Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition Letters 24 (2003) 2409–2419

Pattern Recognition Letters

www.elsevier.com/locate/patrec

# Fast features for face authentication under illumination direction changes

Conrad Sanderson [a,b,*], Kuldip K. Paliwal [b]

[a] *IDIAP, Rue du Simplon 4, CH-1920 Martigny, Switzerland*
[b] *School of Microelectronic Engineering, Griffith University, Nathan, Brisbane, Queensland 4111, Australia*

Received 7 February 2002; received in revised form 20 March 2003

**Abstract**

In this letter we propose a facial feature extraction technique which utilizes polynomial coefficients derived from 2D Discrete Cosine Transform (DCT) coefficients obtained from horizontally and vertically neighbouring blocks. Face authentication results on the VidTIMIT database suggest that the proposed feature set is superior (in terms of robustness to illumination changes and discrimination ability) to features extracted using four popular methods: Principal Component Analysis (PCA), PCA with histogram equalization pre-processing, 2D DCT and 2D Gabor wavelets; the results also suggest that histogram equalization pre-processing increases the error rate and offers no help against illumination changes. Moreover, the proposed feature set is over 80 times faster to compute than features based on Gabor wavelets. Further experiments on the Weizmann database also show that the proposed approach is more robust than 2D Gabor wavelets and 2D DCT coefficients.


*Keywords:* Face authentication; Illumination changes; Polynomial coefficients; Gabor wavelets; Discrete cosine transform; Eigenfaces; Histogram equalization

## 1. Introduction

The field of face recognition can be divided into two areas: face identification and face verification (also known as authentication). A face verification system verifies the claimed identity based on images (or a video sequence) of the claimant's face; this is in contrast to an identification system, which attempts to find the identity of a given person out of a pool of $N$ people.

Verification systems pervade our every day life; for example, Automatic Teller Machines (ATMs) employ simple identity verification where the user is asked to enter their password (known only to the user), after inserting their ATM card; if the password matches the one prescribed to the card, the user is allowed access to their bank account. However, the verification system such as the one used in the ATM only verifies the validity of the

* Corresponding author. Address: School of Microelectronic Engineering, Griffith University, Nathan, Brisbane, Queensland 4111, Australia. Tel.: +41-27-721-7743/+61-73-875-6578; fax: +41-27-721-7712/+61-73-875-5198.
*E-mail address:* conradsand@ieee.org (C. Sanderson).

0167-8655/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0167-8655(03)00070-9

combination of a certain possession (in this case, the ATM card) and certain knowledge (the password). The ATM card can be lost or stolen, and the password can be compromised (e.g. somebody looks over your shoulder while you're keying it in). In order to address this issue, biometric verification methods have emerged where the password can be either replaced by, or used in addition to, biometrics such as the person's speech, face image or fingerprints. More information about the field of biometrics can be found in papers by Bolle et al. (2002), Dugelay et al. (2002) and Woodward (1997).

Generally speaking, a full face recognition system can be thought of as being comprised of three stages:

1. Face localization and segmentation
2. Normalization
3. The actual face identification/verification, which can be further subdivided into:
   (a) Feature extraction
   (b) Classification

The second stage (normalization) usually involves an affine transformation (Gonzales and Woods, 1993) (to correct for size and rotation), but it can also involve an illumination normalization (however, illumination normalization may not be necessary if the feature extraction method is robust against varying illumination). In this letter we shall concentrate on the feature extraction part of the last stage.

There are many approaches to face based systems, ranging from the ubiquitous Principal Component Analysis (PCA) approach (also known as eigenfaces) (Turk and Pentland, 1991), Dynamic Link Architecture (also known as elastic graph matching) (Duc et al., 1999), Artificial Neural Networks (Lawrence et al., 1997), to pseudo-2D Hidden Markov Models (HMM) (Samaria, 1994; Eickeler et al., 2000). Recent surveys on face recognition can be found in papers by Chellappa et al. (1995), Zhang et al. (1997) and Grudin (2000).

The above-mentioned systems differ in terms of the feature extraction procedure and/or the classification technique used. For example, Turk and Pentland (1991) used PCA for feature extraction and a nearest neighbour classifier for recognition. Duc et al. (1999) used biologically inspired 2D Gabor wavelets (Lee, 1996) for feature extraction, while employing the Dynamic Link Architecture as part of the classifier. Eickeler et al. (2000) obtained features using the 2D Discrete Cosine Transform (DCT) and used the pseudo-2D HMM as the classifier.

PCA derived features have been shown to be sensitive to changes in the illumination direction (Belhumeur et al., 1997) causing rapid degradation in verification performance. A study by Zhang et al. (1997) has shown a system employing 2D Gabor wavelet derived features to be robust to moderate changes in the illumination direction; however, Adini et al. (1997) showed that the 2D Gabor wavelet derived features are sensitive to gross changes in the illumination direction.

Belhumeur et al. (1997) proposed robust features based on Fisher's Linear Discriminant; however, to achieve robustness, the system required face images with varying illumination for training purposes.

