Fast feature extraction method for robust face verification

C. Sanderson and K.K. Paliwal

A feature extraction technique for face verification is proposed. It utilises polynomial coefficients derived from 2D discrete cosine transform (DCT) coefficients of neighbouring blocks. Experimental results suggest that the technique is more robust against illumination direction changes than 2D Gabor wavelets, 2D DCT and eigenface methods. Moreover, compared to Gabor wavelets, the proposed technique is over 80 times quicker to compute.

Introduction: Recently there has been much interest in biometric verification systems. A face verification system verifies the claimed identity (a 2 class task) based on images (or a video sequence) of the claimant's face. The claimant is either accepted (classified as a true claimant) or rejected (classified as an impostor).

There are many approaches to facial feature extraction techniques – ranging from the eigenface approach [1], 2D Gabor wavelets [2] to 2D discrete cosine transform (DCT) [3]. PCA derived features have been shown to be sensitive to changes in the illumination direction [4] causing rapid degradation in verification performance. A study by Adini *et al.* [5] shows that the 2D Gabor wavelet derived features are also sensitive to the illumination direction.

As will be shown, 2D DCT-based features are also sensitive to changes in the illumination direction. In this Letter, we introduce the DCT-mod2 feature extraction technique and compare its performance against the 2D DCT, eigenface and 2D Gabor wavelet approaches. We show that the proposed technique is significantly less affected by an illumination direction change.

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

2D DCT feature extraction: Here the given face image is analysed 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. The result is an $N \times N$ matrix $C(v, u)$ containing DCT coefficients:

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

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

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

$\alpha(v) = \sqrt{(1/N)}$ for $v = 0$, and $\alpha(v) = \sqrt{(2/N)}$ for $v = 1, 2, \ldots, N-1$. The coefficients are ordered according to a zig-zag pattern, reflecting the amount of information stored [3]. For block located at (b, a), the DCT feature vector is composed of:

$$x = [\, c_0^{(b,a)} \quad c_1^{(b,a)} \quad \cdots \quad c_{M-1}^{(b,a)} \,]^T \tag{3}$$

where $c_n^{(b,a)}$ denotes the nth DCT coefficient and M is the number of retained coefficients.

Proposed feature extraction method: 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 [6].

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

$$\Delta^h c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k c_n^{(b,a+k)}}{\sum_{k=-K}^{K} h_k k^2} \tag{4}$$

Similarly, we define the nth vertical delta coefficient as:

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

where h is a $2K+1$ dimensional symmetric window vector. In this Letter we shall use $K = 1$ and a rectangular window. We interpret these delta coefficients as transitional spatial information (somewhat akin to edges).

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 (4) to obtain (ignoring the denominator):

$$\Delta^h Y = -X + Z = -(I_X + I_N) + (I_Z + I_N) \tag{6}$$

$$= I_Z - I_X \tag{7}$$

i.e. the noise component is removed.

By inspecting (1) and (2), it is evident that the 0th DCT coefficient $[C(0,0)]$ reflects the average pixel value (or the DC level) inside each block and hence will be the most affected by any illumination change. Moreover, it is evident that the first $[C(0,1)]$ and second $[C(1,0)]$ coefficients represent the average horizontal and vertical pixel intensity change, respectively. As such, they will also be significantly affected by any illumination change. Therefore, we replace the first three coefficients with their horizontal and vertical deltas and form a feature vector representing the given block as follows:

$$x = [\Delta^h c_0 \; \Delta^v c_0 \; \Delta^h c_1 \; \Delta^v c_1 \; \Delta^h c_2 \; \Delta^v \; c_2 \; c_3 \; c_4 \cdots c_{M-1}]^T \tag{8}$$

where the (b, a) superscript is omitted for clarity. Let us term this modified approach as DCT-mod2. It must be noted that in this approach transitional spatial information is combined with local texture information.

Experiments: The experiments were performed on the VidTIMIT database [7], which is composed of video recordings of 43 people, reciting short sentences. It was recorded in three sessions, with one week delay between each session.

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 each video frame.

For PCA, the dimensionality of the face window is reduced to 40 (based on [4]). For 2D DCT and DCT-mod2 methods, each block is 8×8 pixels. Moreover, each block overlaps with horizontally and vertically adjacent blocks by 50%. Based on [3] we have chosen 15 as the dimensionality of baseline DCT feature vectors; hence the dimensionality of DCT-mod2 is 18.

