ROBUST FEATURES FOR FRONTAL FACE AUTHENTICATION IN DIFFICULT IMAGE CONDITIONS

Conrad Sanderson and Samy Bengio *

IDIAP
Rue du Simplon 4, CH-1920 Martigny, Switzerland
conradsand@ieee.org, samy.bengio@idiap.ch

Abstract. In this paper we extend the recently proposed *DCT-mod2* feature extraction technique (which utilizes polynomial coefficients derived from 2D DCT coefficients obtained from horizontally & vertically neighbouring blocks) via the use of various windows and diagonally neighbouring blocks. We also propose *enhanced PCA*, where traditional PCA feature extraction is combined with *DCT-mod2*. Results using test images corrupted by a linear and a non-linear illumination change, white Gaussian noise and compression artefacts, show that use of diagonally neighbouring blocks and windowing is detrimental to robustness against illumination changes while being useful for increasing robustness against white noise and compression artefacts. We also show that the *enhanced PCA* technique retains all the positive aspects of traditional PCA (that is, robustness against white noise and compression artefacts) while also being robust to illumination changes; moreover, *enhanced PCA* outperforms PCA with histogram equalisation pre-processing.

1 Introduction

A face verification (authentication) system verifies the claimed identity based on images (or a video sequence) of the claimant's face. Such systems have forensic and security (i.e., access control) applications. Generally speaking, a full face verification system can be thought of as being comprised of three stages:

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

The second stage (normalization) usually involves an affine transformation [8] (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 work we shall concentrate on the feature extraction part of the last stage.

* The authors thank the Swiss National Science Foundation for supporting this work through the National Centre of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM2).

There are many approaches to facial feature extraction; for example, Turk and Pentland [17] used Principal Component Analysis (PCA), Duc et al. [5] used biologically inspired 2D Gabor wavelets, while Eickeler et al. [7] obtained features using the 2D Discrete Cosine Transform (DCT). Recently, Sanderson & Paliwal [14] used a modified form of DCT feature extraction, termed *DCT-mod2*, which has been shown to be robust against illumination direction changes.

While robustness against illumination direction changes may be of most concern in security systems, in forensic applications [10] other types of image corruption can be important. Here, face images may be obtained in various illumination conditions from various sources: digitally stored video, possibly damaged and/or low quality analogue video tape or TV signal corrupted with "static" noise.

The rest of the paper is organized as follows. In Section 2 we describe and extend the *DCT-mod2* feature extraction technique through the use of various windows. In Section 3 we further extend the *DCT-mod2* approach via the addition of extra features. In Section 4, PCA and *DCT-mod2* feature extraction techniques are combined to form *enhanced PCA*. In Section 5 we describe a Gaussian Mixture Model (GMM) classifier which shall be used as the basis for experiments. In Section 6, the performance of all presented feature extraction techniques is evaluated on images corrupted by a linear & non-linear illumination change, white Gaussian noise (simulating "static" noise) and compression artefacts (simulating compressed digital video). Section 7 is devoted to discussion of results and conclusions.

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 Extension of DCT-mod2 feature extraction

In *DCT-mod2* feature extraction [14] a given face image is analyzed on a block by block basis; each block is $N_P \times N_P$ (here we use $N_P = 8$) and overlaps neighbouring blocks by 50%. Each block is decomposed in terms of 2D Discrete Cosine Transform (DCT) basis functions [8]. A feature vector for each block is then constructed as:

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

where c_n represents the n-th DCT coefficient, M is the number of retained DCT coefficients, while $\Delta^h c_n$ and $\Delta^v c_n$ represent the horizontal and vertical delta coefficients respectively. For a block located at (b, a), the delta coefficients are defined as modified orthogonal polynomial coefficients [16]:

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

$$\Delta^v c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k \; c_n^{(b+k,a)}}{\sum_{k=-K}^{K} h_k k^2} \tag{3}$$

where $\boldsymbol{h}$ is a $2K + 1$ dimensional symmetric window vector and $c_n^{(b,a)}$ is the n-th DCT coefficient for block located at (b, a).

Compared to traditional DCT feature extraction [7, 8], the first three DCT coefficients are replaced by their respective horizontal and vertical deltas, in order to reduce the effects of illumination direction changes.