As will be shown, 2D DCT based features are also sensitive to changes in the illumination direction. In this letter we introduce four new techniques, which are significantly less affected by an illumination direction change: DCT-delta, DCT-mod, DCT-mod-delta and DCT-mod2. We will show that the DCT-mod2 method, which utilizes polynomial coefficients derived from 2D DCT coefficients of spatially neighbouring blocks, is the most suitable. We then compare the robustness and performance of the DCT-mod2 method against three popular feature extraction techniques: eigenfaces (PCA), PCA with histogram equalization and 2D Gabor wavelets.

The rest of the letter is organized as follows. In Section 2 we briefly review the 2D DCT feature extraction technique and describe the proposed feature extraction methods which build from the 2D DCT. In Section 3 we describe a Gaussian Mixture Model (GMM) based classifier which shall be used as the basis for experiments. The performance of the traditional and proposed feature extraction techniques is compared in Section 4, using an artificial illumination direction change.

Section 5 is devoted to experiments on the Weizmann database (Adini et al., 1997) which has more realistic illumination direction changes.

To keep consistency with traditional matrix notation, pixel locations (and image sizes) are described using the row(s) first, followed by the column(s).

## 2. Feature extraction

### 2.1. 2D discrete cosine transform (DCT)

Here the given face image is analyzed on a block by block basis. Given an image block $f(y,x)$, where $y,x = 0,1,\ldots,N_P-1$ (here we use $N_P = 8$), we decompose it in terms of orthogonal 2D DCT basis functions (see Fig. 1). The result is an $N_P \times N_P$ matrix $C(v,u)$ containing 2D DCT coefficients:

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

for $v,u = 0,1,2,\ldots,N_P-1$, where

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{N_P}} & \text{for } v = 0 \\ \sqrt{\frac{2}{N_P}} & \text{for } v = 1,2,\ldots,N_P-1 \end{cases} \quad (2)$$

and

$$\beta(y,x,v,u) = \cos\left[\frac{(2y+1)v\pi}{2N_P}\right]\cos\left[\frac{(2x+1)u\pi}{2N_P}\right] \quad (3)$$

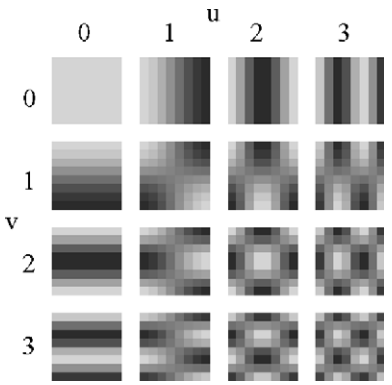


Fig. 1. Several 2D DCT basis functions for $N_P = 8$ (lighter colours represent larger values).

| v \ u | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 1 | 5 | 6 |
| 1 | 2 | 4 | 7 | 12 |
| 2 | 3 | 8 | 11 | 13 |
| 3 | 9 | 10 | 14 | 15 |

Fig. 2. Ordering of 2D DCT coefficients $C(v,u)$ for $N_P = 4$.

The coefficients are ordered according to a zig–zag pattern, reflecting the amount of information stored (Gonzales and Woods, 1993) (see Fig. 2). For block located at $(b,a)$, the 2D DCT feature vector is composed of:

$$\vec{x} = \left[c_0^{(b,a)} c_1^{(b,a)} \cdots c_{M-1}^{(b,a)}\right]^{\mathrm{T}} \quad (4)$$

where $c_n^{(b,a)}$ denotes the $n$th 2D DCT coefficient and $M$ is the number of retained coefficients. To ensure adequate representation of the image, each block overlaps its horizontally and vertically neighbouring blocks by 50% (Eickeler et al., 2000). Thus for an image which has $N_Y$ rows and $N_X$ columns, there are $N_D = (2(N_Y/N_P)-1)\times(2(N_X/N_P)-1)$ blocks. [1]

### 2.2. DCT-delta

In speech based systems, features based on polynomial coefficients (also known as deltas), representing transitional spectral information, have been successfully used to reduce the effects of background noise and channel mismatch (Soong and Rosenberg, 1988).

For images, we define the $n$th *horizontal* delta coefficient for block located at $(b,a)$ as a modified 1st order orthogonal polynomial coefficient (Johnson and Leone, 1977; Soong and Rosenberg, 1988):

$$\Delta^h c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k c_n^{(b,a+k)}}{\sum_{k=-K}^{K} h_k k^2} \quad (5)$$

[1] Thus for a $56 \times 64$ (rows $\times$ columns) image, there are 195 2D DCT feature vectors.

Similarly, we define the $n$th *vertical* delta coefficient as:

$$\Delta^v c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k c_n^{(b+k,a)}}{\sum_{k=-K}^{K} h_k k^2} \quad (6)$$

where $\vec{h}$ is a $2K+1$ dimensional symmetric window vector. In this letter we shall use $K=1$ and a rectangular window (thus $\vec{h} = [1.0\ \ 1.0\ \ 1.0]^{\mathrm{T}}$).

