ELSEVIER

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

Signal Processing 83 (2003) 931–940

SIGNAL
PROCESSING

www.elsevier.com/locate/sigpro

Features for robust face-based identity verification

Conrad Sanderson, Kuldip K. Paliwal*

School of Microelectronic Engineering, Griffith University, Brisbane, Queensland 4111, Australia

Received 2 November 2001; received in revised form 2 July 2002

Abstract

In this paper we propose the discrete cosine transform (DCT) mod 2 feature set, which utilizes polynomial coefficients derived from 2D DCT coefficients obtained from spatially neighboring blocks. Face verification results on the multi-session VidTIMIT database suggest that the DCT-mod 2 feature set is superior (in terms of robustness to illumination direction changes and discrimination ability) to features extracted using three popular methods: eigenfaces principal component analysis, 2D DCT and 2D Gabor wavelets. Moreover, compared to Gabor wavelets, the DCT-mod 2 feature set is over 80 times faster to compute. Additional experiments on the Weizmann database also show that the DCT-mod 2 approach is more robust than 2D Gabor wavelets and 2D DCT coefficients.

Keywords: Face verification; Polynomial coefficients; Discrete cosine transform; Gabor wavelets; Principal component analysis

1. Introduction

Recently, there has been a lot of interest in biometric person verification systems [2,7,29]. A biometric verification (or authentication) system verifies the identity of a claimant based on the person's physical attributes, such as their voice, face or fingerprints. Apart from security applications (e.g., access control), verification systems are also useful in forensic work (where the task is whether a given biometric sample belongs to a given suspect), and law enforcement applications [2,29].

A face verification system verifies the claimed identity (a 2 class task) 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. Past research on face-based systems has concentrated on the identification aspect even though the verification task has the greatest application potential [7].

While identification and verification systems share feature extraction techniques and in many cases a large part of the classifier structure, there is no guarantee that an approach used in the identification scenario would work equally well in the verification scenario.

There are many approaches to face-based systems ranging from the ubiquitous principal component analysis (PCA) approach (also known as eigenfaces) [27], dynamic link architecture (also known as elastic graph matching) [8,16,17], artificial neural networks [18], pseudo-2D hidden Markov models (HMM) [10,23] to simulated retinal vision [25]. Recent reviews of the face recognition area can be found in articles by Chellappa et al. [4], Grudin [14] and Zhang et al. [30].

* Corresponding author. Tel.: +61-7-3875-6578; fax: +61-7-3875-5198.

E-mail addresses: conradsand@ieee.org (C. Sanderson), k.paliwal@me.gu.edu.au (K.K. Paliwal).

0165-1684/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0165-1684(02)00497-8

The above systems differ in terms of the feature extraction procedure and/or the classification technique used. For example, in [27] PCA is used for feature extraction and a nearest neighbor classifier is utilized for recognition. In [8], biologically inspired 2D Gabor wavelets [19] are used for feature extraction, while the elastic graph matching procedure is part of the classifier. Kotropoulos et al. [16] also uses the elastic graph matching procedure, but the feature vectors are comprised of the outputs of multiscale morphological dilation and erosion operations [13]. 2D Gabor wavelets are also used in [25], where the classification strategy employs several support vector machine classifiers [9,28]. In [10], features are derived using the 2D discrete cosine transform (DCT) and the pseudo-2D HMM is the classifier.

PCA-derived features have been shown to be sensitive to changes in the illumination direction [3] causing rapid degradation in verification performance. A study by Zhang et al. [30] has shown a system employing 2D Gabor wavelet-derived features to be robust to moderate changes in the illumination direction. However, a different study by Adini et al. [1] shows that the 2D Gabor wavelet-derived features are sensitive to gross changes in the illumination direction.

Belhumeur et al. [3] 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 paper 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-mod 2 method, which utilizes polynomial coefficients derived from 2D DCT coefficients of spatially neighboring blocks, is the most suitable. We then compare the robustness and performance of the DCT-mod 2 method against two popular feature extraction techniques: eigenfaces (PCA) and 2D Gabor wavelets.

