On Local Features for GMM Based Face Verification

Conrad Sanderson
National ICT Australia (NICTA), Canberra 2601, Australia
CSSIP, University of Adelaide, SA 5005, Australia
conradsand@ieee.org

Marc Saban
IDIAP Research Institute
Martigny, CH-1920, Switzerland
saban@idiap.ch

Yongsheng Gao
School of Microelectronic Engineering
Griffith University, QLD 4111, Australia
yongsheng.gao@griffith.edu.au

Abstract

It has been recently shown that local feature approaches to face verification are considerably more robust than holistic approaches, in terms of translations (caused by automatic face localization) and pose variations. In this paper we first investigate whether features based on **local** *Principal Component Analysis (LPCA) are more discriminative than features based on the 2D Discrete Cosine Transform (2D DCT). We also investigate several methods for modifying the two feature extraction techniques in order to counteract the effects of linear and non-linear illumination changes, without losing discriminative information. Results on the XM2VTS database show that when using a Bayesian classifier based on Gaussian Mixture Models (GMMs), the performances of 2D DCT and LPCA techniques are quite similar, suggesting that the 2D DCT technique is preferable due to its lower computational complexity. When using 8×8 blocks, modifying the 2D DCT and LPCA techniques by removing the first coefficient, which is the most affected by illumination changes, enhances robustness with little change in discrimination ability; removing further coefficients causes a noticeable reduction in performance on clean images and provides little gain in robustness. When using the 2D DCT with 16×16 blocks, the first three coefficients need to be removed in order to achieve good robustness. It is further shown that contrary to previously published results, the use of deltas of low-order coefficients (to alleviate performance losses caused by removing coefficients) can adversely affect robustness.*

1. Introduction

Face recognition systems (here we mean both identification and verification systems) are a particular type of biometric recognition systems. Applications include transaction authentication, surveillance, forensics and various forms of access control, such as immigration checkpoints and access to information [15], [19].

Many techniques have been proposed for face recognition; some examples are systems using Principal Component Analysis (PCA) based feature extraction [22], modular PCA [16], Elastic Graph Matching (EGM) [9], Hidden Markov Models (HMMs) [3], [11] and Gaussian Mixture Models (GMMs) [3], [18].

The abovementioned approaches differ in one major aspect: the degree of constraints placed on spatial relations between face features (such as the distance between the eyes and nose). In PCA based representation, the relations are rigid, meaning that translations or local deformations are not taken into account. In EGM and HMM based systems, the constraints are more relaxed, allowing for a degree of translations and local deformations. In GMM based systems, the constraints are very loose, resulting in good robustness to imperfect face localization [2] and pose changes [20].

Approaches which have relaxed constraints typically utilize local features (that is, features which describe only a *small part* of the face). This is in contrast to approaches with rigid constraints, which typically utilize holistic representations. For HMM and GMM based approaches, local features are often obtained by analyzing a face on a block by block basis. Feature extraction based on the 2D Discrete Cosine Transform (2D DCT) [13] or DCTmod2 [18] is usually applied to each block. In 2D DCT based feature extraction, a given block is decomposed in terms of *pre-defined* orthogonal basis functions. Following the approach used in image compression, low-order coefficients are retained and form a feature vector for each block [11].

In [18] it was shown that robustness to illumination changes can be achieved by removing the first three coefficients; however, this robustness came at the cost of reduction in discrimination performance. It was suggested that instead of throwing out the coefficients, they should be replaced with "deltas", which are differences between coefficients obtained from neighbouring blocks. The results showed that susceptibility to illumination changes was reduced without a corresponding degradation in discrimination performance. While the results in [18] look promising, the experiments had several limitations: (i) a relatively small database was used, (ii) the illumination change was linear in nature, and (iii) the classifier was not optimized for each configuration of the feature extractor, leading to a bias in the results.

In this paper we first evaluate the use of features based on *local* Principal Component Analysis (LPCA), where the basis functions are defined by training datums rather than being pre-defined like in the 2D DCT. Since the feature extraction technique would be specifically tuned for faces, we make the hypothesis that it should provide more discriminative features than the 2D DCT, and hence obtain higher performance.

Secondly, we investigate several methods for modifying the LPCA and 2D DCT feature extraction techniques in order to achieve robustness to illumination changes. Specifically, we investigate the effects of removing low-order coefficients (most likely to be affected by illumination changes), the effects of replacing low-order coefficients with deltas and also the use of deltas by themselves. The limitations of [18] are avoided by using a much larger database (295 persons), a non-linear illumination change (in addition to the linear illumination change), and properly optimizing the classifier in each experiment.