Let us assume that we have three horizontally consecutive blocks $X$, $Y$ and $Z$. Each block is composed of two components: facial information and additive noise; e.g. $X = I_X + I_N$. Moreover, let us also suppose that all of the blocks are corrupted with the same noise (a reasonable assumption if the blocks are small and close or overlapping). To find the deltas for block $Y$, we apply Eq. (5) to obtain (ignoring the denominator):

$$\Delta^h Y = -X + Z \quad (7)$$

$$= -(I_X + I_N) + (I_Z + I_N) \quad (8)$$

$$= I_Z - I_X \quad (9)$$

i.e. the noise component is removed.

By combining the horizontal and vertical delta coefficients an overall delta feature vector is formed. Hence, given that we extract $M$ 2D DCT coefficients from each block, the delta vector is $2M$ dimensional. We shall term this feature extraction method as DCT-delta.

DCT-delta feature extraction for a given block is only possible when the block has vertical and horizontal neighbours; thus processing an image which has $N_Y$ rows and $N_X$ columns and using a 50% block overlap results in $N_{D2} = (2(N_Y/N_P) - 3) \times (2(N_X/N_P) - 3)$ DCT-delta feature vectors. [2]

### 2.3. DCT-mod, DCT-mod2 and DCT-mod-delta

By inspecting Eqs. (1) and (3), it is evident that the 0th DCT coefficient will reflect the average pixel value (or the DC level) inside each block and hence will be the most affected by any illumination change. Moreover, by inspecting Fig. 1 it is evident that the first and second coefficients represent the average horizontal and vertical pixel intensity change, respectively. As such, they will also be significantly affected by any illumination change. Hence we shall study three additional feature extraction approaches (in all cases we assume the baseline 2D DCT feature vector is $M$ dimensional):

1. Discard the first three coefficients from the baseline 2D DCT feature vector. We shall term this *modified* feature extraction method as DCT-mod.
2. Discard the first three coefficients from the baseline 2D DCT feature vector and concatenate the resulting vector with the corresponding DCT-delta feature vector. We shall refer to this method as DCT-mod-delta.
3. Replace the first three coefficients with their horizontal and vertical deltas and form a feature vector representing a given block as follows:

$$\vec{x} = [\Delta^h c_0\ \Delta^v c_0\ \Delta^h c_1\ \Delta^v c_1\ \Delta^h c_2\ \Delta^v c_2\ c_3\ c_4\ \cdots\ c_{M-1}]^{\mathrm{T}} \quad (10)$$

where the $(b,a)$ superscript was omitted for clarity. Let us term this approach as DCT-mod2.

As for DCT-delta, DCT-mod-delta and DCT-mod2 feature extraction for a given block is only possible when the block has vertical and horizontal neighbours; thus processing an image which has $N_Y$ rows and $N_X$ columns and using a 50% block overlap results in $N_{D2} = (2(N_Y/N_P) - 3) \times (2(N_X/N_P) - 3)$ DCT-mod-delta or DCT-mod2 feature vectors. [3]

## 3. GMM based classifier

Given a claim for person $C$'s identity and a set of feature vectors $X = \{\vec{x}_i\}_{i=1}^{N_Y}$ supporting the claim, the average log likelihood of the claimant being the true claimant is calculated using:

[2] Thus for a $56 \times 64$ image, there are 143 DCT-delta feature vectors.

[3] Thus for a $56 \times 64$ image, there are 143 DCT-mod-delta or DCT-mod2 feature vectors.

$$\mathscr{L}(X|\lambda_C) = \frac{1}{N_V}\sum_{i=1}^{N_V} \log p(\vec{x}_i|\lambda_C) \quad (11)$$

$$\text{where } p(\vec{x}|\lambda) = \sum_{j=1}^{N_G} m_j \mathscr{N}(\vec{x}; \vec{\mu}_j, \boldsymbol{\Sigma}_j) \quad (12)$$

$$\lambda = \{m_j, \vec{\mu}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{N_G} \quad (13)$$

Here, $\mathscr{N}(\vec{x}; \vec{\mu}_j, \boldsymbol{\Sigma})$ is a $D$-dimensional Gaussian function with mean $\vec{\mu}$ and diagonal covariance matrix $\boldsymbol{\Sigma}$:

$$\mathscr{N}(\vec{x}; \vec{\mu}_j, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left[\frac{-1}{2}(\vec{x}-\vec{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right] \quad (14)$$

$\lambda_C$ is the parameter set for person $C$, $N_G$ is the number of Gaussians and $m_j$ is the weight for Gaussian $j$ (with constraints $\sum_{j=1}^{N_G} m_j = 1$ and $\forall j : m_j \geqslant 0$).

Given the average log likelihood of the claimant being an impostor, $\mathscr{L}(X|\lambda_{\overline{C}})$, an opinion on the claim is found using:

$$\Lambda(X) = \mathscr{L}(X|\lambda_C) - \mathscr{L}(X|\lambda_{\overline{C}}) \quad (15)$$

The verification decision is reached as follows: given a threshold $t$, the claim is accepted when $\Lambda(X) \geqslant t$ and rejected when $\Lambda(X) < t$.