The rest of the paper is organized as follows. In Section 2 we briefly review the PCA, Gabor and 2D DCT feature extraction techniques and describe the proposed feature extraction methods. In Section 3 we describe a Gaussian mixture model (GMM) classifier which shall be used as the basis for experiments. The performance of the described feature extraction techniques is compared in Section 4, using an artificial illumination direction change. Section 5 is devoted to experiments on the Weizmann database [1] which has more realistic illumination direction changes. The paper is concluded in Section 6.

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. PCA (eigenfaces)

The PCA-derived features are obtained as follows. Given a face image matrix[1] F of size $Y \times X$, we construct a vector representation by concatenating all the columns of F to form a column vector $\vec{f}$ of dimensionality YX. Given a set of training vectors $\{\vec{f}_i\}_{i=1}^{N_P}$ for all persons, we define the mean of the training set as $\vec{f}_\mu$. A new set of mean subtracted vectors is formed using

$$\vec{g}_i = \vec{f}_i - \vec{f}_\mu, \quad i = 1, 2, \ldots, N_P. \tag{1}$$

The mean subtracted training set is represented as matrix $\boldsymbol{G} = [\vec{g}_1 \vec{g}_2 \ldots \vec{g}_{N_P}]$. The covariance matrix is calculated using

$$\boldsymbol{C} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}. \tag{2}$$

Let us construct matrix $\boldsymbol{U} = [\vec{u}_1 \vec{u}_2 \ldots \vec{u}_D]$, containing D eigenvectors of $\boldsymbol{C}$ with largest corresponding eigenvalues. Here, $D < N_P$. A feature vector $\vec{x}$ of dimensionality D is then derived from a face vector $\vec{f}$ using

$$\vec{x} = \boldsymbol{U}^{\mathrm{T}}(\vec{f} - \vec{f}_\mu), \tag{3}$$

i.e., face vector $\vec{f}$ decomposed in terms of D eigenvectors, known as "eigenfaces" [27].

[1] The face images used in our experiments have 56 rows (Y) and 64 columns (X).

2.2. 2D Gabor wavelets

The biologically inspired family of 2D Gabor wavelets is defined as follows [19]:

$$\Psi(y,x,\omega,\theta)=\frac{\omega}{\kappa\sqrt{2\pi}}\,\psi_A(y,x,\omega,\theta) \times\left[\psi_B(y,x,\omega,\theta)-\exp\left\{-\frac{\kappa^2}{2}\right\}\right], \tag{4}$$

where

$$\psi_A(y,x,\omega,\theta)=\exp\left\{-\frac{\omega^2}{8\kappa^2}[4(y\sin\theta+x\cos\theta)^2 +(y\cos\theta-x\sin\theta)^2]\right\} \tag{5}$$

and

$$\psi_B(y,x,\omega,\theta)=\exp\{\mathrm{i}(\omega y\sin\theta+\omega x\cos\theta)\}. \tag{6}$$

Here ω is the radial frequency in radians per unit length and θ is the wavelet orientation in radians. Each wavelet is centered at point $(y,x)=(0,0)$. The family is made up of wavelets for N_ω radial frequencies, each with N_θ orientations. The radial frequencies are spaced in octave steps and cover a range from $\omega_{\min}>0$ to $\omega_{\max}<\pi$, where 2π represents the Nyquist frequency. Typically, $\kappa\approx\pi$ so that each wavelet has a frequency bandwidth of one octave [19].

Feature extraction is done as follows. A coarse rectangular grid is placed over given face image F. At each node of the grid, the inner product of F with each member of the family is computed:

$$P_{j,k}=\int_y\int_x \Psi(y_0-y,x_0-x,\omega_j,\theta_k) \times F(y,x)\,\mathrm{d}x\,\mathrm{d}y \tag{7}$$

for $j=1,2,\ldots,N_\omega$ and $k=1,2,\ldots,N_\theta$. Here, the node is located at (y_0,x_0). An $N_\omega N_\theta$-dimensional feature vector[2] for location (y_0,x_0), is then constructed using the modulus of each inner product [17]

$$\vec{x}=[|P_{1,1}||P_{1,2}|\cdots|P_{1,N_\omega}|\cdots|P_{2,1}||P_{2,2}|\cdots |P_{2,N_\omega}|\cdots|P_{N_\theta,N_\omega}|]^{\mathrm{T}}. \tag{8}$$

Thus, if there are N_G nodes in the grid, we extract N_G feature vectors from one image.