The rest of this paper is organized as follows. In Section 2, we describe the 2D DCT and LPCA feature extraction techniques as well as provide a brief description of deltas. Section 3 provides an overview of the GMM based classifier, while Section 4 is devoted to experiments and discussions. The main findings of the paper are summarized in Section 5.

2. Feature Extraction

In the feature extraction techniques described below, the initial analysis stage is the same: each face window is analyzed block by block; each block has a size of $N \times N$ pixels; unless stated otherwise, $N = 8$; the location of each block is advanced by 4 pixels, resulting in an overlap of neighbouring blocks by 50%[1]. The choice of N and the overlap is based on [11], where a 2D DCT based feature extraction was utilized.

2.1. 2D DCT

Each block, $b(x, y)$, where $x, y = 0, 1, ..., N-1$, is decomposed in terms of pre-defined orthogonal 2D DCT basis functions (see Fig. 1 for an example). The result is a $N \times N$coefficient matrix $C(u, v)$:

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

where $u, v = 0, 1, \cdots, N-1$, $\alpha(v) = 1/N$ for $v = 0$, $\alpha(v) = 2/N$ for $v = 1, 2, \cdots, N-1$ and

$$\beta(x,y,u,v) = \cos\ \frac{(2x+1)u\pi}{2N}\ \cos\ \frac{(2y+1)v\pi}{2N} \qquad (2)$$

The coefficients are ordered according to a zig-zag pattern (see Fig. 3 for an example), which reflects the amount of information stored in each coefficient [13] (i.e. lower order coefficients almost always contain more information). For a block located at (a, b), the baseline 2D DCT feature vector is composed of:

$$\mathbf{x}^{(a,b)} = \left[\ c_0^{(a,b)}\ \ c_1^{(a,b)}\ \ \ldots\ \ c_{M-1}^{(a,b)}\ \right]^T \qquad (3)$$

where $c_n^{(a,b)}$ denotes the n-th 2D DCT coefficient and M is the number of retained coefficients. For the case of N=8, M varies from 1 to 64, depending on the desired dimensionality reduction. If we follow examples from image compression [13], as much as 75% of the highest order coefficients (which represent high frequency information, and is often noise) can be omitted without adversely affecting image quality. Reducing the dimensionality has several advantages; firstly, a smaller dataset is required to adequately train a classifier [10]; secondly, the feature vectors should contain less noise, thus being more discriminative.

A useful aspect of 2D DCT based feature extraction is the ability to physically interpret the basis functions. As can be observed, the 0-th coefficient reflects the sum of pixel values in the block, and as such will be the most affected by any illumination changes. Some robustness could thus be achieved by simply removing it from each feature vector. It can also be observed that the following two coefficients, which represent the horizontal and vertical pixel intensity changes, respectively, also have the potential to be considerably affected by illumination changes.

2.2. Local PCA

As opposed to using Principal Component Analysis (PCA) for *holistic* representation (where processing one face results in one feature vector [22]), we shall apply a PCA based feature extraction technique to each block; we term this method as *local* PCA (LPCA).

The first step is to arrange the raw pixels from a given block into vector format; the pixels are arranged in the zig-zag pattern, as used in the 2D DCT technique. The choice of the pattern in this case is arbitrary; any consistent pattern is suitable. Let us denote the raw pixel vector for a block at (a, b) as $\mathbf{r}^{(a,b)}$. A feature vector, possibly with a lower dimensionality, is then obtained using:

$$\mathbf{x}^{(a,b)} = \mathbf{U}^T\ \ \mathbf{r}^{(a,b)} - \mathbf{r}_\mu \qquad (4)$$

In order to keep the complexity low and to retain the advantage of the GMM classifier being robust to translations of the face [2], the transformation matrix $\mathbf{U}$ and $\mathbf{r}_\mu$ have to be the same for all vectors (i.e. they cannot be dependent on which part of the face each raw pixel vector comes from). As such, $\mathbf{U}$ and $\mathbf{r}_\mu$ are found as follows. A set of training raw pixel vectors is collected from all training face windows; let us define this set as:

$$R = \{\ \mathbf{r}_i\ \}_{i=1}^{N_A} \qquad (5)$$

where the position superscripts have been omitted for clarity. The mean vector, $\mathbf{r}_\mu$, of set R is then found. A covariance matrix is then calculated as

$$\mathbf{C} = \frac{1}{N_A} \sum_{i=1}^{N_A} (\mathbf{r}_i - \mathbf{r}_\mu)(\mathbf{r}_i - \mathbf{r}_\mu)^T \qquad (6)$$