### 3.1. Model training and impostor likelihood

Given a set of training vectors, $X = \{\vec{x}_i\}_{i=1}^{N_v}$ (which may come from several images), the GMM parameters ($\lambda$) for each client model are found by the Expectation Maximization (EM) algorithm (Dempster et al., 1977; Moon, 1996; Duda et al., 2001).

The likelihood of the claimant being an impostor can be found via the use of a composite model, [4] comprised of several GMMs for other clients. The client models in such a composite are referred to as background models (Reynolds, 1995) or cohort models (Furui, 1997). Given $B$ background models, the impostor likelihood is found using:

$$\mathscr{L}(X|\lambda_{\overline{C}}) = \log\left[\frac{1}{B}\sum_{b=1}^{B} \exp \mathscr{L}(X|\lambda_b)\right] \quad (16)$$

The background model set contains models which are the "closest" as well as the "farthest" from the client model (Reynolds, 1995). While it may intuitively seem that only the "close" models are required (which represent the expected impostors), this would leave the system vulnerable to impostors which are very different from the client. This is demonstrated by inspecting Eq. (15) where both terms would contain similar likelihoods, leading to an unreliable opinion on the claim.

In this letter we have utilized the method described by Reynolds (1995) to select the background models for each client.

## 4. Experiments

### 4.1. VidTIMIT audio-visual database

The VidTIMIT database (Sanderson, 2002), is comprised of video and corresponding audio recordings of 43 people (19 female and 24 male), reciting short sentences. It was recorded in 3 sessions, with a mean delay of 7 days between Session 1 and 2, and 6 days between Sessions 2 and 3. There are 10 sentences per person; the first six sentences are assigned to Session 1; the next two sentences are assigned to Session 2 with the remaining two to Session 3. The mean duration of each sentence is 4.25 s, or approximately 106 video frames.

The video of each person is stored as a sequence of high quality JPEG images with a resolution of $384 \times 512$ pixels. The corresponding audio is stored as a mono, 16 bit, 32 kHz WAV file.

### 4.2. Experimental setup

Before feature extraction can occur, the face must first be located (Chen et al., 2001). Furthermore, to account for varying distances to the camera, a geometrical normalization must be

[4] It must be noted that the Universal Background Model (Reynolds et al., 2000) can also be used to find $\mathscr{L}(X|\lambda_{\overline{C}})$.

performed. We treat the problem of face location and normalization as separate from feature extraction.

To find the face, we use template matching with several prototype faces [5] of varying dimensions. Using the distance between the eyes as a size measure, an affine transformation is used (Gonzales and Woods, 1993) to adjust the size of the image, resulting in the distance between the eyes to be the same for each person. Finally a $N_Y \times N_X$ ($N_Y = 56$, $N_X = 64$) pixel face window, $w(y,x)$, containing the eyes and the nose (the most invariant face area to changes in the expression and hair style) is extracted from the image.

For PCA, the dimensionality of the face window is reduced to 40 (choice based on Samaria (1994) and Belhumeur et al. (1997)).

For 2D DCT and 2D DCT derived methods, each block is $8 \times 8$ pixels. Moreover, each block overlaps with horizontally and vertically adjacent blocks by 50%.

For 2D Gabor features, we follow Duc et al. (1999) where the dimensionality of the 2D Gabor feature vectors is 18. The location of the wavelet centers was chosen to be as close as possible to the centers of the blocks used in DCT-mod2 feature extraction.

In our experiments, we use a sequence of images (video) from the VidTIMIT database for person verification. If the sequence has $N_I$ images, then $N_V = N_I$ for PCA derived features, $N_V = N_I N_D$ for 2D DCT and DCT-mod features and $N_V = N_I N_{D2}$ for DCT-delta, DCT-mod-delta, DCT-mod2 and 2D Gabor features.

To reduce the computational burden during modeling and testing, every second video frame was used. For each feature extraction method, client models with $N_G = 8$ (choice based on preliminary experiments) were generated from features extracted from face windows in Session 1. Sessions 2 and 3 were used for testing. Thus for each person an average of 318 frames were used for training and 212 for testing.

Ignoring any edges created by shadows, the main effect of an illumination direction change is that one part of the face is brighter than the rest. [6] Taking this into account, an artificial illumination change was introduced to face windows extracted from Sessions 2 and 3; to simulate more illumination on the left side of the face and less on the right, a new face window $v(y,x)$ is created by transforming $w(y,x)$ using: [7]

$$v(y,x) = w(y,x) + mx + \delta \quad (17)$$
$$\text{for } y = 0,1,\ldots,N_Y - 1, \quad x = 0,1,\ldots,N_X - 1$$
$$\text{where } m = \frac{-\delta}{(N_X - 1)/2},$$
$$\delta = \text{illumination delta (in pixels)}$$

Example face windows for various $\delta$ are shown in Fig. 3. It must be noted that this model of illumination direction change is artificial and restrictive as it does not cover all the effects possible in real life (shadows, [8] etc.), but it is useful for providing suggestive results.

