Local-DCT features for Facial Recognition

Belinda Schwerin, Kuldip Paliwal.
Signal Processing Laboratory, Griffith University, Brisbane, AUSTRALIA.

Abstract

This paper presents the results of a project investigating the use of Discrete-Cosine transform (DCT) to represent facial images of an identity recognition system. The objective was to address the problem of sensitivity to variation in illumination, faced by commercially available systems. The proposed method uses local, block-wise DCT to extract the features of the image. Results show that this appearance based method gives high identification rates, improved tolerance to variation in illumination, and simple implementation.

I. Introduction

Biometric signatures, which characterize a person with respect to their body features, cannot be lost, stolen, or shared with others [1], unlike commonly used access cards and passwords. Consequently, there has been an increasing motivation to development identification systems based on these signatures. Applications for such systems include person identification for use by law enforcement agencies [1], facilitation of human-machine interactions in robotics, and identification for access to secure areas or sensitive information. While a range of commercial systems are currently available, they are documented to demonstrate poor performance under extreme variations in environment [2], and therefore there is a need for further development of these systems. Additionally, for applications such as smart cards for ATMs, training needs to be able to be done over the life of the system. Therefore independence from the larger database is preferable.

Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are baseline methods often used for comparison. They are based on the Karhunen-Loeve Transform (KLT), which was first used for representing faces by Kirby and Sirovich (1990). In PCA[3], Turk and Pentland reduced the calculation complexity of their approach by only selecting the most significant coefficients to represent the face. In LDA[4], the Fisher's linear discriminant was applied to the feature vector space found via KLT to find the most discriminating faces (rather than the most expressive ones). The KLT minimizes the mean squared error for data compression, and has the most information in the fewest number of transform coefficients, however it is sensitive to variations in face pose, position, and illumination. It also has the added disadvantage that its basis set is data-dependent, with determination of this set requiring knowledge of the complete reference database. As a result, training is also computationally expensive.

As an alternative to KLT for representing the feature space, the Discrete-Cosine Transform (DCT) has found popularity due to its comparative concentration of information in a small number of coefficients, and increased tolerance to variation of illumination. The DCT-based approach offers other advantages over KLT-based methods including improved processing time, potential dimensionality reduction and compatibility with data already encoded via DCT coefficients such as jpeg compressed digital video and images. Unlike KLT-based methods, the DCT basis vectors are data independent, making it much more suitable for systems where the reference database is dynamic over the life of the system.

A number of different implementations using DCTs to represent facial images have been developed in recent years. In Hafed and Levine [5] the DCT of the entire, normalized image is extracted. A subset of the most significant coefficients were retained as the feature vector, with nearest-neighbour classification then used to identify the face. Using the first 64 coefficients, which have low to mid range frequencies and the most variance, a very high recognition rate was reported.

The 2-D DCT method described in Sanderson [6] uses a block-by-block approach. The image is broken into 8x8 pixel blocks, overlapping by 50%. The first 15 2D DCT coefficients (ordered in a zigzag pattern –see Figure 1) were found and combined at the decision level. Variations also proposed in Sanderson [6] included discarding of the first three DCT coefficients used to represent each block; and the use of delta functions to replace the first three, illumination sensitive coefficients in each block.

In an alternative method, Ekenel and Stiefelhagen [7] used DCTs for localized appearance-based feature representations, where the detection of face features such as eyes, etc. is not required. In their method, the image was broken into 8x8 pixel non-overlapping blocks, which were then represented by their DCT coefficients. The first coefficient (DC coefficient) was discarded and of the remaining coefficients, only the most significant were kept.

In this paper, results from a project implementing a facial recognition system using local appearance DCT feature extraction are presented. It is shown that the approach used provides improved tolerance to variation in illumination, with a good identification rate using nearest-neighbour classification. In section II, the proposed local-DCT feature extraction method is described. Section III shows results of testing, and finally in section IV, conclusions and recommendations for future work are given.