Matrix $\mathbf{U}$ is then formed as

$$\mathbf{U} = [\ \mathbf{e}_1\ \ \mathbf{e}_2\ \ \cdots\ \ \mathbf{e}_D\ ] \qquad (7)$$

where $\mathbf{e}_n$ is the n-th eigenvector of $\mathbf{C}$. The eigenvectors are ordered, in a descending manner, according to their corresponding eigenvalues; doing so defines orthogonal directions that account for the highest amount of variance. D has the following constraints: $D \leq N_A$ and $D \leq N^2$. If $D = N^2$, no dimensionality reduction occurs, and thus vector $\mathbf{x}^{(a,b)}$ represents a decorrelated version of the raw pixel vector $\mathbf{r}^{(a,b)}$.

The main difference between 2D DCT and LPCA based feature extraction is hence in the definition of the basis functions. They are pre-defined in the 2D DCT, while in LPCA they are *learned*. As such, LPCA basis functions should more representative of face blocks. Moreover, PCA based dimensionality reduction is optimal in a Mean Square Error (MSE) sense [23] (i.e. it preserves the most information), thus LPCA feature vectors could be of lower dimensionality than those from the 2D DCT based technique. A possible disadvantage of the LPCA approach is that the basis functions may not have

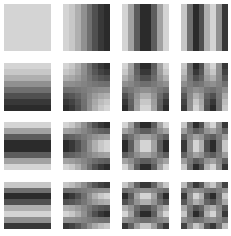

Fig. 1. Graphical interpretation of the first few 2D DCT basis functions for N=8; lighter colors represent larger values.

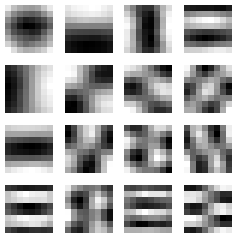

Fig. 2. Graphical interpretation of the first few LPCA basis functions for N=8, calculated on the training section of the XM2VTS database. Lighter colors represent larger values.

u \ v	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. 3. Zig-zag ordering of 2D DCT coefficients for N=4.

Fig. 4. *left*: original face window; *middle*: corrupted with the linear illumination change; *right*: corrupted with the non-linear illumination change; in both cases $\delta = 80$.

[1]For a 56×64 (rows$\times$columns) image, this results in 195 feature vectors.

an easily interpretable meaning in terms of image structures (as opposed to a statistical meaning). Moreover, the basis functions vary depending on which dataset is used for training. As such, throwing out specific elements from a feature vector (as opposed to reducing dimensionality) in order to achieve robustness to illumination changes may not be possible.

2.3. Delta coefficients

It has been previously shown [18] that on a relatively small database, and using a GMM based classifier with a low number of gaussians, simply throwing out the first three coefficients from 2D DCT based feature vectors increases robustness to illumination changes at the *expense* of reducing discrimination ability; this suggests that the first three coefficients are affected by illumination changes but contain a significant amount of discriminant information. To counteract this performance loss, it was proposed to replace (as opposed to throw out) the first few coefficients with their corresponding deltas, adapting a technique from speech processing [21].

The n-th *horizontal* and *vertical* delta coefficients for a block located at (a, b) are defined as a modified polynomial coefficients, respectively:

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

where $\mathbf{h}$ is a $2K$+1 dimensional symmetric window vector. Typically K=1 and a rectangular window is used (thus $\mathbf{h} = [\ 1.0\ 1.0\ 1.0\ ]^T$). Replacing the first three DCT coefficients by their horizontal and vertical deltas corresponds to the DCTmod2 feature extraction method:

$$\mathbf{x} = \left[\ \Delta^h c_0\ \Delta^v c_0\ \Delta^h c_1\ \Delta^v c_1\ \Delta^h c_2\ \Delta^v c_2\ \ c_3\ c_4\ \ldots\ c_{M-1}\ \right]^T \quad (8)$$

where the (a, b) superscript was omitted for clarity. The assumption in DCTmod2 is that the image corruption (e.g. an illumination change) is *constant* for the consecutive blocks that are used for calculating the deltas (i.e. it is locally constant). Under this assumption, the deltas reflect the information in the blocks which is not constant, effectively ignoring the illumination change.