To find the performance, Sessions 2 and 3 were used for obtaining example opinions of known impostor and true claims. Four utterances, each from 8 fixed persons (4 male and 4 female), were used for simulating impostor accesses against the remaining 35 persons. As per Reynolds (1995), 10 background person models were used for the impostor likelihood calculation. For each of the remaining 35 persons, their four utterances were

[5] A "mother" prototype face was constructed by averaging manually extracted and size normalized faces from all people in the VidTIMIT database; prototype faces of various sizes were constructed by applying an affine transform to the "mother" prototype face.

[6] As evidenced by the images presented by Kotropoulos et al. (2000), which were obtained under real life conditions.

[7] Please note that many authors (for example, Weiss, 2001; Forsyth and Ponce, 2003) describe light changes as a multiplicative effect on image brightness. In our experiments we have treated the face image simply as an information source. The transformation described by Eq. (17) is in effect an empirical information transformation method; it has been designed to transform the face information to approximate the face images presented by Kotropoulos et al. (2000).

[8] However, the face images presented by Belhumeur et al. (1997) show that only extreme illumination direction conditions produce significant shadows, where even humans have trouble recognizing faces.

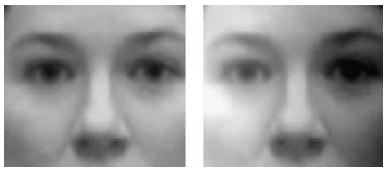

Fig. 3. Examples of varying light illumination; left: $\delta = 0$ (no change), right: $\delta = 80$.

used separately as true claims. In total there were 1120 impostor and 140 true claims. The decision threshold was then set so the a posteriori performance was as close as possible to the Equal Error Rate (EER) (i.e. where the False Acceptance rate (FA%) is equal to the False Rejection rate (FR%)). This protocol is described in more detail in (Sanderson, 2002).

In the first experiment, we found the performance of the 2D DCT approach on face windows with $\delta = 0$ (i.e. no illumination change) while varying the dimensionality of the feature vectors. The results are presented in Fig. 4, where it can be observed that the performance improves immensely as the number of dimensions is increased from 1 to 3. Increasing the dimensionality from 15 to 21 provides only a relatively small improvement, while significantly increasing the amount of computation time required to generate the models. Based on this we have chosen 15 as the dimensionality of baseline 2D DCT feature vectors; hence the dimensionality of DCT-delta feature vectors is 30, DCT-mod is 12, DCT-mod-delta is 42 and DCT-mod2 is 18.

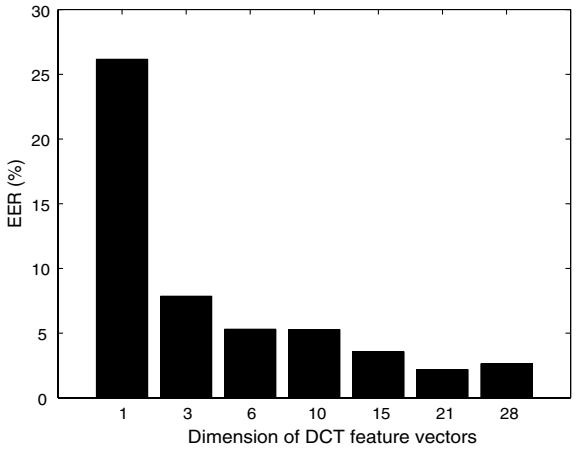


Fig. 4. Performance for varying dimensionality of 2D DCT feature vectors.

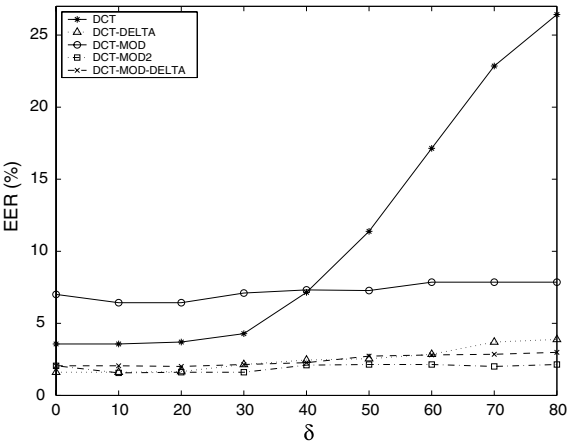


Fig. 5. Performance of 2D DCT and proposed feature extraction techniques.

In the second experiment we compared the performance of 2D DCT and all of the proposed techniques for increasing $\delta$; results are shown in Fig. 5.

In the third experiment we compared the performance of PCA, PCA with histogram equalization pre-processing, [9] 2D DCT, Gabor and DCT-mod2 features for varying $\delta$; results are presented in Fig. 6.

