IDIAP RESEARCH REPORT

On Local Features for Face Verification

Marc Saban [a] Conrad Sanderson [b]

IDIAP–RR 04-36

June 2004

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

(a) **saban @ idiap.ch**; this work was performed during Marc Saban's internship at IDIAP; Marc Saban is currently completing an engineering degree at the Eurecom Institute, Sophia Antipolis, France.
(b) **conradsand @ ieee.org**; Dept. Electrical and Electronic Engineering, University of Adelaide, SA 5005, Australia.

IDIAP Research Report 04-36

ON LOCAL FEATURES FOR FACE VERIFICATION

Marc Saban Conrad Sanderson

JUNE 2004

Abstract. We compare four local feature extraction techniques for the task of face verification, namely (ordered in terms of complexity): raw pixels, raw pixels with mean removal, 2D Discrete Cosine Transform (DCT) and *local* Principal Component Analysis (PCA). The comparison is performed in terms of discrimination ability and robustness to illumination changes. We also evaluate the effectiveness of several approaches to modifying standard feature extraction methods in order to increase performance and robustness to illumination changes. Results on the XM2VTS database suggest that when using a Gaussian Mixture Model (GMM) based classifier, the raw pixel technique provides poor discrimination and is easily affected by illumination changes; the mean removed raw pixel technique provides performance that is fairly close to 2D DCT and local PCA, but is considerably affected by illumination changes. The performance of 2D DCT and local PCA techniques is quite similar, suggesting that the 2D DCT technique is to be preferred over the local PCA technique, due to the lower complexity of the 2D DCT. Both 2D DCT and local PCA techniques are considerably more robust to illumination changes compared to the raw pixel techniques. Modifying the 2D DCT and local PCA techniques by removing the first coefficient, which is deemed to be the most affected by illumination changes, clearly enhances robustness; removing more than the first coefficient causes a noticeable reduction in performance on clean images and provides no further gains in robustness. Compared to just throwing out the first coefficient, the use of deltas can achieve a small increase in performance and robustness. Lastly, we suggest that it is more appropriate to use analysis blocks of size 8×8 (as opposed to 16×16) with 2D DCT decomposition; out of the 64 resulting coefficients, the second through to 21-st (resulting in 20 dimensional feature vectors) are the most robust to illumination changes while providing good discriminatory information.

Contents

List of Figures

List of Tables

1 Introduction

Several fields related to security have received special attention the past few years, especially since the unpleasantness in New York in 2001; biometric identity verification is one of these. The aim of identity verification is to discriminate between two cases: either the person claiming a given identity is the true claimant or the person is an impostor. This is not to be confused with automatic identification, which, in the closed set case, consists in determining the identity of a given person out of a set of people. Both biometric verification and identification systems can be considered to fall in the general area of biometric person recognition.

Apart from security applications, verification systems can also be used in transaction authentication, secure teleworking and forensics. We are already using many verification systems in our every day life. For example, Automatic Teller Machines (ATMs) utilize a basic identity verification; the user claims his/her identity through the presentation of his/her card; the verification is accomplished by asking to enter a password, known by the user. This is in fact a combination of something that the user has and something that he/she knows. The problem is that the security of this scheme can be compromised quite easily.

Several biometric verification systems have been proposed in order to augment (or replace) the above card/password mechanism; these systems don't attempt to check what you know, but rather who you are. Various methods of biometric verification have been proposed in the literature; this includes the use of fingerprints, iris scans, voices, faces and palm prints. A couple of these are already used in several international airports, such as Schipol in Amsterdam or JFK in New York [24]. Further introductory and review material about the biometrics field can be found in [5, 18, 32, 38, 45].

In this report we shall focus on face verification. The main perceived advantage of using faces, compared to other methods, is that this verification approach can be largely *non-intrusive*; in other words, it can require little or no collaboration on the user's part to be effective [32]. A complete appearance based face verification system can be decomposed into several steps:

1. *Localization*: here the position of the face is found;
2. *Normalization*: this step usually involves an geometric transformation to correct the size and rotation, and/or an illumination normalization;
3. *Feature extraction*: information relevant to discrimination purposes is extracted;
4. *Classification*: information from the previous step is compared against one or more models (also known as templates) and a decision on the claim is reached.

For the purposes of this study, we shall assume that we are dealing with static frontal images and that the face is perfectly localized. Thus, we will concentrate on the last two steps: feature extraction and classification. For information about face localization the reader is directed to the following publications: [11, 47, 46].