It must be noted that utilizing deltas in a feature vector for a given block is only possible when the block has vertical and horizontal neighbours[2]. Moreover, the use of deltas effectively increases the area used when obtaining each feature vector. The increase is dependent on the amount of overlap; the smaller the overlap, the larger the effective spatial area. For a 50% overlap (i.e. 4 pixels), the effective width and height increase from 8 pixels to 8+4+4 = 16 pixels. However, since we are utilizing only horizonal and vertical deltas, the effective area increases from a total of 64 pixels to 192 pixels (rather than 256, which would result from a 16×16 block).

3. Classifier

Face verification can be treated as a two-class classification problem; the two classes correspond to the cases where a given face belongs to the claimed identity, or to an impostor. We utilize a Bayesian classifier based on Gaussian Mixture Models (GMMs). For each person, two GMMs are utilized: the first is a representative of the distribution of training vectors for that particular person's face, while the second is a representative of the distribution of training feature vectors for all training faces; the second GMM is commonly known as a generic model, a world model, or a universal background model [17].

Suppose that we have the following scenario. We are presented with a face image and also a claim that this face belongs to person C. To classify the face, a set of feature vectors, $X = \{\mathbf{x}_i\}_{i=1}^{N_V}$, is first extracted. By assuming that each vector is independent and identically distributed, the likelihood of the face belonging to person C is found with:

$$\mathcal{L}(X|\lambda_C) = \prod_{i=1}^{N_V} p(\mathbf{x}_i|\lambda_C) \quad (9)$$

where

$$p(\mathbf{x}|\lambda) = \sum_{g=1}^{N_G} w_g \, \mathcal{N}(\mathbf{x}, \mu_g, \mathbf{\Sigma}_g) \quad (10)$$

$$\lambda = \{w_g, \mu_g, \mathbf{\Sigma}_g\}_{g=1}^{N_G} \quad (11)$$

and $\mathcal{N}(\mathbf{x}; \mu, \mathbf{\Sigma})$ is a D-dimensional Gaussian function with mean μ and diagonal covariance matrix $\mathbf{\Sigma}$. λ_C is the parameter set for person C, N_G is the number of gaussians and w_g is the weight for Gaussian g (with constraints $\sum_{g=1}^{N_G} w_g = 1$ and $\forall\ g : w_g \geq 0$).

The generic model is then used to find the likelihood of the face belonging to an impostor, i.e. $\mathcal{L}(X|\lambda_{generic})$. An opinion on the face belonging to person C is found with:

$$\mathcal{O}(X) = \log \mathcal{L}(X|\lambda_C) - \log \mathcal{L}(X|\lambda_{generic}) \quad (12)$$

Note that in (12) we assumed non-informative prior probabilities of the two classes. The final decision for the given face is then reached as follows: given a threshold t, the face is classified as belonging to person C when $\mathcal{O}(X) \geq t$ and classified as belonging to an impostor when $\mathcal{O}(X) < t$.

Given a set of training vectors, the GMM parameters (λ) for each face are found by adapting the generic model using a form of Maximum *a Posteriori* (MAP) adaptation [12], [3]. The parameters for the generic model are found using the Expectation Maximization (EM) algorithm [10], [7] using information from all training faces. The higher the N_G, the more precise the model (assuming a large enough training dataset); moreover, even though diagonal covariance matrices are utilized, it is possible to model correlated datasets as long as $N_G \geq 2$ [17].

4. Evaluation

4.1. XM2VTS Database

The XM2VTS database [14] is composed of 295 subjects, which are divided into three types: 200 *clients*, 25 *evaluation impostors* and 70 *test impostors*. Each subject attended four recording sessions taken at one month intervals; during each session two images were taken. We used Config. I of the Lausanne Protocol [14], which further partitions the images into three disjoint sections: training, evaluation and testing.

For all experiments, the training section was utilized as a source of images for training the face models; the evaluation section was used for tuning classifier parameters (such as the number of gaussians and the threshold). Once the optimum parameters were found, the test section was used for final performance measurement.

In each experiment, the classifier was given a model of a client's face, images of that face and impostor faces; each given face was classified as either belonging to the client (i.e. a true face), or belonging to someone else (i.e. an impostor

[2] For a 56×64 image, and a 4 pixel overlap, this results in 143 vectors.

face). When using the evaluation section, the above procedure resulted in a total of 600 true face presentations and 40000 impostor face presentations. When using the test section, there was a total of 400 true face presentations and 112000 impostor face presentations.