Since *DCT-mod2* feature extraction for a given block is only possible when the block has vertical and horizontal neighbours, processing an image which has N_Y rows and N_X columns results in $N_D = (2\frac{N_Y}{N_P} - 3) \times (2\frac{N_X}{N_P} - 3)$ feature vectors[1].

In [14], delta coefficients were calculated using $K = 1$ and a rectangular window (i.e., $\boldsymbol{h} = [\ 1\ 1\ 1\ ]^T$), which amounted to finding the differences between DCT coefficients obtained from neighbouring blocks. In this paper we shall extend the *DCT-mod2* approach with $K = 2$ (which increases the number of blocks used in deriving a *DCT-mod2* feature vector) and various windows.

By inspecting Eqns. (2) & (3) and assuming that a rectangular window is used, it can be seen that for $K = 2$, DCT coefficients from blocks with $k = -2$ and $k = 2$ have the largest contribution to the final value; since this may not be optimal, we shall study two additional windows:

- Window B, where $\boldsymbol{h} = [\ 0.5\ 1.0\ 1.0\ 1.0\ 0.5\ ]^T$, causing all DCT coefficients to have equal contribution
- Window C, where $\boldsymbol{h} = [\ 0.25\ 1.0\ 1.0\ 1.0\ 0.25\ ]^T$, causing the DCT coefficients from the outer blocks to have smaller contribution

We shall refer to the rectangular window ($\boldsymbol{h} = [\ 1\ 1\ 1\ 1\ 1\ ]^T$) as Window A.

3 Proposed DCT-mod3 feature extraction

In [14], *DCT-mod2* feature extraction has been shown to be robust against a horizontal illumination direction change. Since an illumination direction change can occur in any direction, we propose to extend the *DCT-mod2* approach by including delta coefficients for both diagonal directions, defined as follows:

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

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

A feature vector for each block is then constructed as:

$$\begin{aligned} \boldsymbol{x} = [\ & \Delta^d c_0\ \Delta^e c_0\ \Delta^d c_1\ \Delta^e c_1\ \Delta^d c_2\ \Delta^e c_2 \quad \Delta^h c_0\ \Delta^v c_0\ \Delta^h c_1\ \Delta^v c_1\ \Delta^h c_2\ \Delta^v c_2 \\ & c_3\ \ c_4\ \ \cdots\ \ c_{M-1}\ ]^T \end{aligned} \tag{6}$$

where the (b, a) superscript was omitted for clarity. We shall term this approach as *DCT-mod3*.

We will evaluate the performance of the *DCT-mod3* approach for $K = 1$ and $K = 2$ with various windows (as explained in Section 2 for the *DCT-mod2* approach).

4 Proposed Enhanced PCA

In standard PCA based feature extraction (also known as eigenfaces [17]), a given face image is represented by matrix F containing grey level pixel values; F is converted to a face vector, $\boldsymbol{f}$, by concatenating all the columns; a D-dimensional feature vector, $\boldsymbol{x}$, is then obtained by:

$$\boldsymbol{x} = \mathbf{U}^T(\boldsymbol{f} - \boldsymbol{f}_\mu) \tag{7}$$

[1] Thus for a 56×64 image, there are $11 \times 13 = 143$ vectors.

where $\mathbf{U}$ contains D eigenvectors (with largest corresponding eigenvalues) of the training data covariance matrix, and $\boldsymbol{f}_\mu$ is the mean of training face vectors.

PCA derived features have been shown to be sensitive to changes in the illumination direction [1] causing rapid degradation in verification performance [14]. In the proposed *enhanced PCA* approach, a given face image is processed using *DCT-mod2* feature extraction to produce pseudo-image $\hat{F}$, which is then used in place of F by traditional PCA feature extraction. Since *DCT-mod2* feature vectors are robust to illumination changes, features obtained via the *enhanced PCA* should also be immune to illumination changes. Formally, the pseudo image is constructed as follows:

$$\hat{F} = \begin{bmatrix} \boldsymbol{c}^{\,(\Delta b,\Delta a)} & \boldsymbol{c}^{\,(\Delta b,2\Delta a)} & \boldsymbol{c}^{\,(\Delta b,3\Delta a)} & \cdots \\ \boldsymbol{c}^{\,(2\Delta b,\Delta a)} & \boldsymbol{c}^{\,(2\Delta b,2\Delta a)} & \boldsymbol{c}^{\,(2\Delta b,3\Delta a)} & \cdots \\ \boldsymbol{c}^{\,(3\Delta b,\Delta a)} & \boldsymbol{c}^{\,(3\Delta b,2\Delta a)} & \boldsymbol{c}^{\,(3\Delta b,3\Delta a)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{8}$$

where $\boldsymbol{c}^{\,(n\Delta b,n\Delta a)}$ denotes the *DCT-mod2* feature vector (generated using $K = 1$) for block located at $(n\Delta b, n\Delta a)$, while Δb and Δa are block location advancement constants for rows and columns respectively. Since $N_P = 8$ and we are using a 50% overlap, Δb and Δa are equal to 4. Because each *DCT-mod2* feature vector is $M + 3$ dimensional, matrix $\hat{F}$ has $(M+3)(2\frac{N_Y}{N_P} - 3)$ rows and $(2\frac{N_X}{N_P} - 3)$ columns.

The proposed *enhanced PCA* method will be compared against the standard approach (no pre-processing) as well as pre-processing using histogram equalisation [8], often used in an attempt to reduce the effects of varying illumination conditions [9, 11].

5 GMM Based Classifier

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

$$\mathcal{L}(X|\lambda_C) = \frac{1}{N_V} \sum_{i=1}^{N_V} \log p(\boldsymbol{x}_i|\lambda_C) \tag{9}$$

$$\text{where} \quad p(\boldsymbol{x}|\lambda) = \sum_{j=1}^{N_G} m_j \, \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{10}$$

$$\lambda = \{m_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{N_G} \tag{11}$$

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

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[\frac{-1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right] \tag{12}$$

λ_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 $\forall\, j : m_j \geq 0$).

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

$$\Lambda(X) = \mathcal{L}(X|\lambda_C) - \mathcal{L}(X|\lambda_{\overline{C}}) \tag{13}$$

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

5.1 Model Training

Given a set of training vectors, $X = \{\boldsymbol{x}_i\}_{i=1}^{N_V}$ (which may come from several images), the GMM parameters (λ) for each client model are found by adapting a Universal Background Model (UBM) using a form of *maximum a posteriori* (MAP) adaptation [12]. The UBM is trained with the Expectation Maximization (EM) algorithm [3, 6] using training data from all clients.

Since the UBM is a good representation of many clients, it is also used to find the likelihood of the claimant being an impostor, i.e.:

$$\mathcal{L}(X|\lambda_{\overline{C}}) = \mathcal{L}(X|\lambda_{\text{UBM}}) \tag{14}$$

6 Experiments

6.1 VidTIMIT Audio-Visual Database

The VidTIMIT database [15], 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 Session 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 seconds, or approximately 106 video frames.

6.2 Experiment Setup

Before feature extraction can occur, the face must first be located [2]. 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 [8] 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 based methods, the dimensionality of the face window is reduced to 32 (choice based on preliminary experiments and [1, 13]).

For ***DCT-mod2*** and ***DCT-mod3*** the number of retained DCT coefficients is $M = 15$ (choice based on [7, 14]) resulting in 18 and 24 dimensional vectors, respectively.

In our experiments, we use a sequence of images (video); if the sequence has N_I images, then $N_V = N_I$ for the PCA based approaches and $N_V = N_I N_D$ for ***DCT-mod2*** & ***DCT-mod3*** approaches.

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 636 images were used for training and 424 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. 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:

$$
\begin{aligned}
v(y,x) &= w(y,x) + mx + \delta \qquad (15)\\
&\text{for} \quad y = 0,1,...,N_Y-1 \quad \text{and} \quad x = 0,1,...,N_X-1\\
\text{where} \quad m &= \frac{-\delta}{(N_X-1)/2}\\
\delta &= \text{illumination delta (in pixels)}
\end{aligned}
$$