In the fourth experiment, we have evaluated the effects of varying block overlap used during DCT-mod2 feature extraction (in all other experiments, the overlap was fixed at 50%). Varying the overlap has two effects: the first is that as overlap is increased the spatial area used to derive one feature vector is decreased; the second effect is that the number of feature vectors extracted from an image grows in an exponential manner as the overlap is increased. Results are shown in Fig. 7.

[9] Histogram equalization (Castleman, 1996; Gonzales and Woods, 1993) is often used in an attempt to reduce the effects of varying illumination conditions (Koh et al., 2002; Moon and Phillips, 2001).

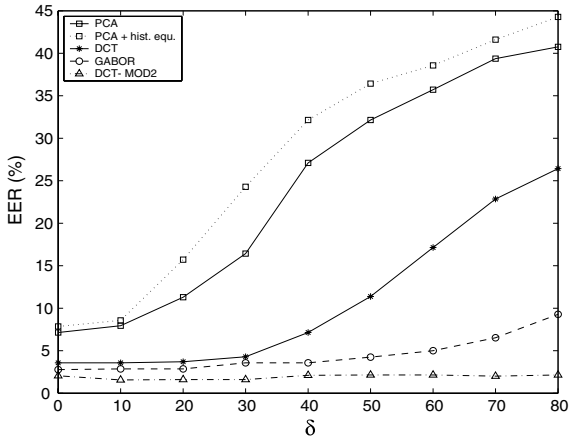


Fig. 6. Performance of PCA, PCA with histogram equalization pre-processing, 2D DCT, Gabor and DCT-mod2 feature sets.

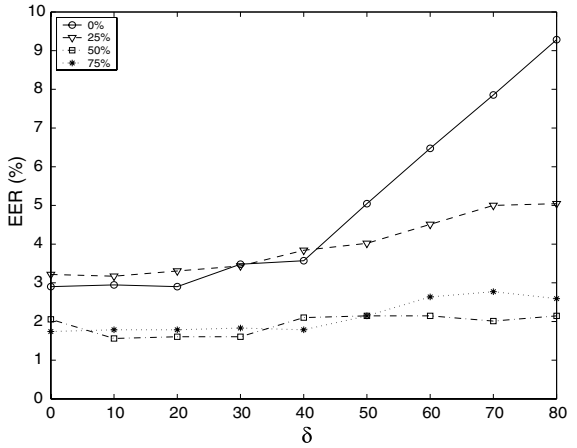


Fig. 7. Performance of DCT-mod2 for varying overlap.

Computational burden is an important factor in practical applications, where the amount of required memory and speed of the processor have direct bearing on the final cost. Hence in the final experiment we compared the average time taken to process one face window by PCA, 2D DCT, 2D Gabor and DCT-mod2 feature extraction techniques. It must be noted that apart from having the transformation data pre-calculated (e.g. $\beta$ 2D DCT basis functions), no thorough hand optimization of the code was done. Nevertheless, we feel that this experiment provides figures which are at least indicative. Results are listed in Table 1.

Table 1
Average time taken per face window (results obtained using Pentium III 500 MHz, Linux 2.2.18, gcc 2.96)

| Method | Time (msec) |
|---|---|
| PCA | 11 |
| 2D DCT | 6 |
| 2D Gabor | 675 |
| DCT-mod2 | 8 |

### 4.3. Discussion

As can be observed in Fig. 4, the first three 2D DCT coefficients contain a significant amount of person dependent information; thus ignoring them (as in DCT-mod) implies a reduction in performance. This is verified in Fig. 5 where the DCT-mod features have worse performance than 2D DCT features when there is little or no illumination direction change ($\delta \leqslant 30$). We can also see that the performance of DCT features is fairly stable for small illumination direction changes but rapidly degrades for $\delta \geqslant 40$ (in contrast to DCT-mod features which have a relatively static performance).

The remaining feature sets (DCT-delta, DCT-mod-delta and DCT-mod2) do not have the performance penalty associated with the DCT-mod feature set. Moreover, all of them have similarly better performance than 2D DCT features; we conjecture that the increase in performance can be attributed to the effectively larger spatial area used when obtaining the features. DCT-mod2 edges out DCT-delta and DCT-mod-delta in terms of stability for large illumination direction changes ($\delta \geqslant 50$). Additionally, the dimensionality of DCT-mod2 (18) is lower than DCT-delta (30) and DCT-mod-delta (42).

The results suggest that delta features make the system more robust as well as improve performance; they also suggest that it is only necessary to use deltas of coefficients representing the average pixel intensity and low frequency features (i.e. the 0th, first and second 2D DCT coefficients) while keeping the remaining DCT coefficients unchanged; hence out of the four proposed feature extraction techniques, the DCT-mod2 approach is the most suitable.

Using 0% or 25% block overlap in DCT-mod2 feature extraction (Fig. 7) results in a performance

degradation as $\delta$ is increased, implying that the assumption that the blocks are corrupted with the same noise has been violated (see Section 2.2). Increasing the overlap from 50% to 75% had little effect on the performance at the expense of extracting significantly more feature vectors.