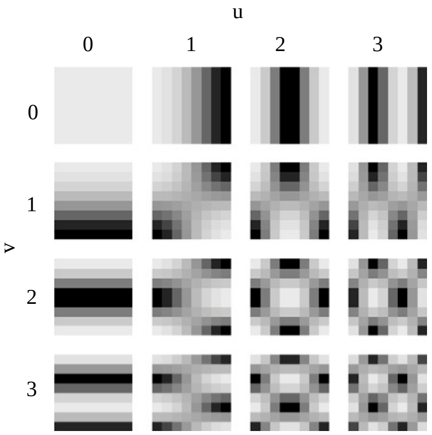

Fig. 1. Several DCT basis functions for $N=8$. Lighter colors represent larger values.

2.3. 2D DCT

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

$$C(v,u)=\alpha(v)\alpha(u)\sum_{y=0}^{N-1}\sum_{x=0}^{N-1}f(y,x)\beta(y,x,v,u) \tag{9}$$

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

$$\alpha(v)=\begin{cases}\sqrt{\dfrac{1}{N}} & \text{for } v=0,\\[2ex] \sqrt{\dfrac{2}{N}} & \text{for } v=1,2,\ldots,N-1\end{cases} \tag{10}$$

[2] Typically, $N_\omega=3$ and $N_\theta=6$, resulting in an 18 dimensional vector.

$$\begin{array}{cc} & \overbrace{\begin{array}{cccc}\mathbf{0}&\mathbf{1}&\mathbf{2}&\mathbf{3}\end{array}}^{u} \\ v\left\{\begin{array}{c}\mathbf{0}\\ \mathbf{1}\\ \mathbf{2}\\ \mathbf{3}\end{array}\right. & \begin{array}{|c|c|c|c|}\hline 0&1&5&6\\ \hline 2&4&7&12\\ \hline 3&8&11&13\\ \hline 9&10&14&15\\ \hline\end{array}\end{array}$$

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

and

$$\beta(y,x,v,u)=\cos\left[\frac{(2y+1)v\pi}{2N}\right]\cos\left[\frac{(2x+1)u\pi}{2N}\right]. \tag{11}$$

The coefficients are ordered according to a zig-zag pattern, reflecting the amount of information stored [13] (see Fig. 2). For block located at (b,a), the DCT feature vector is composed of

$$\vec{x}=[c_0^{(b,a)}c_1^{(b,a)}\ldots c_{M-1}^{(b,a)}]^{\mathrm{T}}, \tag{12}$$

where $c_n^{(b,a)}$ denotes the nth DCT coefficient and M is the number of retained coefficients. To ensure adequate representation of the image, each block overlaps its horizontally and vertically neighboring blocks by 50% [10]. Thus for an image which has Y rows and X columns, there are $N_D=(2Y/N-1)\times(2X/N-1)$ blocks.[3]

2.4. 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 [11,26].

For images, we define the nth *horizontal* delta coefficient for block located at (b,a) as a modified first-order orthogonal polynomial coefficient

$$\Delta^h c_n^{(b,a)}=\frac{\sum_{k=-K}^{K} kh_k c_n^{(b,a+k)}}{\sum_{k=-K}^{K} h_k k^2}. \tag{13}$$

Similarly, we define the nth *vertical* delta coefficient as

$$\Delta^v c_n^{(b,a)}=\frac{\sum_{k=-K}^{K} kh_k c_n^{(b+k,a)}}{\sum_{k=-K}^{K} h_k k^2}, \tag{14}$$

where h is a $2K+1$-dimensional symmetric window vector. In this work we shall use $K=1$ and a rectangular window.

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 are close or overlapping). To find the deltas for block Y, we apply Eq. (13) to obtain (ignoring the denominator)

$$\Delta^h Y=-X+Z, \tag{15}$$

$$=-(I_X+I_N)+(I_Z+I_N), \tag{16}$$

$$=I_Z-I_X, \tag{17}$$

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 DCT coefficients from each block, the delta vector is $2M$ dimensional. We shall term this feature extraction method as DCT-delta. We interpret these delta coefficients as transitional spatial information (somewhat akin to edges).