Various approaches to appearance based face recognition (here we mean both identification and verification) have been investigated; they can be roughly divided into *holistic* and *non-holistic* (i.e. *local feature*) approaches. Examples of holistic approaches include systems based on Principal Component Analysis (PCA) feature extraction [42], and Linear Discriminant Analysis (LDA) [2]. Examples of non-holistic approaches include systems based on modular-PCA [33], Elastic Graph Matching (EGM) [16, 26]. 1D Hidden Markov Models (HMMs) [35], pseudo-2D HMMs [19, 31] and Gaussian Mixture Models (GMMs) [8, 39, 7]. As an in-depth review of face recognition literature is beyond the scope of this report, the reader is directed to the following review articles: [9, 23, 25, 48].

Local feature extraction can be based on the use raw pixels [35], 2D-Discrete Cosine Transform (2D-DCT) [22, 39] and Gabor wavelets [27]. Recently it has been shown that systems based on a combination of featured derived from 2D DCT coefficients and a GMM classifier are relatively robust to out-of-plane rotations [6] and to translation errors made by the face localization stage [8].

In this report we compare the performance of four local feature extraction techniques on clean face images, as well as face images corrupted with artificial linear and non-linear illumination changes; we also evaluate the effectiveness of several approaches to modifying standard feature extraction methods in order to increase performance and robustness to illumination changes. The four feature extraction techniques are: raw pixels, raw pixels with mean removal, 2D DCT and local PCA. In all experiments we shall utilize a GMM based classifier.

The rest of the report is organized as follows. The XM2VTS database is summarized in Section 2; performance measures are presented in Section 3; Section 4 describes two artificial image corruption techniques; in Section 5 we provide an overview of the four local feature extraction techniques as well as provide a brief description of deltas (a method used in modifying standard feature extraction); in Section 6 we provide an overview of the Gaussian Mixture Model based classifier; Section 7 is devoted to experiments and discussions. Conclusions and suggestions are given in Section 8.

2 XM2VTS database

The XM2VTS database [30] is composed of 295 subjects, which are divided into three sets: 200 clients, 25 evaluation impostors and 70 test impostors. Each subject attended four recording sessions taken at one month intervals (hence there is intra-personal variability, such as different expressions, hair-cuts and make-up).

XM2VTS is divided into three sets with respect to the *Lausanne Protocol* (LP) [28]: a training set, an evaluation set and a test set. Two configurations are defined in the LP, as shown on Figure 1. For all our experiments, Configuration I was used, leading to the following setup:

Training examples per client:	3
Evaluation client accesses:	600
Evaluation impostor accesses:	40,000 ($25 \times 8 \times 200$)
Test client accesses:	400 (200×2)
Test impostor accesses:	112,000 ($70 \times 8 \times 200$)

The training set is used to train the Universal Background Model (UBM) (explained in Section 6), as well as the client models derived from the UBM. The evaluation set was used to tune various classifier hyper-parameters, such as the number of gaussians and the decision threshold. Finally, the test set was used to measure the performance of the system.

To reduce the effects of intra personal variations, *closely cropped* [10, 36] greyscale face windows were extracted from original images. In each face window the location of the eyes is fixed; the size of the window is 56×64 (rows$\times$columns) pixels. An example of an image from the database as well as the corresponding face window is shown in Figure 2.

2.1 Caveats

The XM2VTS database was designed for research and development of systems where one assumes that the client will be cooperative and where the illumination conditions are controlled. Examples of more challenging databases are the PIE [40] and BANCA [1] databases. While the PIE database contains illumination changes, they were simulated with a flash system and are hence artificial in nature; moreover, the PIE database contains images of only 68 subjects which were taken in only one session; apart from expression changes, there is no other intra-personal variation. The BANCA databases includes more realistic illumination changes as well as intra-personal variations, however, its experiment protocols specify that only 52 subjects (out of 208) can be used in one experiment at a time.

