speed of algorithms executed on different computing platforms because of the interactions of a large number of factors such as CPU speed, memory and cache size, compiler efficiency and even the programmer's skill. However, the training and classification times of the DCT-based method are much faster than the other approaches.

Note that the classification time for the DCT-based method is around 700 times faster than the convolutional NN approach [10]. The classification speed of the convolutional NN approach is itself about 200 times faster than Hidden Markov Models approach. The above speed comparison is conservative. Furthermore, for the above comparison, input images to the convolutional network were a quarter of full resolution.

In this case, the best average recognition rate is 93.7% obtained by retaining 1/16 DCT coefficients and using the network with two hidden layers (with 75 and 45 nodes respectively).

The complete recognition run takes less than 0.0002 second on a PC with 1000MHz CPU, while the training time need about 30 minute.

7. Conclusions

An efficient and very fast approach to face recognition is presented, which combines image compression and NN techniques together. The compression is achieved by applying a DCT to the face images and truncating the unimportant components. For face images, high frequency DCT components are negligibly small and can be truncated without loss of the most important facial features such as hair, eyes, and mouth outline and location. In this approach, the compressed coefficients are used for arbitrate NN classification. The experiments reported above demonstrate that for the ORL database, using only 1/16 of all the available DCT coefficients produces a recognition rate comparable to the best results reported to date while the processing speed is more than 3 orders of magnitude faster. Since the classification process is done on the transform coefficients, that lead to increase the computational load of the system, using a proposed approach for image compression technique can solve this disadvantage.

8. References

1. Cipolla, R., and Pentland, A., ***"Computer Vision for Human-Machine Interaction"***, Cambridge University Press 1998.

2. Turk, M., and Pentland, A., ***"Eigenfaces for Recognition"***, Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.

3. Chellappa, R., Wilson, C., and Sirohey, S., ***"Human and Machine Recognition of Faces: A Survey"***, Proceedings of the IEEE, Vol. 83, No. 5, 1995, pp. 705-740.

4. Zhang, J., Yan, Y., and Lades, M., ***"Face Recognition: Eigenface, Elastic Matching, and Neural Nets"***, Proceedings of the IEEE, Vol. 85, No. 9, 1997, pp. 1423-1435.

5. Brunelli, R., and Poggio, T., ***"Face Recognition: Features Versus Templates"***, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, 1993, pp. 1042-1052.

6. Gao, Y., and Leung, M., ***"Face Recognition using Line Edge Map"***, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6, June 2002.

7. Harmon, L., Goldstein, and Lesk, A., ***"Identification of Human Faces"***, Proceeding. IEEE, Vol. 59, 1971, pp. 748-760.

8. Samaria, F., and Fallside, F., ***"Face Identification and Feature Extraction using Hidden Markov Model"***, Image Processing Theory and Application, 1993.

9. Baluja, S., Rowley, and Kanade, T., ***"NN-Based Face Detection"***, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, 1998, pp. 23-38.

10. Lawrence, S., Lee C., Giles, Tsoi, A., and Back, A., ***"Face Recognition: A Convolutional NN Approach"***, IEEE Transactions on NNs, vol. 8, no. 1, 1997, pp. 98-113.

11. Samaria, F., ***"Face Recognition using Hidden Markov Models"***, Ph.D. Thesis, Trinity College, University of Cambridge, Cambridge, 1994.

12. Dawood, S., ***"Face Recognition using Neocognitron NN"***, M.Sc. Thesis, University of Mosul, 2000.

13. Nebauer, C., ***"Evaluation of Convolutional NNs for Visual Recognition"***, IEEE Transactions on NNs, Vol. 9, No. 4, 1998, pp. 685-696.

14. Bormans, J., Christopoulos, A. Skodras, and Cornelis, J., ***"Efficient Computation of the Two-Dimensional Fast Cosine Transform"***, In SPIE Hybrid Image and Signal Processing IV, (Orlando, Florida, USA), 1994, pp. 229-237.

15. Abdel-Mottaleb, M., and Shneier, ***"Exploiting the JPEG Compression Scheme for Image Retrieval"***, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, 1996, pp. 849-853.

16. AL-Fahdawi, K., ***"Face Recognition using NN"***, Ph.D. Thesis, University of Technology, 2004.

17. AL-Fahdawi, K., ***"High Performance Technique for Image Data Compression"***, M.Sc. Thesis, AL-Mustansiriya University, 1997.

18. Cho, N., and Lee, S., ***"A Fast 4*4 DCT Algorithm for the Recursive 2-D DCT"***, IEEE Trans. On Signal Processing, Vol. 40, No. 9, September 1992, pp. 2166-2174.

19. Gonzalez, R., and Woods, R., **"Digital Image Processing"**, Addison-Wesley, 1992.

20. Umbaugh, S. E., ***"Computer Vision and Image Processing"***, A Practical Approach using CVIP Tools, Prentice-Hall International, Inc., 1998.

21. Roweth, D. Forrest, Stroud, N., Wallace, D., and Wilson, G., ***"NN Models"***, Parallel Computing, Elsevier Science Publishers B. V. (North-Holland), 1988.

22. Fausett, L., ***"Fundamentals of NNs: Architectures, Algorithms and Applications"***, Prentice- Hall Inc., 1994.

23. Patel, R., ***"Face Detection using NN"***, http://www.ri.cmu.edu/pubs/pub_926.html.

24. Smith, S., ***"Digital Signal Processing"***, California Technical Publishing USA, 1999.

25. Jahren, C., Lindsey, C., Lindbland, T., and Osterud, K., ***"Eye Identification for Face Recognition with NNs"***, Pattern Recognition, 1996, pp.150-163.