DCT-delta feature extraction for a given block is only possible when the block has vertical and horizontal neighbors. Thus, processing an image which has Y rows and X columns and using a 50% block overlap results in $N_{D2}=(2Y/N-3)\times(2X/N-3)$ DCT-delta feature vectors.[4]

2.5. *DCT-mod, DCT-*mod 2 *and DCT-mod-delta*

By inspecting Eqs. (9) and (11), it is evident that the 0th DCT coefficient will reflect the average pixel value (or the DC level) inside each block and hence

[3] Thus for a 56 × 64 image, there are 195 DCT feature vectors.

[4] Thus for a 56 × 64 image, there are 143 DCT-delta feature vectors.

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 DCT feature vector is M dimensional):

(1) Discard the first three coefficients from the baseline DCT feature vector. We shall term this *modified* feature extraction method as DCT-mod.
(2) Discard the first three coefficients from the baseline 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, i.e.,

$$\vec{x} = [\Delta^h c_0 \Delta^v c_0 \Delta^h c_1 \Delta^v c_1 \Delta^h c_2 \Delta^v c_2 c_3 c_4 \ldots c_{M-1}]^{\mathrm{T}}, \tag{18}$$

where the (b,a) superscript was omitted. Let us term this approach as DCT-mod 2.

Thus, in the DCT-mod-delta and DCT-mod 2 approaches transitional spatial information is combined with local texture information.

3. GMM classifier

The distribution of feature vectors for each person is modeled by a GMM. Given a set of training vectors, an N_M-mixture GMM is trained using a *k-means* clustering algorithm followed by 10 iterations of the expectation maximization (EM) algorithm [6,9,20].

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

$$\mathscr{L}(X|\lambda_C) = \frac{1}{N_V}\sum_{i=1}^{N_V} \log p(\vec{x}_i|\lambda_C), \tag{19}$$

where

$$p(\vec{x}|\lambda) = \sum_{j=1}^{N_M} m_j \mathscr{N}(\vec{x}; \vec{\mu}_j, \mathbf{\Sigma}_j) \tag{20}$$

and

$$\lambda = \{m_j, \vec{\mu}_j, \mathbf{\Sigma}_j\}_{j=1}^{N_M}. \tag{21}$$

Here λ_C is the model for person C. N_M is the number of mixtures, m_j is the weight for mixture j (with constraint $\sum_{j=1}^{N_M} m_j = 1$), and $\mathscr{N}(\vec{x}; \vec{\mu}, \mathbf{\Sigma})$ is a multi-variate Gaussian function with mean $\vec{\mu}$ and diagonal covariance matrix $\mathbf{\Sigma}$:

$$\mathscr{N}(\vec{x}; \vec{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[\frac{-1}{2}(\vec{x}-\vec{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right]. \tag{22}$$

Given a set $\{\lambda_b\}_{b=1}^{B}$ of B background person models for person C, the average log likelihood of the claimant being an impostor is found using

$$\mathscr{L}(X|\lambda_{\bar{C}}) = \log\left[\frac{1}{B}\sum_{b=1}^{B} \exp \mathscr{L}(X|\lambda_b)\right]. \tag{23}$$

The set of background person models is found using the method described in [21]. An opinion on the claim is found using

$$\Lambda(X) = \mathscr{L}(X|\lambda_C) - \mathscr{L}(X|\lambda_{\bar{C}}). \tag{24}$$

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

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

4. Experiments

4.1. VidTIMIT audio-visual database

The VidTIMIT database [24] is comprised of video and corresponding audio recordings of 43 people

(19 female and 24 male), reciting short sentences selected from the NTIMIT corpus [15]. It was recorded in three sessions, with a mean delay of 7 days between Sessions 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 first two sentences for all persons are the same, with the remaining eight generally different for each person. The mean duration of each sentence is 4.25 s, or approximately 106 video frames.