By comparing the performance of PCA, PCA with histogram equalization pre-processing, 2D DCT, 2D Gabor and DCT-mod2 feature sets (Fig. 6), it can be seen that the DCT-mod2 approach is the most immune to illumination direction changes (the performance is virtually flat for varying $\delta$). The performance of PCA derived features rapidly degrades as $\delta$ increases, while the performance of 2D Gabor features is stable for $\delta \leqslant 40$ and then gently deteriorates as $\delta$ increases. We can also see that use of histogram equalization as pre-processing for PCA increases the error rate in all cases, and most notably offers no help against illumination changes. The results thus suggest that we can order the feature sets, based on their robustness and performance, as follows: DCT-mod2, 2D Gabor, 2D DCT, PCA, and lastly, PCA with histogram equalization pre-processing.

From Table 1 we can see that 2D Gabor features are the most computationally expensive to calculate, taking about 84 times longer than DCT-mod2 features. This is due to the size of the 2D Gabor wavelets as well as the need to compute both real and imaginary inner products. Compared to 2D Gabor features, PCA, 2D DCT and DCT-mod2 features take a relatively similar amount of time to process one face window.

It must be noted that when using the GMM classifier in conjunction with the 2D Gabor, 2D DCT or DCT-mod2 features, the spatial relation between major face features (e.g. eyes and nose) is lost. However, excellent performance is still obtained, implying that the use of more complex classifiers which preserve spatial relation, such as a pseudo-2D HMM (Eickeler et al., 2000) and elastic graph matching (Duc et al., 1999), is not necessary. Moreover, due to the loss of the spatial relations, the GMM classifier theoretically has some inbuilt robustness to translation (which may be caused by inaccurate face localization).

It must also be noted that using the introduced illumination change, the center portion of the face (column wise) is largely unaffected; the size of the portion decreases as $\delta$ increases. In the PCA approach one feature vector describes the entire face, hence any change to the face would alter the features obtained. This is in contrast to the other approaches (2D Gabor, 2D DCT and DCT-mod2), where one feature vector describes only a small part of the face. Thus a significant percentage (dependent on $\delta$) of the feature vectors is largely unchanged, automatically leading to a degree of robustness.

## 5. Experiments on the Weizmann database

The experiments described in Section 4 utilized an artificial illumination direction change. In this section we shall compare the performance of 2D DCT, 2D Gabor and DCT-mod2 feature sets on the Weizmann database (Adini et al., 1997), which has more realistic illumination direction changes.

It must be noted that the database is rather small, as it is comprised of images of 27 people; moreover, for the direct frontal view, there is only one image per person with uniform illumination (the training image) and two test images where the illumination is either from the left or right; all three images were taken in the same session. As such, the database is not suited for verification experiments, but some suggestive results can still be obtained.

The experimental setup is similar to that described in Section 4. However, due to the small amount of training data, an alternative GMM training strategy is used. Rather than training the client models directly using the EM algorithm, each model is derived from a Universal Background Model (UBM) by means of maximum a posteriori (MAP) adaptation (Gauvain and Lee, 1994; Reynolds et al., 2000). The UBM is trained via the EM algorithm using pooled training data from all clients. Moreover, due to the small number of persons in the database, the UBM is also used to calculate the impostor likelihood (rather than using a set of background models). A detailed description of this training and testing strategy is presented by Reynolds et al. (2000).

Table 2
Results on the Weizmann database, quoted in terms of HTER

| Feature type | Illumination direction | | |
|---|---|---|---|
| | Uniform | Left | Right |
| 2D DCT | 3.49 | 48.15 | 48.15 |
| 2D Gabor | 0.36 | 33.34 | 33.34 |
| DCT-mod2 | 0 | 25.93 | 22.65 |

Since PCA based feature extraction produces one feature vector per image, there is insufficient training data to reliably train the client models; thus PCA based feature extraction is not evaluated in this section.

For each illumination type, the client's own training image was used to simulate a true claim. Images from all other people were used to simulate impostor claims. In total, for each illumination type, there were 27 true claims and 702 impostor claims. The a posteriori decision threshold was set to obtain performance as close as possible to EER. Results are presented in Table 2, in terms of Half Total Error Rate (HTER), defined as $\mathrm{HTER} = (\mathrm{FA}\% + \mathrm{FR}\%)/2$.

As can be observed, no method is immune to the changes in the illumination direction. However DCT-mod2 features are the least affected, followed by 2D Gabor features and lastly 2D DCT features.

## 6. Conclusion

In this letter we proposed four new facial feature extraction techniques, which are resistant the effects of illumination direction changes; out of the proposed methods, the DCT-mod2 technique, which utilizes polynomial coefficients derived from 2D DCT coefficients of spatially neighbouring blocks, is the most suitable. Face verification results on the VidTIMIT database suggest that the DCT-mod2 feature set is superior (in terms of robustness to illumination direction changes and discrimination ability) to features extracted using four popular methods: eigenfaces (PCA), PCA with histogram equalization pre-processing, 2D DCT and 2D Gabor wavelets; the results also suggest that histogram equalization pre-processing increases the error rate and offers no help against illumination changes. Moreover, the DCT-mod2 feature set is over 80 times faster to compute than features based on Gabor wavelets. Further experiments on the Weizmann database have also shown that the DCT-mod2 approach is more robust than 2D Gabor wavelets and 2D DCT coefficients.