Since the above model of illumination direction change is rather restrictive, a second, Gaussian shaped (non-linear), artificial illumination was also used, as defined below:

$$
\begin{aligned}
v(y,x) &= w(y,x) + 2\delta \left(\exp\left[\frac{-1}{2} \boldsymbol{p}^{\,T} \mathbf{A}^{-1} \, \boldsymbol{p} \right] - \frac{1}{2} \right) \qquad (16)\\
&\text{for} \quad y = 0,1,...,N_Y-1 \quad \text{and} \quad x = 0,1,...,N_X-1\\
\text{where} \quad \boldsymbol{p} &= [\, y \;\; x \,]^T \; - \; [\, (N_Y-1)/2 \;\; (N_X-1)/2 \,]^T\\
\mathbf{A} &= \begin{bmatrix} (N_Y/4)^2 & 0 \\ 0 & (N_X/4)^2 \end{bmatrix}\\
\delta &= \text{illumination delta (in pixels)}
\end{aligned}
$$

For experiments involving compression artefacts, face windows extracted from Sessions 2 and 3 were processed by a JPEG codec [18] (simulating compressed digital video). The JPEG codec reduces the bitrate of a given image at the expense of introducing distortion in the form of compression artefacts; the distortion is measured in terms of Peak Signal to Noise Ratio (PSNR). The average PSNR of the corrupted images was 31.13 dB. Similarly, for TV "static" noise experiments, face windows extracted from Sessions 2 and 3 were corrupted by additive white Gaussian noise, resulting in the PSNR being equal to 22.5 dB. Example face windows are shown in Figure 1.

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. 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* performance was as close as possible to the Equal Error Rate (EER) (i.e., where the false acceptance rate is equal to the false rejection rate) [4]. This protocol is described in more detail in [15].

In the first experiment, we found the performance of the *DCT-mod2* approach for $K = 1$ and $K = 2$ with various windows (as described in Section 2). Results are presented in Table 1.

The second experiment is similar to the first; here we used the *DCT-mod3* approach. Results are presented in Table 2.

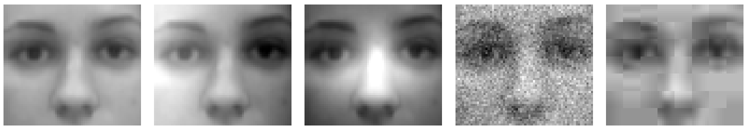

Fig. 1. From left to right: original image, corrupted with linear illumination change (δ=80), corrupted with Gaussian illumination change (δ=80), corrupted with white Gaussian noise (PSNR=22.5 dB), corrupted with compression artefacts (PSNR=31.4 dB).

In the final experiment we evaluated the *enhanced PCA* approach and compared it against the standard PCA approach without pre-processing and with histogram equalisation pre-processing. Results are presented in Table 3.

7 Discussion and Conclusions

As can be seen in Table 1, extending the *DCT-mod2* approach with $K = 2$ and various windows mainly causes worse performance (when compared to $K = 1$) for the case of Gaussian illumination change; these results indicate that in order to achieve robustness against illumination changes, delta coefficients [see Eqns. (2) & (3)] should only be calculated using directly neighbouring blocks; this is also suggested by the results for $K = 2$, where Window C obtains better results than Windows A & B (recall that for Window C the DCT coefficients from the outer blocks have smaller contribution). The results also show that *DCT-mod2* features are somewhat affected by compression artefacts and are significantly affected by white Gaussian noise.

DCT-mod3 features (Table 2) obtain comparable performance to *DCT-mod2* features on clean images and a improvement in the error rate for images corrupted by compression artefacts & white noise (especially for $K = 2$); however, it must be noted that for the case of white noise the performance is still quite poor. For images corrupted by either the linear or Gaussian illumination change, the performance is significantly worse than *DCT-mod2*, indicating that use of diagonal delta coefficients [see Eqns. (4) & (5)] is detrimental to robustness; further analysis is required to determine the cause.

In Table 3 we can see that the standard and *enhanced PCA* approaches are robust against white noise (in contrast to *DCT-mod2* and *DCT-mod3* approaches). We can also see that use of histogram equalisation as pre-processing for PCA increases the error rate in all cases, and most notably offers no help against illumination changes; this is in contrast to *enhanced PCA* which is significantly more robust against illumination

Type	**clean**	**lin. illum.**	**Gaus. illum.**	**white noise**	**compr.**
K=1	3.57	5.85	13.57	43.75	9.96
K=2, Win A	2.86	5.00	27.10	42.05	9.38
K=2, Win B	3.48	5.00	24.29	43.53	10.00
K=2, Win C	3.57	4.87	21.43	42.72	10.00

Table 1. Performance of *DCT-mod2* feature extraction. Results are quoted in terms of EER.