The recording was done in a noisy office environment using a broadcast quality digital video camera. The video of each person is stored as a sequence of JPEG images with a resolution of 384 × 512 (rows × columns) pixels. The corresponding audio is stored as a mono, 16 bit, 32 kHz WAV file.

For more information on the database, please visit *http://spl.me.gu.edu.au/vidtimit/* or contact the authors.

4.2. Experimental setup

Before feature extraction can occur, the face must first be located [5]. Furthermore, to account for varying distances to the camera, a geometrical normalization must be 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 of varying dimensions. Using the distance between the eyes as a size measure, an affine transformation is used [13] to adjust the size of the image, resulting in the distance between the eyes to be the same for each person. Finally, a 56 × 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 DCT and DCT derived methods, each block is 8 × 8 pixels. Moreover, each block overlaps with horizontally and vertically adjacent blocks by 50%. For PCA, the dimensionality of the face window is reduced to 40 (choice based on the work by Samaria [23] and Belhumeur et al. [3]). For Gabor wavelet features, we heed the choice of Duc et al. [8] with $N_\omega = 3$, $N_\theta = 6$, $\omega_1 = \pi/2$, $\omega_2 = \pi/4$, $\omega_3 = \pi/8$ and $\theta_k = \pi(k-1)/N_\theta$ (where $k = 1, 2, \ldots, N_\theta$). Hence, the dimensionality of the Gabor feature vectors is 18.

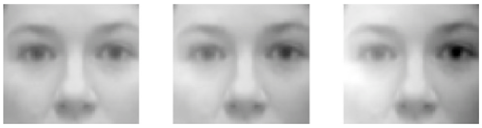

Fig. 3. Examples of varying light illumination; left: $\delta = 0$ (no change); middle: $\delta = 40$; right: $\delta = 80$.

The location of the wavelet centers was chosen to be as close as possible to the centers of the blocks used in DCT-mod 2 feature extraction.

To reduce the computational burden during modeling and testing, every second video frame was used. For each feature extraction method, eight mixture client models (GMMs) were generated from features extracted from face windows in Session 1.

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

$$v(y,x) = w(y,x) + mx + \delta \quad \text{for } y = 0, 1, \ldots, 55 \quad \text{and} \quad x = 0, 1, \ldots, 63, \tag{25}$$

where

$$m = \frac{-\delta}{63/2} \quad \text{and} \quad \delta = \text{illumination delta (in pixels)}. \tag{26}$$

Example face windows for various δ are shown in Fig. 3. It must be noted that the above artificial illumination change does not cover all the effects of illumination changes possible in real life (shadows, etc.).

To find the performance, Sessions 2 and 3 were used for obtaining example opinions of known impostor and true claims. Four utterances, each from eight fixed persons (4 male and 4 female), were used for simulating impostor accesses against the remaining 35 persons. As in [21], 10 background person models were used for the impostor likelihood calculation. For each of the remaining 35 persons, their four utterances were 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

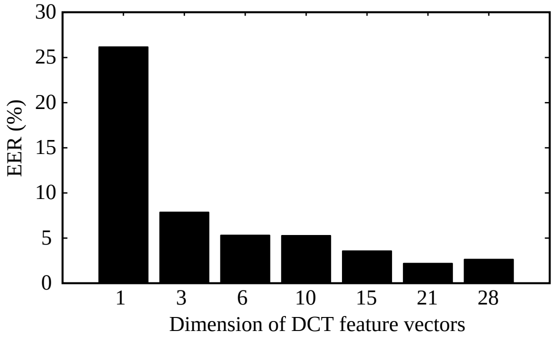

Fig. 4. Performance for varying dimensionality of DCT feature vectors.

performance is as close as possible to equal error rate (EER) (i.e., where the false acceptance rate is equal to the false rejection rate).

In the first experiment, we found the performance of the 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. 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 DCT feature vectors—hence the dimensionality of DCT-delta is 30, DCT-mod is 12, DCT-mod-delta is 42 and DCT-mod 2 is 18.

In the second experiment we compared the performance of DCT and all of the proposed techniques for increasing δ. Results are shown in Fig. 5.

In the third experiment we compared the performance of PCA, DCT, Gabor and DCT-mod 2 features for varying δ. Results are presented in Fig. 6.