4.2. Performance Measures

Verification systems make two types of errors: a False Acceptance (FA), which occurs when the system accepts an impostor face, or a False Rejection (FR), which occurs when the system refuses a true face. The performance is generally measured in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR), defined as:

$$\text{FAR} = \frac{\text{number of FAs}}{\text{number of impostor face presentations}} \quad (13)$$

$$\text{FRR} = \frac{\text{number of FRs}}{\text{number of true face presentations}} \quad (14)$$

To aid the interpretation of performance, the two error measures are often combined using the Half Total Error Rate (HTER), defined as HTER=(FAR+FRR)/2; the HTER is a special case of the Decision Cost Function [1], [8]. A special case of the HTER, known as the Equal Error Rate (EER), occurs when the system is adjusted (e.g. via tuning a threshold) so that FAR=FRR on a particular dataset.

4.3. Illumination Changes

In order to simulate illumination changes, we have applied (individually) two image transformations to each *test* face window. The first transformation is linear in nature, while the second is non-linear.

The linear illumination change simulates the effect of one half of the face being brighter than the other half. An original face window, $w(x,y)$, with N_X columns and N_Y rows, is corrupted to obtain a new face window, $v(x,y)$, using:

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

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

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

While these illumination changes are artificial and do not represent situations such as self-shadowing, we believe they are useful in providing suggestive results. Throughout the experiments δ was set to 80, representing quite challenging conditions. Fig. 4 shows the effects of the two illumination changes.

4.4. Experiments and Discussion

For the purposes of this study, we assumed that we are dealing with static frontal images and that each face has been correctly localized and size normalized (that is, the location of the eyes is the same in each image). Examples of face localization approaches can be found in [24]. To reduce the effects of intra-personal variations, *closely cropped* [4] greyscale face windows were extracted from original images; the size of each window is 56×64 (rows×columns) pixels (following [18]). An example face window is shown in Fig. 4.

The classifier parameters (number of gaussians and the threshold) were selected to minimize the EER on the evaluation set (i.e. the dataset which is *not* used for final performance measurement). The number of gaussians was varied from 1 to 512, doubling the number of gaussians in each step (e.g. 1, 2, 4, $\cdots$, 512). The threshold found on the evaluation section was used on the test section to obtain the final performance figure (i.e. in terms of HTER).

We evaluated the performance of the 2D DCT and LPCA feature extraction techniques on clean face images, as well as face images corrupted with the linear and non-linear illumination changes defined in Section 4.3. We also evaluated the effectiveness of several approaches to modifying the above mentioned feature extraction methods in order to increase robustness to illumination changes. These approaches are:

- Removing lower order coefficients (which represent basis functions that are most likely to be affected by illumination changes)
- Replacing lower order coefficients with their corresponding horizontal and vertical deltas
- Using only horizontal and vertical deltas

We first found the optimal dimensionality on the evaluation section of the database; this dimensionality was then used as a baseline for further experiments. Each dimensionality was based on the cumulative amount of coefficients along the diagonals traced by the zig-zag pattern (see Fig. 3 for an example).

The results in Table 1 suggest that when using blocks of size 8×8, the optimal dimensionality for both 2D DCT and LPCA is 21 (which amounts to keeping approx. 33% of the coefficients). The performances of the two techniques are quite similar, suggesting that the 2D DCT technique is to be preferred due to its lower complexity. The basis functions in 2D DCT are pre-defined while in LPCA they first have to be learned; moreover, at the best dimensionality, the 2D DCT based technique requires less gaussians than the LPCA based technique.

By comparing Figures 1 and 2, it can be seen that the first few LPCA basis functions are quite similar to the 2D DCT basis functions, partly explaining the similar performance of the two approaches. Moreover, the nature of the first three LPCA basis functions makes them susceptible to illumination changes, thus removing the corresponding coefficients from each vector should achieve a degree of robustness.