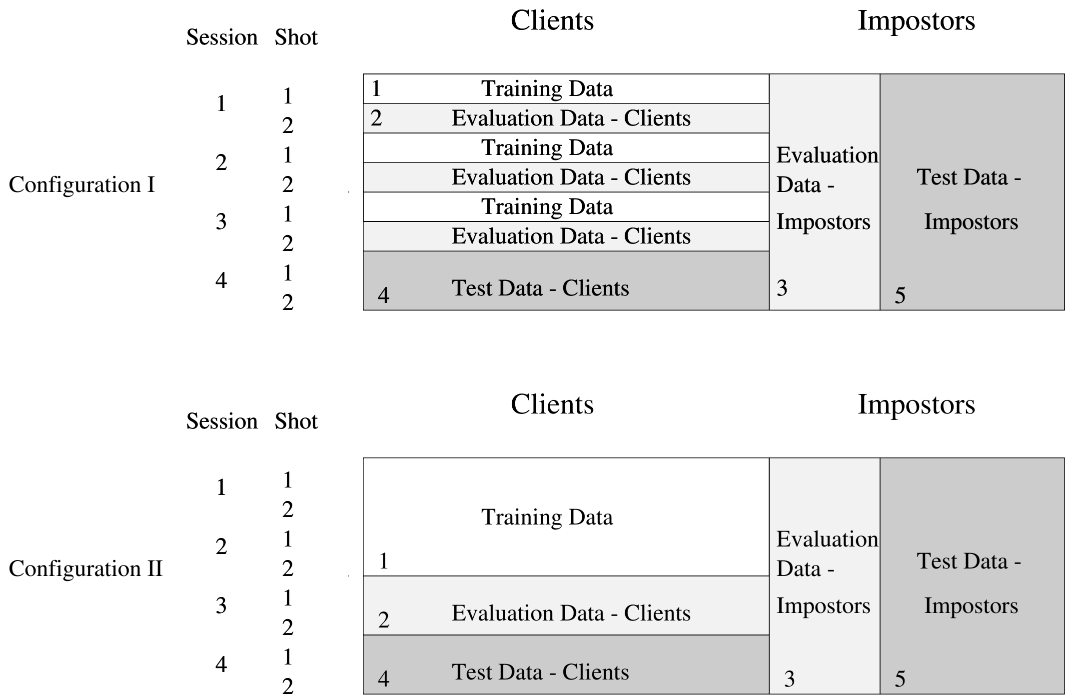

Figure 1: Partitioning of the XM2VTS database according to Lausanne protocol configuration I (top) and II (bottom).

Figure 2: Example of an image from the XM2VTS database (left) and the corresponding face window (right).

3 Performance measures

There are two types of errors that can occur in an identity verification system: a *false acceptance* (FA), which occurs when the system accepts an *impostor*, or a *false rejection* (FR), which occurs when the system refuses a *true client*. The performance of verification systems is generally measured in terms of *false acceptance rate* (FAR) and *false rejection rate* (FRR), defined as follows:

$$\text{FAR} = \frac{\text{number of FAs}}{\text{number of impostor accesses}} \tag{1}$$

$$\text{FRR} = \frac{\text{number of FRs}}{\text{number of client accesses}} \tag{2}$$

To aid the interpretation of performance, the two error measures are often combined into one measure, called the Half Total Error Rate (HTER):

$$\text{HTER} = \frac{\text{FAR} + \text{FRR}}{2} \tag{3}$$

The HTER is a particular case of the Decision Cost Function (DCF) [4, 15]:

$$\text{DCF} = \text{Cost(FR)} \cdot P(\text{client}) \cdot \text{FRR} \ + \ \text{Cost(FA)} \cdot P(\text{impostor}) \cdot \text{FAR} \tag{4}$$

where $P(\text{client})$ is the prior probability that a client will use the system, $P(\text{impostor})$ is the prior probability that an impostor will use the system, Cost(FR) is the cost of a false rejection and Cost(FA) is the cost of a false acceptance. For the HTER, we have $P(\text{client})$=$P(\text{impostor})$=0.5 and the costs are set to 1.

It is often impossible to get perfect performance (that is, both FAR and FRR are zero). Thus, there is a choice to be done: do we prefer a smaller FAR and a larger FRR, or the opposite? For high security needs, it may be preferable to have a FAR as low as possible.

Apart from expressing the performance in terms of HTER or DCF, the performance, in terms of FAR and FRR, can be visualized using a *receiver operating characteristic* (ROC) curve [43], or the the *detection error trade-off* (DET) curve [29], which is a non-linear version of the ROC curve; every point of these curves corresponds to a given decision threshold. Note that in these curves each threshold is found on *test* data, thus optimistically biasing the resultant performance measurement. Recently, a more appropriate graphical representation, called the *expected performance curve*, has been proposed [3].

4 Artificial corruption techniques

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, $w(y,x)$, is corrupted to obtain a new face window, $v(y,x)$, of dimensions $N_X = 64$ and $N_Y = 56$ in our case, using:

$$\begin{aligned} v(y,x) &= w(y,x) + mx + \delta \\ \text{for} \;\; y &= 0,1,...,N_Y-1 \quad \text{and} \quad x = 0,1,...,N_X-1 \\ \text{where} \;\; m &= \frac{-\delta}{(N_X-1)/2} \\ \delta &= \text{illumination delta (in pixels)} \end{aligned} \tag{5}$$