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, DCT, Gabor and DCT-mod 2 feature extraction techniques. It must be noted that apart from having the transformation data pre-calculated (e.g., β 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.

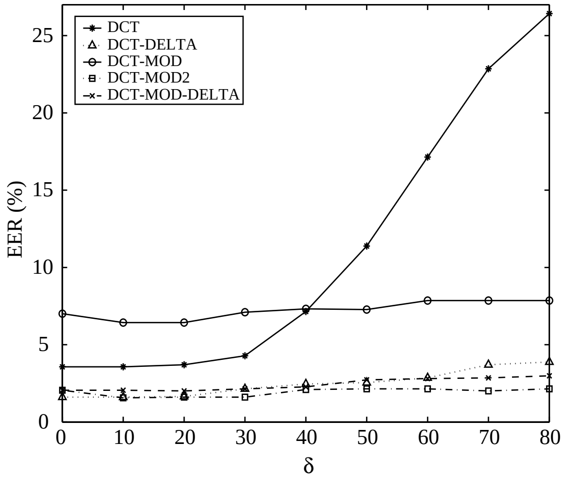

Fig. 5. Performance of DCT and proposed feature extraction techniques.

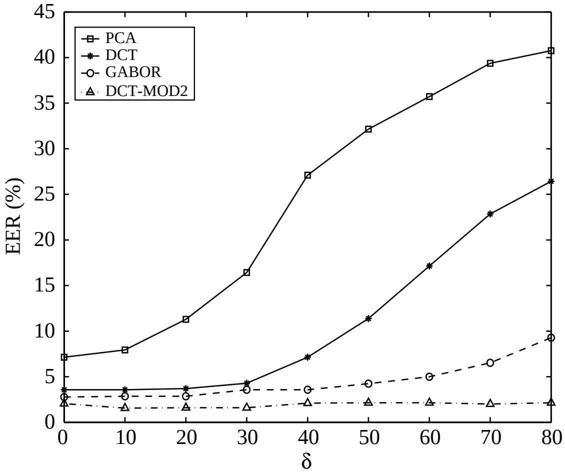

Fig. 6. Performance of PCA, DCT, Gabor and DCT-mod 2 feature extraction techniques.

4.3. Discussion

We can see in Fig. 4 that the first three DCT coefficients contain a significant amount of person dependent information. Thus ignoring them (as in DCT-mod) implies a reduction in performance. This

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

Method	Time (ms)
PCA	11
DCT	6
Gabor	675
DCT-mod 2	8

is verified in Fig. 5, where the DCT-mod features have worse performance than DCT features when there is little or no illumination change ($\delta \leqslant 30$). Performance of DCT features is fairly stable for small illumination changes but degrades for $\delta \geqslant 40$. This is in contrast to DCT-mod features which have a relatively static performance.

The remaining proposed features (DCT-delta, DCT-mod-delta and DCT-mod 2) do not have the performance penalty present in DCT-mod. Moreover, all of them have similarly better performance than DCT features. DCT-mod 2 edges out DCT-delta and DCT-mod-delta in terms of stability for large illumination changes ($\delta \geqslant 50$). Additionally, the dimensionality of DCT-mod 2 is lower than DCT-delta and DCT-mod-delta.

The results suggest that delta features make the system more robust as well as improve performance. The results also suggest that it is only necessary to use deltas of coefficients representing the DC level and low frequency features (i.e., the 0th, 1st and 2nd DCT coefficients) while keeping the remaining DCT coefficients unchanged. Hence out of the four proposed feature extraction techniques, the DCT-mod 2 approach is the most suitable.

Comparing PCA, DCT, Gabor and DCT-mod 2 (Fig. 6), we can see that the DCT-mod 2 approach is the most immune to illumination changes—the performance is virtually flat for varying δ. The performance of PCA-derived features rapidly degrades as δ increases. Performance of Gabor features is stable for $\delta \leqslant 40$ and then gently deteriorates as δ increases. The results suggests that we can order the features, based on their robustness and performance, as follows: DCT-mod 2, Gabor, DCT, and lastly, PCA.