978-1-4244-4242-3/08/$25.00 ©2008 IEEE

II. METHODS

Discrete-Cosine Transform

The main features of the DCT which make it attractive for face recognition are:

1. data compression (energy compaction) property, commonly used in image/video compression, and
2. efficient computation, due to its relationship to the Fourier transform.

The DCT was first introduced by Ahmed, Natarajan, and Rao in 1974, with categorization into 4 variations by Wang (1984). This invertible transform represents a finite signal as a sum of sinusoids of varying magnitudes and frequencies. For images, the 2-dimensional DCT is used, with the most visually significant information being concentrated in the first few DCT coefficients (ordered in a zigzag pattern from the top left corner - see Figure 1).

0	1	5	6	14
2	4	7	13	15
3	8	12	16	21
9	11	17	20	22
10	18	19	23	24

Figure 1: Labelling of DCT coefficients of 5x5 image block.

Information concentration in the top left corner is due to the correlation between image and DCT properties. For images, the most visually significant frequencies are of low-mid range, with higher frequencies giving finer details of the image. For the 2-D DCT, coefficients correspond to sinusoids of frequencies increasing from left to right and top to bottom. Therefore those in the upper corner are the coefficients associated with the lowest frequencies. The first coefficient, known as the DC coefficient, represents the average intensity of the block, and is the most affected by variation in illumination.

For the DCT, any image block $A_{m,n}$, of size MxN, can be written as the sum of the MN functions of the form:

$$\alpha_p \alpha_q \cos\left(\frac{\pi(2m+1)p}{2M}\right)\cos\left(\frac{\pi(2n+1)q}{2N}\right),$$

$$\text{for } 0 \le \mathrm{p} \le \mathrm{M}-1 \quad , \quad 0 \le q \le N-1 .$$

These are the basis functions of the DCT.

The DCT coefficients $B_{p,q}$ are the weights applied to each basis function for reconstruction of the block. The coefficients are calculated as:

$$B_{p.q} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\left(\frac{\pi(2m+1)p}{2M}\right)\cos\left(\frac{\pi(2n+1)q}{2N}\right)$$

for $0 \le p \le M\text{-}1$, $0 \le q \le N\text{-}1$,

where $$\alpha_p = \begin{cases} \sqrt{\frac{1}{M}} & , \quad p = 0 \\ \sqrt{\frac{2}{M}} & , \quad 1 \le p \le M\text{-}1 \end{cases},$$

$$\alpha_q = \begin{cases} \sqrt{\frac{1}{N}} & , \quad q = 0 \\ \sqrt{\frac{2}{N}} & , \quad 1 \le q \le N\text{-}1 \end{cases},$$

and A_{mn} is the MxN input matrix.

The result is MN 2-D DCT coefficients.

Proposed Local-DCT method:

The proposed feature extraction method uses a local appearance based approach. This provides a number of advantages. Firstly, the use of local features allows spatial information about the image to be retained. Secondly, for an appropriately selected block size, illumination affecting the block can be assumed quasi-stationary. Consequently, the resulting recognition system can be designed for improved performance in environments where variation in illumination occurs across images.

The local-DCT method operates on normalized images where faces are aligned (e.g. by eyes) and of the same size. To extract the features of the image, blocks of size bxb are used. Neighbouring blocks overlap by $(b\text{-}1)\times(b\text{-}1)$ pixels, scanning the image from left to right and top to bottom. The number of blocks processed is then:

$$(M-(b-1))\,(N-(b-1)) .$$

For each block, the 2-dimensional DCT is then found and c_f coefficients are retained. The coefficients found for each block are then concatenated to construct the feature vector representing the image. For c_f retained coefficients and square image, the dimensionality of this feature vector is then:

$$d = (N-(b-1))^2 \times c_f$$

where d = feature vector dimensionality
NxN = original image resolution
(D = original image dimensionality = N^2),
bxb = block size, and
c_f = number of coefficients retained from each block.