## References

Adini, Y., Moses, Y., Ullman, S., 1997. Face recognition: The problem of compensating for changes in illumination direction. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 721–732.

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 711–720.

Bolle, R.M., Connell, J.H., Ratha, N.K., 2002. Biometric perils and patches. Pattern Recognit. 35 (12), 2727–2738.

Castleman, K.R., 1996. Digital Image Processing. Prentice-Hall, USA.

Chellappa, R., Wilson, C.L., Sirohey, S., 1995. Human and machine recognition of faces: a survey. Proc. IEEE 83 (5), 705–740.

Chen, L.-F., Liao, H.-Y., Lin, J.-C., Han, C.-C., 2001. Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof. Pattern Recognit. 34 (7), 1393–1403.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39 (1), 1–38.

Duc, B., Fischer, S., Bigün, J., 1999. Face authentication with Gabor information on deformable graphs. IEEE Trans. Image Process. 8 (4), 504–516.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. John Wiley & Sons, USA.

Dugelay, J.-L., Junqua, J.-C., Kotropoulos, C., Kuhn, R., Perronnin, F., Pitas, I., 2002. Recent advances in biometric person authentication. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. IV, Orlando, pp. 4060–4062.

Eickeler, S., Müller, S., Rigoll, G., 2000. Recognition of JPEG compressed face images based on statistical methods. Image Vision Comput. 18 (4), 279–287.

Forsyth, D.A., Ponce, J., 2003. Computer Vision: A Modern Approach. Prentice Hall, USA.

Furui, S., 1997. Recent advances in speaker recognition. Pattern Recognition Lett. 18 (9), 859–872.

Gauvain, J.-L., Lee, C.-H., 1994. Maximum a posteriori estimation for multivariate Gaussian Mixture observations of Markov chains. Proc. IEEE Trans. Speech Audio Process. 2 (2), 291–298.

Gonzales, R.C., Woods, R.E., 1993. Digital Image Processing. Addison-Wesley, Reading, Massachusetts.

Grudin, M.A., 2000. On internal representations in face recognition systems. Pattern Recognit. 33 (7), 1161–1177.

Johnson, N.L., Leone, F.C., 1977. Statistics and Experimental Design in Engineering and the Physical Sciences, vol. 1. John Wiley & Sons, USA.

Koh, L.H., Ranganath, S., Venkatesh, Y.V., 2002. An integrated automatic face detection and recognition system. Pattern Recognit. 35 (6), 1259–1273.

Kotropoulos, C., Tefas, A., Pitas, I., 2000. Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. Pattern Recognit. 33 (12), 1935–1947.

Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D., 1997. Face recognition: A convolutional neural-network approach. IEEE Trans. Neural Networks 8 (1), 98–113.

Lee, T.S., 1996. Image representation using 2D Gabor wavelets. IEEE Trans. Pattern Anal. Machine Intell. 18 (10), 959–971.

Moon, H., Phillips, P.J., 2001. Computational and performance aspects of PCA-based face-recognition algorithms. Perception 30, 303–321.

Moon, T.K., 1996. Expectation–maximization Algorithm. IEEE Signal Process. Mag. 13 (6), 47–60.

Reynolds, D.A., 1995. Speaker identification and verification using Gaussian Mixture Speaker Models. Speech Commun. 17 (1–2), 91–108.

Reynolds, D., Quatieri, T., Dunn, R., 2000. Speaker verification using adapted gaussian mixture models. Digital Signal Process. 10 (1–3), 19–41.

Samaria, F., 1994. Face Recognition Using Hidden Markov Models. PhD Thesis, University of Cambridge.

Sanderson, C., 2002. The VidTIMIT Database. IDIAP Communication 02-06, Martigny, Switzerland (see www.idiap.ch).

Soong, F.K., Rosenberg, A.E., 1988. On the use of instantaneous and transitional spectral information in speaker recognition. IEEE Trans. Acoustics, Speech Signal Process. 36 (6), 871–879.

Turk, M., Pentland, A., 1991. Eigenfaces for Recognition. J. Cognitive Neurosci. 3 (1), 71–86.

Weiss, Y., 2001. Deriving intrinsic images from image sequences. In: Proceedings of 8th IEEE International Conference on Computer Vision, Vancouver.

Woodward, J.D., 1997. Biometrics: Privacy's foe or privacy's friend? Proc. IEEE 85 (9), 1480–1492.

Zhang, J., Yan, Y., Lades, M., 1997. Face recognition: Eigenface, elastic matching, and neural nets. Proc. IEEE 85 (9), 1422–1435.