In the second experiment, we evaluated the effects of the linear and non-linear illumination changes. We also evaluated the effects of removing removing lower order coefficients. Tables 2 and 3 show the results for the 2D DCT and LPCA, respectively. The results show that the LPCA technique is

dim.	8×8 2D DCT			8×8 LPCA		
	best N_G	EER	HTER	best N_G	EER	HTER
1	4	31.83	26.12	4	31.67	26.12
3	128	17.23	13.94	128	18.16	14.04
6	256	12.99	10.83	256	12.33	10.66
10	256	8.17	6.96	512	6.71	7.83
15	256	5.67	5.08	256	6.33	5.20
∗ 21	256	**4.83**	4.91	512	**5.68**	**5.00**
28	256	5.01	**4.79**	512	5.93	5.12
36	256	5.46	**4.79**	128	6.16	5.54
43	128	6.16	6.17	128	6.33	5.78
49	128	6.34	6.42	256	6.98	6.45
54	256	6.66	5.78	128	7.66	7.16
58	256	6.85	6.14	128	7.67	6.79
61	256	6.50	6.20	128	8.03	7.11
63	256	6.83	6.97	128	7.49	6.74
64	256	7.50	7.25	128	7.69	6.99

Table 1. Performance of 2D DCT and LPCA based feature extraction techniques for varying dimensionality. "best N_G" indicates the number of gaussians which achieves the lowest EER on the validation set. The HTER is then calculated on the test set.

dim.	modified 8×8 2D DCT				
	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER
21 (baseline)	256	**4.83**	4.91	8.61	9.86
21 - 1	256	5.17	**4.37**	**4.76**	**6.29**
21 - 3	256	7.50	6.50	6.34	6.78
21 - 6	256	10.17	8.12	8.77	8.68
21 - 10	128	15.00	12.06	12.50	12.70

Table 2. Performance of modified **2D DCT** based method on clean faces and faces corrupted with the linear and non-linear illumination changes. The method was modified by removing elements from the *start* of the 21 dimensional baseline feature vectors.

dim.	modified 8×8 LPCA				
	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER
21 (baseline)	512	5.68	5.00	13.68	11.29
21 - 1	512	**5.50**	**4.09**	**6.52**	**8.53**
21 - 3	256	7.83	6.38	7.01	8.68
21 - 6	512	10.02	8.65	8.99	9.38
21 - 10	512	14.67	12.39	13.09	12.95

Table 3. As per Table 2, but using **LPCA** based feature extraction.

somewhat more affected by illumination changes than the 2D DCT method. For both 2D DCT and LPCA, removing the first coefficient from each feature vector considerably enhances robustness to illumination changes, with little effect on the performance on clean images. Removing more than the first coefficient causes a noticeable reduction in performance on clean images and provides little gain in robustness.

In the third experiment we evaluated the effects of replacing coefficients (as opposed to throwing them out) with their corresponding horizontal and vertical deltas. By comparing Tables 2 and 4, it can be observed that the use of deltas of the 0-th coefficient has little effect on the performance on clean faces and considerably increases the error rates on faces corrupted by illumination changes. This can be partly explained by the breakdown of the assumption of locally constant illumination changes (as described in Sec. 2.3). Compared to throwing out the first three coefficients, using deltas of the first three coefficients results in an improvement in performance on clean faces, while achieving similar performance on corrupted faces (implying that in this case the GMM based classifier effectively tolerates the non-robustness of the deltas of the 0-th coefficient). While not shown here, the results of this experiment for LPCA are very similar to 2D DCT.

dim.	modified 8×8 2D DCT + deltas				
	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER
21 (baseline)	256	4.83	4.91	8.61	9.86
21 - 1 + 2	256	5.33	4.68	7.34	17.98
21 - 3 + 6	128	4.51	4.56	5.08	6.01
21 - 6 + 12	256	**4.50**	4.75	5.11	6.62
21 - 10 + 20	256	4.67	**4.17**	**4.49**	**5.93**

Table 4. Performance of modified **2D DCT** based method on clean faces and faces corrupted with the linear and non-linear illumination changes. The baseline method was modified by *replacing* the elements from the *start* of the 21 dimensional baseline vectors with their corresponding horizontal and vertical deltas.

dim.	8×8 2D DCT deltas only				
	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER
2 (1+1)	32	14.02	12.72	27.11	46.20
6 (3+3)	128	5.33	5.90	9.43	30.44
12 (6+6)	512	**3.83**	4.23	5.66	14.43
20 (10+10)	512	4.16	**4.01**	**4.78**	**7.18**

Table 5. Performance of 2D DCT based method on clean faces and faces corrupted with the linear and non-linear illumination changes. The baseline method were modified by keeping only a specified amount of horizontal and vertical deltas.

In the fourth experiment we appraised the performance and robustness of feature vectors which contain only horizontal and vertical deltas. The results in Table 5 confirm that deltas of the first element from 2D DCT vectors are considerably affected by illumination changes. As more deltas are utilized, the performance and robustness increases, suggesting that only deltas of higher order coefficients are useful. Again, while not shown here, the results of this experiment for LPCA are very similar to 2D DCT.