Since the above model of illumination direction change is rather restrictive, a second, gaussian shaped (non-linear), artificial illumination change, simulating a spot-light in the middle of the face, was also used:

$$\begin{aligned} v(y,x) &= w(y,x) + 2\delta \left(\exp\left[\frac{-1}{2} \vec{p}^{\,T} \mathbf{A}^{-1} \, \vec{p} \right] - \frac{1}{2} \right) \\ \text{for} \;\; y &= 0,1,...,N_Y-1 \quad \text{and} \quad x = 0,1,...,N_X-1 \\ \text{where} \;\; \vec{p} &= [\, y \;\; x \,]^T \; - \; [\, (N_Y-1)/2 \;\; (N_X-1)/2 \,]^T \\ \mathbf{A} &= \begin{bmatrix} (N_Y/4)^2 & 0 \\ 0 & (N_X/4)^2 \end{bmatrix} \\ \delta &= \text{illumination delta (in pixels)} \end{aligned} \tag{6}$$

Figure 3 shows the effects of the two illumination changes. While these illumination changes are artificial and do not represent situations such as shadowing, we believe they are useful in providing suggestive results.

Figure 3: *left*: original face window; *middle*: corrupted with the linear illumination change; *right*: corrupted with the non-linear illumination change; in both cases $\delta = 80$.

5 Feature Extraction

In all the feature extraction techniques described below, the initial analysis stage is the same: each face window is analyzed a block-by-block basis; each block is $N \times N$ pixels; unless stated otherwise, $N = 8$; the location of each block is advanced by 4 pixels, resulting in an overlap of neighboring blocks by 50%[1]. The choice of N and the overlap is based on [19], where a 2D DCT based feature extraction was utilized.

5.1 Raw pixels

We start off with a "naive" feature extraction technique, which essentially packs local raw pixel values into a feature vector. the pixels from a given block are arranged in a zig-zag pattern, which is the same as used in the 2D DCT based technique, described in Section 5.2 (however, note that when dealing with raw pixels, any consistent pattern is suitable). For a block located at (b, a), the raw pixel feature vector is composed of:

$$\vec{x}^{(b,a)} = \left[\, p_0^{(b,a)} \; p_1^{(b,a)} \; \ldots \; p_{N^2-1}^{(b,a)} \, \right]^T \tag{7}$$

where $p_n^{(b,a)}$ is the n-th pixel value according to the zig-zag pattern. We shall term this technique as *raw pixel* feature extraction. There are several main drawbacks to this technique; firstly, the vector elements can be highly correlated; secondly, as a side-effect of the high correlation, an illumination change has the potential to affect all the elements.

We propose to remove some correlation between the elements of each raw pixel feature vector can by subtracting the mean from each element; this has a beneficial side-effect: the mean removal can act be interpreted as a form of illumination normalization. Formally, the mean removed raw pixel feature vector is composed of:

$$\vec{x}^{(b,a)} = \left[\, p_0^{(b,a)} - p_\mu^{(b,a)} \;\; p_1^{(b,a)} - p_\mu^{(b,a)} \;\; \cdots \;\; p_{N^2-1}^{(b,a)} - p_\mu^{(b,a)} \, \right]^T \tag{8}$$

where

$$p_\mu^{(b,a)} = \frac{1}{N^2} \sum_{i=0}^{N^2-1} p_i^{(b,a)} \tag{9}$$

We shall term this method as *mean removed raw pixel* feature extraction.

[1] For a face window which has N_Y rows and N_X columns, there are usually $(2\frac{N_Y}{N} - 1) \times (2\frac{N_X}{N} - 1)$ blocks; hence for a 56×64 (rows$\times$columns) window, there are usually 195 feature vectors.

5.2 2D DCT

Each block, $b(y,x)$, where $y,x = 0,1,...,N-1$, is decomposed in terms of pre-defined orthogonal 2D DCT basis functions (see Figure 4 for an example). The result is a N×N matrix $C(v,u)$ containing 2D DCT coefficients:

$$C(v,u) = \alpha(v)\alpha(u) \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} b(y,x)\beta(y,x,v,u) \quad \text{for } v,u = 0,1,2,...,N-1 \tag{10}$$

where

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } v = 0 \\ \sqrt{\frac{2}{N}} & \text{for } v = 1,2,...,N-1 \end{cases} \tag{11}$$

and

$$\beta(y,x,v,u) = \cos\left[\frac{(2y+1)v\pi}{2N}\right] \cos\left[\frac{(2x+1)u\pi}{2N}\right] \tag{12}$$

The coefficients are ordered according to a zig-zag pattern, an example of which is given in Figure 5; the zig-zag pattern reflects the amount of information stored in each coefficient [22], with lower order coefficients deemed to contain more information. For a block located at (b,a), the baseline 2D-DCT feature vector is composed of:

$$\vec{x}^{(b,a)} = \left[\ c_0^{(b,a)} \ \ c_1^{(b,a)} \ \ ... \ \ c_{M-1}^{(b,a)}\ \right]^T \tag{13}$$

where $c_n^{(b,a)}$ 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.