It must be noted that using the introduced illumination change, the center portion of the face (columnwise) is largely unaffected. The size of the portion decreases as δ 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 (Gabor, DCT and DCT-mod 2), where one feature vector describes only a small part of the face. Thus a significant percentage (dependent on δ) of the feature vectors is virtually unchanged, automatically leading to a degree of robustness.

It must also be noted that when using the GMM classifier in conjunction with the Gabor, DCT or DCT-mod 2 features, the spatial relation between major face features (e.g., eyes and nose) is lost. However, excellent performance is still obtained.

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

5. Experiments on the Weizmann database

The experiments described in Section 4 utilize an artificial illumination direction change. In this section, we shall compare the performance of DCT, Gabor and DCT-mod 2 feature sets on the Weizmann Database [1], 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 4.2. 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

Table 2
Results on the Weizmann database, quoted in terms of approximate EER (%)

Method	Illumination direction		
	Uniform	Left	Right
DCT	3.49	48.15	48.15
Gabor	0.36	33.34	33.34
DCT-mod 2	0	25.93	22.65

maximum a posteriori adaptation [12,22]. 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 in [22].

Since PCA based feature extraction produces one feature vector per image (see Section 2.1), 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.

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

6. Conclusion

In this paper we have proposed four new facial feature extraction techniques, which are robust to an illumination direction change. Out of the proposed methods, the DCT-mod 2 method, which utilizes polynomial coefficients derived from 2D DCT coefficients of spatially neighboring blocks, is the most suitable. Face verification results on the multi-session VidTIMIT database suggest that the DCT-mod 2 feature set is superior (in terms of robustness to illumination direction changes and discrimination ability) to features extracted using three popular methods: eigenfaces (PCA), 2D DCT and 2D Gabor wavelets. Moreover, compared to Gabor wavelets, the DCT-mod 2 feature set is over 80 times faster to compute. Additional experiments on the Weizmann Database also showed that the DCT-mod 2 approach is more robust than 2D Gabor wavelets and 2D DCT coefficients.

References

[1] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 721–732.

[2] W. Atkins, A testing time for face recognition technology, Biometric Technol. Today 9 (3) (2001) 8–11.

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

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

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

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

[7] G.R. Doddington, M.A. Przybycki, A.F. Martin, D.A. Reynolds, The NIST speaker recognition evaluation—overview, methodology, systems, results, perspective, Speech Commun. 31 (2–3) (2000) 225–254.

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

[9] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, New York, 2001.

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

[11] S. Furui, Cepstral analysis technique for automatic speaker verification, IEEE Trans. Acoust. Speech Signal Process. 29 (2) (1981) 254–272.

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

[13] R.C. Gonzales, R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, MA, 1993.

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

[15] C. Jankowski, A. Kalyanswamy, S. Basson, J. Spitz, NTIMIT: a phonetically balanced, continuous speech telephone bandwidth speech database, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Vol. 1, Albuquerque, 1990, pp. 109–112.

[16] C. Kotropoulos, A. Tefas, I. Pitas, Frontal face authentication using morphological elastic graph matching, IEEE Trans. Image Process. 9 (4) (2000) 555–560.

[17] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput. 42 (3) (1993) 300–311.

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

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

[20] T.K. Moon, Expectation-maximization algorithm, IEEE Signal Process. Mag. 13 (6) (1996) 47–60.

[21] D.A. Reynolds, Speaker identification and verification using Gaussian mixture speaker models, Speech Commun. 17 (1–2) (1995) 91–108.

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

[23] F. Samaria, Face recognition using hidden Markov models, Ph.D. Thesis, University of Cambridge, 1994.

[24] C. Sanderson, K.K. Paliwal, Noise compensation in a person verification system using face and multiple speech features, Pattern Recognition 36 (2) (2003) 293–302.

[25] F. Smeraldi, J. Bigün, Retinal vision applied to facial features detection and face authentication, Pattern Recognition Lett. 23 (4) (2002) 463–473.

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

[27] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.

[28] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

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

[30] J. Zhang, Y. Yan, M. Lades, Face recognition: eigenface, elastic matching, and neural nets, Proc. IEEE 85 (9) (1997) 1423–1435.