As mentioned in Section 2.3, one of the effects of using deltas is an increase in the effective area used when obtaining each feature vector. The results from the third experiment suggest that performance can be increased through the use of deltas, implying that the use of a larger block size may be beneficial. Instead of using the indirect method of deltas to increase the area, in the fifth experiment we evaluated the performance and robustness of feature vectors derived from 2D DCT using 16×16 blocks (compared to 8×8 in previous experiments). The location advance of each 16×16 block is the same as for 8×8 blocks (i.e. 4 pixels), resulting in an overlap of neighbouring blocks by 75%. Results in Table 6 suggest that the optimum baseline dimensionality is 21, which is the same as for 8×8 blocks; moreover, the performance on clean faces is slightly better than for 8×8 blocks.

In the final experiment, we evaluated the effects of the two illumination changes on the performance of the 16×16 2D DCT based feature extraction technique. As for 8×8 blocks, we also evaluated the effects of removing low-order coefficients. The results in Table 7 show that removing just the first coefficient is insufficient to achieve robustness to non-linear illumination changes. Good robustness is achieved by removing the first three coefficients, though it comes at the cost of a small performance degradation on clean images. Removing more coefficients causes a noticeable reduction in performance on clean images, with little change in the robustness. By comparing Tables 2 and 7 it can be observed that the performance and robustness of 16×16 2D DCT with the first three coefficients removed (resulting in 18 dimensional vectors) is similar to the performance of 8×8 2D DCT with the first coefficient removed (i.e. 20 dimensional vectors).

dim.	16×16 2D DCT		
	best N_G	**EER**	**HTER**
1	2	31.67	31.39
3	256	20.00	16.26
6	256	12.67	10.65
10	256	6.33	6.64
15	512	4.34	4.22
21	256	**4.00**	**4.02**
28	256	4.67	4.49
36	256	5.00	4.53
66	128	6.00	6.02
136	128	8.99	7.79
256	64	12.17	12.61

Table 6. Performance of 16×16 2D DCT based method on clean faces.

dim.	modified 16×16 2D DCT				
	clean			linear	non-lin.
	best N_G	**EER**	**HTER**	**HTER**	**HTER**
21 (baseline)	256	4.00	**4.02**	**5.06**	8.99
21 - 1	256	**3.87**	4.34	5.10	8.81
21 - 3	256	5.03	5.05	5.28	**5.42**
21 - 6	256	7.51	6.81	7.14	7.50
21 - 10	512	10.17	8.91	9.40	10.01

Table 7. Performance of modified 16×16 2D DCT based method on clean faces and faces corrupted with the linear and non-linear illumination changes. The dimensionality was reduced by removing elements from the *start* of the 21 dimensional baseline feature vectors.

5. Conclusions

In the context of a face verification system utilizing local features, we first investigated whether features based on *local* Principal Component Analysis (LPCA) are more discriminative than features based on the 2D Discrete Cosine Transform (2D DCT). As opposed to holistic feature extraction techniques, local features describe only a small part of the face. The evaluation was performed in terms of discrimination ability and robustness to linear and non-linear illumination changes. We also investigated several methods of modifying the two feature extraction techniques in order to increase robustness to illumination changes; these are: removal of coefficients which are deemed to be most affected by illumination changes, replacing coefficients with deltas (to alleviate performance losses caused by removing coefficients) and using only deltas.

Results on the XM2VTS database show that when using a Gaussian Mixture Model (GMM) based classifier, and a block size of 8×8, the performances of 2D DCT and LPCA techniques are similar, suggesting that the 2D DCT technique is to be preferred due to its lower computational complexity. The basis functions in 2D DCT are pre-defined while in LPCA they first have to be learned.

Modifying the 2D DCT and LPCA techniques by removing the first coefficient, which is the most affected by illumination changes, clearly enhances robustness. When utilizing analysis blocks of size 8×8, removing further coefficients causes a noticeable reduction in performance on clean images and provides little gain in robustness. When using the 2D DCT with 16×16 blocks, the first three coefficients need to be removed in order to achieve good robustness.

The experiments further show that contrary to previously published results, deltas of low order coefficients are considerably affected by illumination changes. In particular, the new results strongly suggest that deltas of the 0-th coefficient should never be used. This is attributed to the breakdown of the assumption of locally constant illumination changes, as used in the definition of the deltas.