Type	**clean**	**lin. illum.**	**Gaus. illum.**	**white noise**	**compr.**
K=1	4.29	7.14	17.86	40.71	8.53
K=2, Win A	2.86	21.38	32.99	33.57	4.87
K=2, Win B	2.32	14.91	37.14	35.13	4.96
K=2, Win C	3.53	11.43	30.00	39.87	5.71

Table 2. Performance of *DCT-mod3* feature extraction. Results are quoted in terms of EER.

Type	**clean**	**lin. illum.**	**Gaus. illum.**	**white noise**	**compr.**
standard	3.57	27.14	32.19	3.57	3.57
hist. equ.	4.29	32.86	36.34	7.14	4.33
enhanced	5.31	7.14	18.57	5.67	6.03

Table 3. Performance of PCA based feature extraction. Results are quoted in terms of EER.

changes at the expense of slightly higher error rates for the case of clean images and images corrupted with white noise & compression artefacts.

While the additive white noise greatly distorts the image, the average pixel intensity remains largely the same. Thus the robustness of the PCA based approaches stems from the dot product operation [see Eqn. (7)], where a given face is projected onto an eigenface. The final dot product remains largely the same for both clean and corrupted images (similar reasoning can be applied for the case of images corrupted with compression artefacts). In contrast, *DCT-mod2* & *DCT-mod3* feature sets describe only a small section of the face and hence are easily affected by additive noise. Due to the diagonal delta coefficients, *DCT-mod3* feature set describes a slightly larger section of the face than *DCT-mod2* and is thus (slightly) more robust against white Gaussian noise.

Based on the obtained results it can be argued that out of all the presented feature extraction techniques, *enhanced PCA* is overall the most robust method.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence **19** (1997) 711–720.
2. Chen, L-F., Liao, H-Y., Lin, J-C., Han, C-C.: 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** (2001) 1393–1403.
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statistical Soc., Ser. B **39** (1977) 1–38.
4. Doddington, G.R., Przybycki, M.A., Martin, A.F., Reynolds, D.A.: The NIST speaker recognition evaluation - Overview, methodology, systems, results, perspective. Speech Communication **31** (2000) 225–254.
5. Duc, B., Fischer, S., Bigün, J.: Face Authentication with Gabor Information on Deformable Graphs. IEEE Trans. Image Processing **8** (1999) 504–516.
6. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, USA, 2001.
7. Eickeler, S., Müller, S., Rigoll, G.: Recognition of JPEG Compressed Face Images Based on Statistical Methods. Image and Vision Computing **18** (2000) 279–287.
8. Gonzales, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, 1993.
9. Koh, L.H., Ranganath, S., Venkatesh, Y.V.: An integrated automatic face detection and recognition system. Pattern Recognition **35** (2002) 1259–1273.
10. Lockie, M. (editor): Facial verification bureau launched by police IT group. Biometric Technology Today **10** (No. 3) (2002) 3–4.
11. Moon, H., Phillips, P.J.: Computational and performance aspects of PCA-based face-recognition algorithms. Perception **30** (2001) 303–321.
12. Reynolds, D., Quatieri, T., Dunn, R.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing **10** (2000) 19–41.
13. Samaria, F.: Face Recognition Using Hidden Markov Models. PhD Thesis, University of Cambridge, 1994.
14. Sanderson, C., Paliwal, K.K.: Polynomial Features for Robust Face Authentication. Proc. Intern. Conf. on Image Processing, Rochester, New York, 2002, pp. 997-1000 (Vol. 3).
15. Sanderson, C.: The VidTIMIT Database. IDIAP Communication 02-06, Martigny, Switzerland, 2002.
16. Soong, F.K., Rosenberg, A.E.: On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition. IEEE Trans. Acoustics, Speech and Signal Processing **36** (1988) 871–879.
17. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86.
18. Wallace, G.K.: The JPEG still picture compression standard. IEEE Trans. Consumer Electronics **38** (1992) xviii-xxxiv.