Compared to the raw pixel feature extraction technique, an obvious advantage of this feature extraction is thus the ability to reduce the dimensionality; if we follow examples from image compression [22] as much as 75% of the highest order coefficients (which represent high frequency information, which is often noise) can be omitted without adversely affecting image quality. Reducing the dimensionality has several advantages; firstly, less data is required to adequately train a classifier [17]; secondly, the feature vectors should contain less noise, thus being more discriminative.

Another advantage of the 2D DCT based feature extraction technique is the ability to physically interpret the basis functions; as can be observed, the 0-th coefficient will be the most affected by any illumination change, thus simply removing it from the feature vector can result in some robustness. It can also be argued that the following two coefficients, due to the nature of the corresponding basis functions, would also be significantly affected by illumination changes.

5.3 Local PCA

As opposed to using Principal Component Analysis (PCA) for holistic representation (as in [42]), we shall apply a PCA based feature extraction technique to each block; we shall term this method as *local PCA*.

The first step is exactly the same as for the raw pixel feature extraction. Let us denote the feature vector resulting from raw pixel feature extraction for a block at (b,a) as $\vec{r}^{(b,a)}$; a new feature vector, possibly with a lower dimensionality, is then obtained using:

$$\vec{x}^{(b,a)} = U^T \left(\vec{r}^{(b,a)} - \vec{r}_\mu \right) \tag{14}$$

In order to keep the complexity low and to retain the advantage of the GMM classifier being robust to translations of the face [8], the transformation matrix U and $\vec{r}_\mu$ have to be the same for all vectors (i.e. they

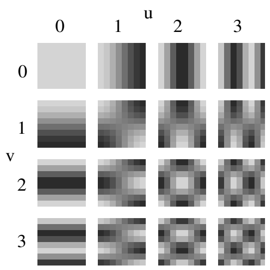

Figure 4: Several 2D-DCT basis functions for N=8; lighter colors represent larger values.

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

Figure 5: Example of coefficient ordering according to the zig-zag pattern for $N = 4$ (i.e. coefficients from a 4×4 block).

cannot be dependent on which part of the face each raw pixel vector comes from). As such, U^T and $\vec{r}_\mu$ are found as follows. A set of training raw pixel feature vectors is collected from all training face windows; let us define this set as:

$$R = \{ \, \vec{r}_i \, \}_{i=1}^{N_A} \tag{15}$$

where the position superscripts have been omitted for clarity. The mean vector of set F is then found, which we will denote as $\vec{r}_\mu$. A covariance matrix is then calculated:

$$C = \frac{1}{N_A} \sum_{i=1}^{N_A} (\vec{r}_i - \vec{r}_\mu)\,(\vec{r}_i - \vec{r}_\mu)^T \tag{16}$$

Matrix U is then formed:

$$U = [\ \vec{e}_1 \ \ \vec{e}_2 \ \ \cdots \ \ \vec{e}_D \ ] \tag{17}$$

where $\vec{e}_n$ is the n-th eigenvector of 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 $D = N^2$ then no dimensionality reduction occurs; in that case, vector $\vec{x}^{(b,a)}$ represents a decorrelated version of the raw pixel vector $\vec{r}^{(b,a)}$.

The main difference between 2D DCT based representation and the local PCA based representation is the definition of the basis functions; in 2D DCT they are pre-defined, while in local PCA they are *learned*; as such, the basis functions are more representative of face blocks. Moreover, PCA based dimensionality reduction is optimal in a Mean Square Error (MSE) sense [44] (i.e. it preserves the most information), thus local PCA feature vectors could be of lower dimensionality than those from the 2D DCT based technique. However, there is no guarantee that the resulting feature vectors are optimal for discrimination purposes (this also applies to 2D DCT based techniques).

A possible disadvantage of the local PCA approach is that the basis functions may not have an easily interpretable meaning in terms of image structures (as opposed to a statistical meaning); moreover, the basis functions vary depending on which data set 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. To evaluate this hypothesis, we have calculated local PCA basis functions on the training section of the XM2VTS database; the first few are shown in Figure 6. It can be seen that the basis functions are quite similar to the 2D DCT basis functions (shown in Figure 4); thus it can be argued that, when using the XM2VTS database, the first three elements of a feature vector (resulting from local PCA feature extraction) would be the most affected by illumination changes.

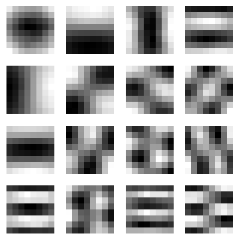

Figure 6: Graphical interpretation of the first few local PCA basis functions for N=8, calculated on the training section of the XM2VTS database and arranged in a zig-zag pattern. Lighter colors represent larger values.