Acknowledgements

The authors thank S. Bengio and J. Mariéthoz (IDIAP) for useful suggestions and the Swiss National Science Foundation for supporting this work through the National Center of Competence in Research (NCCR) on Interactive Multi-Modal Information Management (IM2). National ICT Australia is funded by the Australian Government's Backing Australia's Ability initiative, in part through the Australian Research Council. The implementation of the experiments was aided by the Newmat C++ matrix library [6] and the Torch machine learning library [5].

References

[1] S. Bengio, J. Mariéthoz, "The Expected Performance Curve: a New Assessment Measure for Person Authentication", *Proc. Odyssey 2004: The Speaker and Language Recognition Workshop*, 2004, pp. 279-284.

[2] F. Cardinaux, C. Sanderson, S. Marcel, "Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS", In: *Proc. 4th Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Guildford, 2003, pp. 911-920.

[3] F. Cardinaux, C. Sanderson, S. Bengio, "Face Verification Using Adapted Generative Models", In: *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition (AFGR)*, Seoul, 2004, pp. 825-830.

[4] L-F. Chen et al., "Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof", *Pattern Recognition*, Vol. 34, No. 7, 2001, pp. 1393-1403.

[5] R. Collobert, S. Bengio, J. Mariéthoz, "Torch: a modular machine learning software library", IDIAP Research Report 02-46, Martigny, Switzerland, 2002. Available at www.idiap.ch

[6] R. Davies, Newmat C++ matrix library, Available at www.robertnz.net

[7] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Royal Statistical Soc., Ser. B*, Vol. 39, No. 1, 1977, pp. 1-38.

[8] G.R. Doddington et al., "The NIST speaker recognition evaluation - Overview, methodology, systems, results, perspective", *Speech Communication*, Vol. 31, No. 2-3, 2000, pp. 225-254.

[9] B. Duc, S. Fischer, J. Bigün, "Face Authentication with Gabor Information on Deformable Graphs", *IEEE Trans. Image Processing*, Vol. 8, No. 4, 1999, pp. 504-516.

[10] R. Duda, P. Hart, D. Stork, *Pattern Classification*, John Wiley & Sons, USA, 2001.

[11] S. Eickeler, S. Müller, G. Rigoll, "Recognition of JPEG Compressed Face Images Based on Statistical Methods", *Image and Vision Computing*, Vol 18, No. 4, 2000, pp. 279-287.

[12] J.-L. Gauvain, C.-H. Lee, "Maximum *a Posteriori* Estimation for Multivariate Gaussian Mixture Observations of Markov Chains", *IEEE Trans. Speech and Audio Processing* Vol. 2, No. 2, 1994, pp. 291-298.

[13] R. Gonzales, R. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Massachusetts, 1992.

[14] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, "XM2VTSDB: The Extended M2VTS Database", In: *Proc. 2nd Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Washington, D.C., 1999, pp. 72-77.

[15] J. Ortega-Garcia, J. Bigun, D. Reynolds, J. Gonzales-Rodriguez, "Authentication gets personal with biometrics", *IEEE Signal Processing Magazine*, Vol. 21, No. 2, 2004, pp. 50-62.

[16] A. Pentland, B. Moghaddam, T. Starner, "View-Based and Modular Eigenspaces for Face Recognition", In: *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, Seattle, 1994, pp. 84-91.

[17] D. Reynolds, T. Quatieri, R. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models", *Digital Signal Processing*, Vol. 10, No. 1-3, 2000, pp. 19-41.

[18] C. Sanderson, K.K. Paliwal, "Fast features for face authentication under illumination direction changes", *Pattern Recognition Letters*, Vol. 24, No. 14, 2003, pp. 2409-2419.

[19] C. Sanderson, K.K. Paliwal, "Identity verification using speech and face information", *Digital Signal Processing*, Vol. 14, No. 5, 2004, pp. 449-480.

[20] C. Sanderson, S. Bengio, "Extrapolating Single View Face Models for Multi-View Recognition", In: *Proc. Int. Conf. Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Melbourne, 2004, pp. 581-586.

[21] F. Soong, A. Rosenberg, "On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition", *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 36, No. 6, 1988, pp. 871-879.

[22] M. Turk, A. Pentland, "Eigenfaces for Recognition", *J. Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

[23] X. Wang, *Feature Extraction and Dimensionality Reduction in Pattern Recognition and Their Application in Speech Recognition*, PhD Thesis, Griffith University, Australia, 2002. Available at http://adt.caul.edu.au

[24] M.-H. Yang, D. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, 2002, pp. 34-58.