The retained coefficients from each block are those which contain the most significant information about the image. Since the first coefficient is the most affected by variation in illumination, it is excluded. Coefficient selection, choice of block size, and number of coefficients retained is discussed in part III.

A brief description of other methods used for comparison is given below.

Whole Image method

The image intensities themselves can be used to construct a feature vector by concatenating the columns of the image.

The resulting vector has dimensionality MN for image resolution MxN.

PCA Eigenfaces

For a reference database of R images, each of size NxN, the columns of each image are concatenated to form column vectors $\vec{f}_i$, for i=1,2,…,R. The average face $\vec{f}_m$ is then calculated, and subtracted from each of the training faces in the database to give zero-mean vectors $\vec{g}_i = \vec{f}_i - \vec{f}_m$, for i = 1,2,…,R. The mean training set is then represented as the matrix G = [$\vec{g}_1\ \vec{g}_2 \ldots \vec{g}_R$], and the covariance matrix is then $C = G\,G^T$. For reduced computation, the eigenvectors of C can be equivalently found by multiplying G by the eigenvectors of $G^T G$. For a feature space with d dimensions, the d most significant eigenvectors (with the largest eigenvalues) are selected and describes the feature space $U = [\vec{u}_1\ \vec{u}_2 \ldots \vec{u}_d]$, where $\vec{u}_i$ are the eigenfaces. Finally, the zero-mean vectors of each face are projected onto this feature space to find their feature vectors $\vec{\beta}_i = U^T \vec{g}_i$, i = 1,2,…,R.

LDA Fisherfaces

For a database of R images, comprising P different people, and image dimensionality of N^2, the LDA method requires that R ≥ N^2 for the within-class scatter matrix to be singular. For smaller databases, PCA must first be applied to reduce the dimensionality of each face to R, then the LDA method can be applied as follows.

For a database of faces represented in column vectors $\vec{f}_i$, for i=1,2,…,R, the average of all faces $\vec{f}_m$, and the average face for each person $\vec{p}_j$, j = 1,2,…,P, training faces are represented as the difference from their average face $\vec{h}_i = \vec{f}_i - \vec{p}_j$. The scatter matrices for each person (or class) can then be constructed as $S_j = \sum \vec{h}_i\,\vec{h}_i^{\,T}$, for each image vector $\vec{h}_i$ of person j. The within-class scatter matrix S_W and the between-class scatter matrix S_B can then be found:

$$S_W = \sum_j S_j \ , \text{ for } j = 1,2,\ldots,P.$$

$$S_B = \sum_j 2\,(\vec{p}_j - \vec{f}_m)\,(\vec{p}_j - \vec{f}_m)^T \ , \text{ for } j = 1,2,..,P.$$

The matrix W which maximizes $\left|W^T S_B W\right| / \left|W^T S_W W\right|$ has columns which are the eigenvectors of $S_W^{-1} S_B$. Since the columns of W are eigenvectors satisfying : $S_B w_i = \lambda_i S_w w_i$, the eigenvalues λ_i are found by solving $\left|S_B - \lambda_i S_w\right| = 0$ and the eigenvectors are found by solving $(S_B - \lambda_i S_w) w_i = 0$. Matrix W then defines the LDA-space. The feature vector for each person in this space is then found by projecting faces onto it.

III. Experiments & Discussion

The Purdue AR Face database [8], contains 100 individuals, each with 14 images (excluding those with occlusions such as sunglasses and scarves) of various illuminations and facial expressions. For each person, 7 of these are used for training and 7 for testing. The original images have a resolution of 165x120 pixels. Before testing, each has its illumination normalized via histogram equalization, and is then resized to an asymmetrical resolution of 80x80 pixels (or other required testing size). Figure 2 shows a sample image set from the database.