5.4 Delta coefficients

It has been previously shown [39] that on a relatively small database, and using a GMM based classifier with a low number of gaussians, simply throwing out the first few coefficients from the a 2D DCT based feature vector increases robustness to illumination changes at the *expense* of reducing discrimination ability; this suggests that the first few coefficients are affected by illumination changes but contain a significant amount of discriminant information; to ameliorate the 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 [41].

The n-th *horizontal* delta coefficient for block located at (b, a) was defined as a modified polynomial coefficient:

$$\Delta^h c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k \; c_n^{(b,a+k)}}{\sum_{k=-K}^{K} h_k k^2} \tag{18}$$

Similarly, the n-th *vertical* delta coefficient was defined as:

$$\Delta^v c_n^{(b,a)} = \frac{\sum_{k=-K}^{K} k h_k \; c_n^{(b+k,a)}}{\sum_{k=-K}^{K} h_k k^2} \tag{19}$$

where $\vec{h}$ is a $2K+1$ dimensional symmetric window vector. Typically $K = 1$ and a rectangular window is used (thus $\vec{h} = [\; 1.0 \; 1.0 \; 1.0 \;]^T$). For 2D DCT based feature extraction, replacing the first three DCT coefficients (deemed to be the most affected by illumination changes) by their horizontal and vertical deltas corresponds to the *DCT-mod2* feature extraction (where the “mod” stands for ”modified”):

$$\vec{x} = \left[\; \Delta^h c_0 \; \Delta^v c_0 \; \Delta^h c_1 \; \Delta^v c_1 \; \Delta^h c_2 \; \Delta^v c_2 \;\; c_3 \; c_4 \; \ldots \; c_{M-1} \;\right]^T \tag{20}$$

where the (b, a) superscript was omitted for clarity. Extensions of the delta approach to also utilize diagonally neighboring blocks have been proposed in [37], although no considerable improvement was observed on a relatively small database.

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 neighbors; thus processing an image which has N_Y rows and N_X columns usually results in $(2\frac{N_Y}{N} - 3) \times (2\frac{N_X}{N} - 3)$ feature vectors. It must also be noted that the use of deltas effectively increases the spatial 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 increases from 8 pixels to 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).

6 Gaussian Mixture Model based classifier

Face verification can be treated as a two-class classification problem; the two classes correspond to the cases where the claimed identity is true and false, respectively. To solve this problem, we utilize a classifier based on Gaussian Mixture Models[2] (an instance of a Bayesian classifier [17]). For each client, two GMMs are utilized: the first to model the distribution of training feature vectors for that particular client, and the second to model the general distribution of training feature vectors for all training faces; the second GMM is commonly known as a Universal Background Model (UBM), a world model, or a generic model; it is used as an approximation of the *impostor* distribution.

To verify a given claim, a set of feature vectors, $X = \{\vec{x}_i\}_{i=1}^{N_V}$, is first extracted from a given face window; the likelihood of the claimant being the true claimant is then found:

$$\mathcal{L}\left(X|\lambda_C\right) = \prod_{i=1}^{N_V} p\left(\vec{x}_i|\lambda_C\right) \tag{21}$$

where

$$\begin{aligned} p\left(\vec{x}|\lambda\right) &= \sum_{g=1}^{N_G} w_j \, \mathcal{N}\left(\vec{x}, \vec{\mu}_g, \mathbf{\Sigma}_g\right) & (22) \\ \lambda &= \{w_g, \vec{\mu}_g, \mathbf{\Sigma}_g\}_{g=1}^{N_G} & (23) \end{aligned}$$

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

$$\mathcal{N}(\vec{x}; \vec{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{\Sigma}^{-1} \left(\vec{x} - \vec{\mu}\right)\right) \tag{24}$$

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 $\forall\, g : w_g \geq 0$). Given the likelihood of the claimant being an impostor, $\mathcal{L}(X|\lambda_{\overline{C}})$, an opinion on the claim is found using:

$$\Lambda(X) = \log \mathcal{L}\left(X|\lambda_C\right) - \log \mathcal{L}\left(X|\lambda_{\overline{C}}\right) \tag{25}$$

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

Given a set of training vectors, $X = \{\vec{x}_i\}_{i=1}^{N_V}$, the GMM parameters ($\lambda$) for each client model are found by adapting a Universal Background Model (UBM) using a form of maximum *a posteriori* (MAP) adaptation [21, 34]. The UBM is trained with the Expectation Maximization (EM) algorithm [17, 14] using training data from all clients. Using adaptation allows us to define client models with a limited amount of training data [20].