For Gabor features, we follow [2] where the dimensionality of the Gabor feature vectors is 18. The location of the wavelet centres was chosen to be as close as possible to the centres of the blocks used in DCT-mod2 feature extraction.

For classification we have utilised the Gaussian mixture model approach [8]. For each feature extraction method, client models were generated from features extracted from face windows in session 1. An 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:

$$v(y,x) = w(y,x) + mx + \delta \tag{9}$$

for $y = 0, 1, \ldots, 55$ and $x = 0, 1, \ldots, 63$, where $m = -\delta/(63/2)$ and δ = illumination delta (in pixels). Example face windows are shown in Fig. 1. 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, etc.), but it is useful in providing suggestive results.

Fig. 1 *Examples of varying light illumination*

Sessions 2 and 3 were used to find verification performance. For each δ there were 1120 impostor and 140 true claims. The decision threshold was then set so the *a posteriori* 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) [8].

In the first experiment, we compared the performance of PCA, DCT, Gabor and DCT-mod2 features for varying δ. Results are presented in Fig. 2. 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 second experiment we compared the average time taken to process one face window by all feature extraction techniques. Results are listed in Table 1.

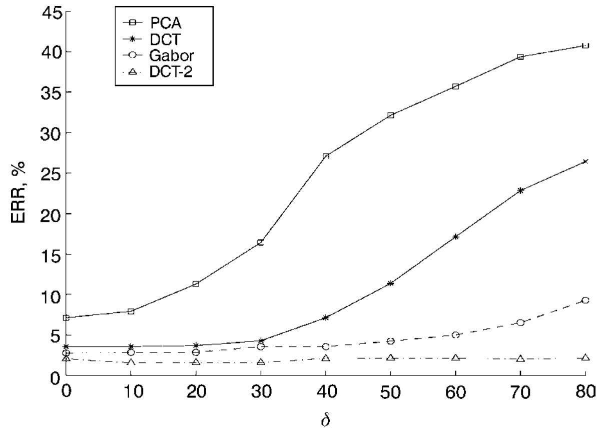

Fig. 2 *Performance of feature extraction techniques*

Table 1: Average time taken per face window (Pentium 3, 500 MHz)

Method	PCA	DCT	Gabor	DCT-mod2
Time [ms]	11	6	675	8

Conclusions: Comparing PCA, DCT, Gabor and DCT-mod2 (Fig. 2), we can see that the DCT-mod2 approach is the most immune to illumination direction changes – the performance is virtually flat for varying δ. The performance of eigenface features rapidly degrades as δ increases. Performance of Gabor features is stable for $\delta \leq 40$ and then gently deteriorates as δ increases. From Table 1 we can see that Gabor features are the most computationally expensive to calculate, taking about 84 times longer than DCT-mod2 features. Compared to Gabor features, PCA, DCT and DCT-mod2 features take a similar amount of time to process one face window.

20 June 2002
Electronics Letters Online No: 20021186
DOI: 10.1049/el:20021186

C. Sanderson and K.K. Paliwal (*School of Microelectronic Engineering, Griffith University, Nathan, QLD 4111, Australia*)

E-mail:conradsand@ieee.org

References

1 TURK, M., and PENTLAND, A.: 'Eigenfaces for recognition', *J. Cogn. Neurosci.*, 1991, **3**, (1), pp. 71–86
2 DUC, B., FISCHER, S., and BIGÜN, J.: 'Face authentication with Gabor information on deformable graphs', *IEEE Trans. Image Proc.*, 1999, **8**, (4), pp. 504–516
3 EICKLER, S., MÜLLER, S., and RIGOLL, G.: 'Recognition of JPEG compressed face images based on statistical methods', *Image Vis. Comput.*, 2000, **18**, (4), pp. 279–287
4 BELHUMEUR, P.N., HESPANHA, J.P., and KRIEGMAN, D.J.: 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720
5 ADINI, Y., MOSES, Y., and ULLMAN, S.: 'Face recognition: the problem of compensating for changes in illumination direction', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 721–732
6 SOONG, F.K., and ROSENBERG, A.E.: 'On the use of instantaneous and transitional spectral information in speaker recognition', *IEEE Trans. Acoust, Speech Signal Process*, 1988, **36**, (6), pp. 871–879
7 SANDERSON, C., and PALIWAL, K.K.: 'Noise compensation in a person verification system using face and multiple speech features', *Pattern Recognit.* (to appear)
8 REYNOLDS, D.A.: 'Speaker identification and verification using Gaussian mixture speaker models', *Speech Commun.*, 1995, **17**, (1–2), pp. 91–108