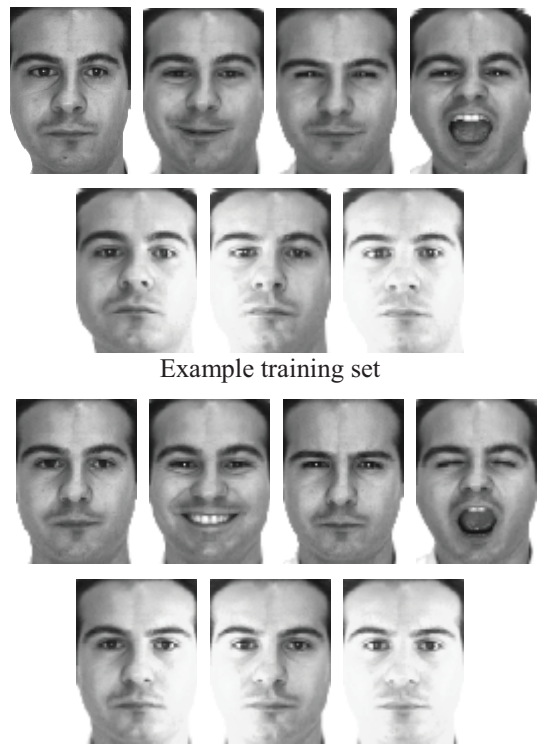

Figure 2: Example image set of one person in the AR database.

Experiments performed on this database implement the identified feature extraction method, then nearest-neighbour classification with the Euclidean distance measure to identify the class of each test image. The identification rate, which gives the percentage of test images correctly identified, is then calculated.

Coefficients of Significance

The local-DCT method uses a selection of the 2-dimensional DCT coefficients from each block to construct the feature vector for an image. For the best identification rates, the coefficients selected should be the most significant in representing important and distinguishing image details. Using one coefficient to represent each block, all images of the database were represented as

feature vectors, and then the recognition rates were determined. Table 1 shows the results for each possible choice of DCT coefficient using a block size of 10x10 and image size of 80x80 pixels.

Coefficients in the upper left corner are confirmed to be the most significant, as indicated by their higher recognition rates. These correspond to the low frequency coefficients of the transform in horizontal and vertical directions.

Table 1 also shows a lower recognition rate for coefficient B(0,1) (coefficient 1 in figure 1) compared with B(1,0) (coefficient 2). The difference observed here is attributed to the reshaping of the original image to make it asymmetrical. Further investigation of this difference will be considered in future work.

p \ q	0	1	2	3	4	5	6	7	8	9
0	70.9	75.1	69.9	60.3	47.7	36.1	25.4	15.7	11.0	7.1
1	85.0	81.3	72.3	56.4	43.9	31.6	19.9	11.9	8.3	3.9
2	76.6	79.3	73.1	58.0	43.0	29.1	18.6	11.9	8.1	4.6
3	63.4	71.0	66.4	50.6	37.1	25.7	16.6	9.3	6.6	4.4
4	53.0	59.7	58.9	44.4	30.4	22.3	12.7	7.7	4.7	3.3
5	48.4	51.3	45.0	34.1	22.6	15.3	9.7	6.0	4.3	2.7
6	40.4	39.4	37.3	24.9	18.3	12.1	6.9	4.7	3.9	1.9
7	32.9	31.3	28.1	18.9	12.6	8.7	5.1	3.6	3.0	1.1
8	24.3	23.6	21.0	12.3	10.1	5.7	3.1	2.1	1.3	1.9
9	11.6	14.9	12.1	9.4	7.4	4.4	2.3	1.6	1.6	1.7

Table 1: Identification rates (%) using the DCT coefficient B(p,q) to represent each block in the feature vector. Results shown correspond to a block size of 10x10 pixels, and image dimensionality D=6400.

Block size

Each block is a subset of the image, with the number of identity distinguishing details (such as edges) contained in each block, related to the fraction of the image it includes. Where block size is too small (compared to the whole image), the likelihood of it containing details of significance is reduced. If too large, it may contain too many details which are smoothed by use of low frequency DCT basis functions.