As mentioned before, the UBM is also used to find the likelihood of the claimant being an impostor, i.e.:

$$\mathcal{L}(X|\lambda_{\overline{C}}) = \mathcal{L}(X|\lambda_{\text{UBM}}) \tag{26}$$

There are various hyper-parameters to tune when using GMMs, such as the number of gaussians and the threshold. In our experiments, the hyper-parameters were selected to minimize the Equal Error Rate[3] (EER) on the validation set of the XM2VTS database (i.e. the data set which is *not* used for final performance evaluation).

[2]A GMM can be interpreted as a simplified version of a HMM; specifically, a GMM can be interpreted as a multi-state ergodic HMM (where all state transitions are equal and each state is represented by a gaussian) or as a single state HMM, with the state represented by multiple gaussians.

[3]The Equal Error Rate occurs when the False Acceptance Rate is equal to the False Rejection Rate.

The threshold is then used on the test set to obtain the final performance figure (i.e. in terms of HTER). In our experiments the number of gaussians was varied from 1 to 512, doubling the number of gaussians in each step (e.g. 1, 2, 4, 8, $\cdots$). Obviously the more gaussians are utilized, the more complex the resulting classifier is; time restrictions prevented us from running experiments with more gaussians[4].

As can be observed in Eqn. (21), each feature vector is treated independently, indicating that most of the spatial information from the face is lost. It has been shown that to some extent this information can be restored through embedding positional information into each vector [20]; we shall not utilize this extension here. Lastly, it must be noted that even though diagonal covariance matrices are utilized, correlated data can still be modeled as long as $N_G \geq 2$ [34].

7 Experiments and Discussions

In this section we evaluate the performance of the raw pixel, 2D DCT and local PCA 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. We also evaluate the effectiveness of 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 deemed to be most affected by illumination changes)
- Replacing lower order coefficients with their corresponding horizontal and vertical deltas
- Using only horizontal and vertical deltas

For the 2D DCT and local PCA feature extraction techniques, we first found the optimal dimensionality on the validation set 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 Figure 5 for an example). We also compared the performance of the 2D DCT and local PCA techniques against the "naive" raw pixel feature extraction technique, for which dimensionality reduction is not possible.

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

The 2D DCT and local PCA easily outperform the raw pixel feature extraction technique, at both the full dimensionality (64) and their optimal dimensionality (21). The performance advantage of 2D DCT and local PCA at the full dimensionality is most likely due to the decorrelation properties of these two techniques. Recall that the GMM classifier utilizes diagonal covariance matrices, and as such it is preferable to use decorrelated vectors.

The results in Table 1 further show that removing the mean from each raw pixel vector causes a dramatic improvement in the performance; this performance is fairly close to 2D DCT and local PCA at their best dimensionality. Considering that we are utilizing a classifier with diagonal covariance matrices, the difference in performance between raw pixels and mean removed raw pixels is consistent with the view that the classifier prefers decorrelated vectors.

[4]Even though the maximum number of gaussians was set to 512, results from experiments in Section 7 show that the optimum number of gaussians was quite often less than 512.

dimensionality	raw pixel			2D DCT			local PCA		
	best N_G	**EER**	**HTER**	**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	32	14.83	12.42	256	7.50	7.25	128	7.69	6.99
64 (mean removed)	64	5.86	5.79	-	-	-	-	-	-

Table 1: Performance for raw pixel, 2D DCT and local PCA based feature extraction techniques. "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.

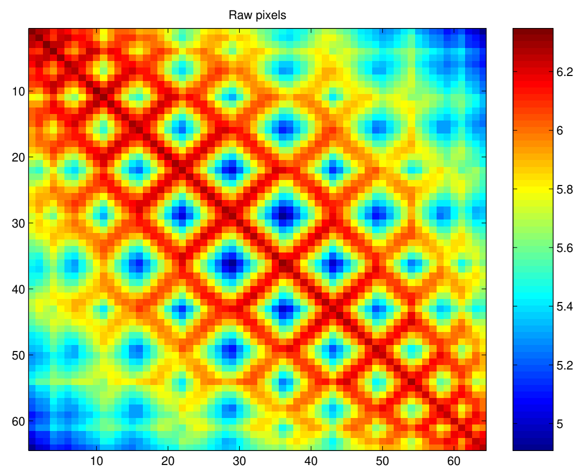

Figure 7: Squared covariance matrix, on a log scale, for raw pixel vectors.

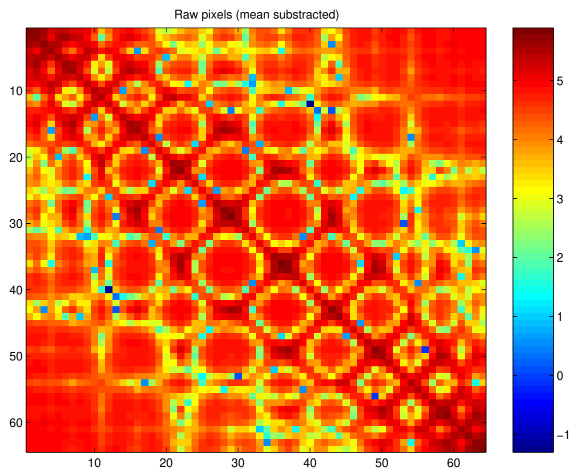

Figure 8: Squared covariance matrix, on a log scale, for mean removed raw pixel vectors.