To determine the best block size for use with an 80x80 pixel image, identification rates where again determined using one DCT coefficient to represent each block. Rates were found over a range of block sizes. As shown by Figure 3(a), a block size of 10x10 pixels gives the best recognition rate.

Figure 3(b) shows identification rates where experiments were repeated using different image sizes. Again we find better performance for an 80x80 image size when using a 10x10 pixel block size. Similar results occur for other image sizes, as shown in Figure 3(c).

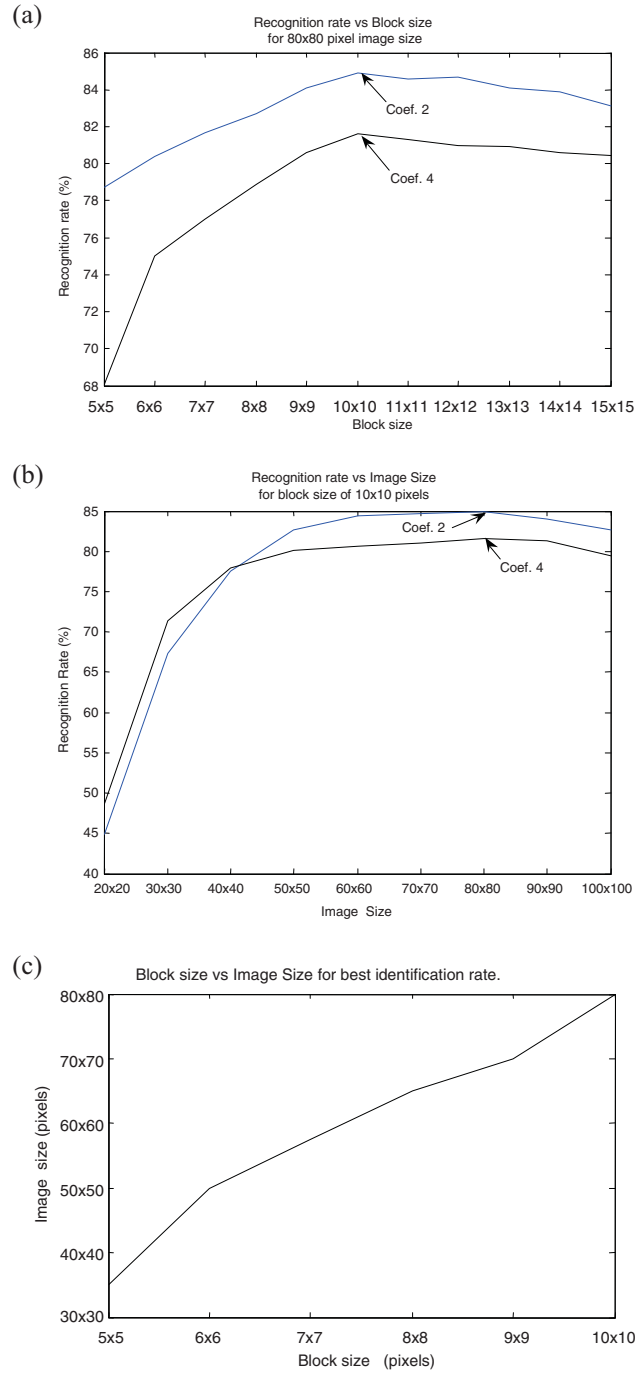

Figure 3: (a) Identification rates (%) using images resized to 80x80 pixels, for different block sizes; (b) Identification rates using a block size of 10x10 pixels, for different image sizes; (c) Image size giving the highest identification rates for each block size.

Multiple coefficients

By use of additional coefficients to represent each pixel, more details are retained. Again using 80x80 pixel images, with a 10x10 pixel block size, feature vectors for each image were found using two, three and four coefficients to

represent each block. Table 2 (original images) shows a selection of the results recorded.

Using one coefficient to represent each block, coefficient 2 gave the best identification rate of 85%. Two coefficient combinations which included coefficient 2, generally showed an increase in identification rate to around 86%. Three coefficient combinations which included coefficients 2 and 4 showed identification rates as high as 87.4%.

Results in table 2 also show that the use of more than three coefficients actually reduces the identification rate, with percentages consistent with those achieved using just two coefficients, with the achievable identification rate being limited by the significance of the coefficients retained.

1 coefficient / block:

Coef. Number	0	1	2	3	4	5	6	7	8
Original Images	70.9	75.1	85.0	76.6	81.3	69.9	60.3	72.3	79.3
Illum.added δ=80	31.6	60.1	70.1	62.3	70.3	58.6	44.1	55.7	68.7

2 coefficients / block:

Coef. Number	0,1	0,2	1,2	1,3	1,4	2,3	2,4	2,5	2,8
Original Images	73.9	75.0	85.1	81.0	79.0	82.7	86.1	86.4	86.3
Illum.added δ=80	40.3	44.0	78.3	76.6	70.3	70.6	74.4	76.9	73.7

3 coefficients / block:

Coef. Number	0,2,4	1,2,3	1,2,4	1,2,5	1,3,4	2,3,4	2,4,5	2,4,8
Original Images	76.6	84.1	85.9	84.9	82.0	84.4	87.4	86.3
Illum.added δ=80	47.7	78.9	79.6	79.6	78.3	74.3	76.7	75.7

4 coefficients / block:

Coef. Number	1,2,4,5	1,2,4,6	1,2,4,8
Original Images	85.3	85.9	85.9
Illum.added δ=80	80.0	80.3	80.0

Table 2: Identification rates (%) using the labelled combinations of DCT coefficients to represent each block in the feature vector. Rates where variation in illumination is added to test images (δ=80) are also shown. Results shown are a selection of those for an image size of 80x80 pixels (dimensionality D=6400), and 10x10 pixel block size.

Effect of variation of illumination

To simulate the problems associated with testing in different environments, illumination (varying across the image) was added to the test images of the AR database. This was done by changing the intensity of the image's pixels proportionally to their column displacement from the image centre. An example of the effect of this transform (for different δ) on an image is shown in Figure 4.

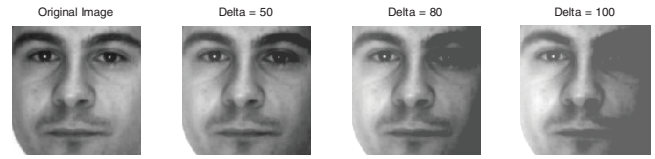

Figure 4 : Images where illumination across the image has been varied.

For variation of illumination from left to right [6]:

New Image (y,x) = Old Image (y,x) + m x + δ

where $m = \frac{-\delta}{(N_X - 1)/2}$,

$x = 0,1,\ldots,N_X - 1$,

and $y = 0,1,\ldots,N_Y - 1$.

Identification rates using different DCT coefficients (and coefficient combinations) to represent each block for delta factor δ=80 are shown in Table 2. Table 3 shows the effect of increasing delta on identification rates using the indicated coefficients to represent each block.

Level of Illum. variation added	δ=0	δ=20	δ=40	δ=60	δ=80	δ=100
Using coef. 2	85.0	83.6	82.3	79.0	70.1	52.4
Using coef. 1,2	85.1	85.1	85.1	82.7	78.3	66.4
Using coef. 2,4	85.9	85.7	84.3	80.9	74.3	57.6
Using coef. 1,2,4	85.9	85.9	85.1	82.9	79.6	68.9

Table 3: Identification rates (%) using the indicated coefficients and level of illumination variation added to test images (δ). An image size of 80x80 pixels and 10x10 pixel block size was used.

These results show that use of the DC coefficient 0 is unreliable for practical implementations, with significantly reduced performance for testing of non-uniformly illuminated images. While the identification rate found using the coefficient combination (2,4,5) was the highest using the original images, once varying illumination was added across the test images, performance dropped to 76.7%. Coefficient combinations (1,2) or (1,2,4), on the other hand, showed much less sensitivity to illumination variation, with the three coefficient combination still having a 79.6% recognition rate.

Comparison to other methods

The local-DCT method was then compared to baseline methods, including whole image (where image intensities

form the feature vector), PCA Eigenfaces [3] and LDA Fisherface [4] methods. Nearest-neighbour classification with the Euclidean distance measure was again used to determine the recognition rate of each method, both with and without illumination variation across the test images.

Feature Extraction method :	Whole Image (d=6400)	PCA (d=99)	LDA (d=600)	Local-DCT Coef:1,2 (d=9940)	Local-DCT Coef:1,2,4 (d=14910)
For D=6400, δ=0:	78.0%	73.7%	82.0%	85.1%	85.9%
For D=6400, δ=80:	63.0%	58.6%	67.7%	78.3%	79.6%

Table 4: Recognition rates using different methods for the same AR database for image size of 80x80 pixels. Local-DCT results shown use a 10x10 pixel block size and the indicated coefficients. Testing of the sensitivity to variation of illumination is indicated by δ=80.

From the results of Table 4, the local-DCT method shows a higher identification rate than other methods tested. In particular, there is a significant improvement where test images are affected by illumination variation, with an increase of more than 10% over LDA and 20% over PCA methods.

Improvements are acknowledged to be at the expense of feature vector size. While this method increases the memory requirements, this is inexpensive for most current computing systems. Implementation is also very simple compared to other methods. The method also offers the advantage of data independence, with changes to a database (such as the addition of new persons), having no affect on the basis vectors used for feature extraction. As opposed to this, PCA and LDA methods require recomputation of basis vectors then recalculation of features across the entire database, as and when any new person is added to it. As a result, the cost of larger feature length is offset by other implementation and performance advantages.

IV. Conclusions

The local-DCT method presented showed identification rates of up to 85.0% using only 1 coefficient, compared to 82% for LDA on the same database. Those coefficients which were most significant in representing the details of an image block were identified as the low frequency coefficients 1, 2, and 4. Using multiple significant coefficients, the detection rate could be increased to over 86% on a 80x80 pixel image. Evaluation of this feature extraction method also showed a relationship between the size of the block used and the size of the image for the best recognition rate. For an image size of 80x80 pixels, a 10x10 was identified to give the best rate. While this method does not reduce the dimensionality of the resulting feature vector, it is shown to be much less sensitive to variation in illumination across images than other methods tested.

Future work will compare the local-DCT method presented here to other DCT-based approaches, using a number of facial image databases.

V. References

[1] Kar, S. and Hiremath, S., et al, "A Multi-Algorithmic Face Recognition System", IEEE, p 321-326, 2006.

[2] Ling, Yangrong, et al, "Sirface vs Fisherface : Recognition Using Class Specific Linear Projection", IEEE, p III-885 – III-888, 2003.

[3] Turk, M., & Pentland A., 1991, "Face Recognition Using Eigenfaces", IEEE p586-591.

[4] Belhumeur, P., etal, 1997, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 19, No 7, 1997.

[5] Hafed, Z., and Levine, M., 2001, "Face Recognition using the Discrete Cosine Transform", International Journal of Computer Vision, 43(3), p167-188.

[6] Sanderson, C., 2002, "Automatic Person Verification Using Speech & Face Information", Dissertation presented to School of Microelectronic Engineering, Griffith University.

[7] Ekenel, H.K,. and Stiefelhagen, R., "Local appearance based face recognition using discrete cosine transform", University of Karlsruhe, Germany.

[8] Martinez, A.M. & Benavente, R., Purdue Robot Vision Lab, "The AR Face Database", CVC Technical Report #24, June 1998, http://cobweb.ecn.purdue.edu/RVL/database.htm .