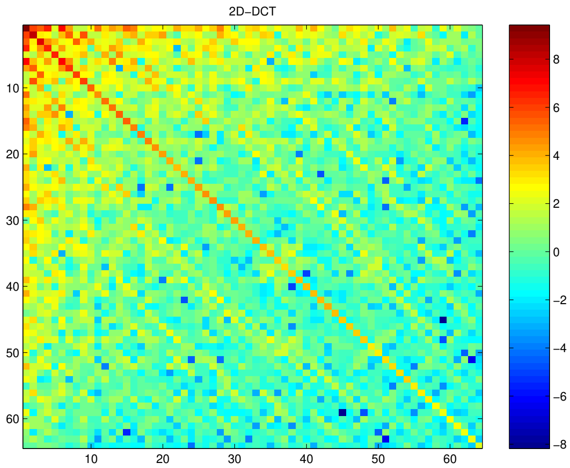

Figure 9: Squared covariance matrix, on a log scale, for 2D DCT vectors.

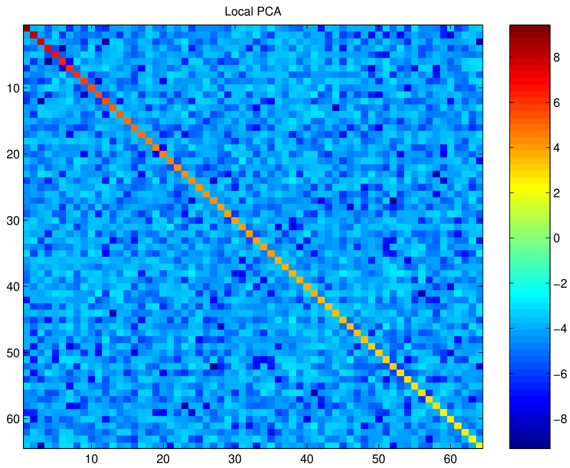

Figure 10: Squared covariance matrix, on a log scale, for local PCA vectors.

dim.	**raw pixel**					**mean removed raw pixel**				
	clean			**linear**	**non-lin.**	**clean**			**linear**	**non-lin.**
	best N_G	**EER**	**HTER**	**HTER**	**HTER**	**best** N_G	**EER**	**HTER**	**HTER**	**HTER**
64	32	14.83	12.42	45.58	42.90	64	5.86	5.79	9.04	17.87

Table 2: Performance for raw pixel, and mean removed raw pixel feature extraction techniques on clean faces and faces corrupted with the linear and non-linear illumination changes.

dim.	**2D DCT**					**local PCA**				
	clean			**linear**	**non-lin.**	**clean**			**linear**	**non-lin.**
	best N_G	**EER**	**HTER**	**HTER**	**HTER**	**best** N_G	**EER**	**HTER**	**HTER**	**HTER**
21	256	**4.83**	4.91	8.61	9.86	512	5.68	5.00	13.68	11.29
21 - 1	256	5.17	**4.37**	**4.76**	**6.29**	512	**5.50**	**4.09**	**6.52**	**8.53**
21 - 3	256	7.50	6.50	6.34	6.78	256	7.83	6.38	7.01	8.68
21 - 6	256	10.17	8.12	8.77	8.68	512	10.02	8.65	8.99	9.38
21 - 10	128	15.00	12.06	12.50	12.70	512	14.67	12.39	13.09	12.95

Table 3: Performance for 2D DCT and local PCA based feature extraction techniques on clean faces and faces corrupted with the linear and non-linear illumination changes. The baseline techniques were modified by removing elements from the *start* of the 21 dimensional baseline feature vectors.

Figures 7 to 10 represent the overall covariance matrices of feature vectors for each feature extraction technique; the feature vectors from the training section of the database were used to calculate the covariance matrices. As expected, the local PCA feature vectors are the most decorrelated, followed by 2D DCT vectors, mean removed raw pixel vectors and finally the raw pixel vectors. It must be noted that even though the local PCA feature vectors are more decorrelated compared to 2D DCT vectors, there is no observed advantage in terms of performance.

In the second experiment we evaluated the effects of linear and non-linear illumination changes on the performance of all feature extraction techniques. Table 2 shows the results for the raw pixel and mean removed raw pixel techniques, while Table 3 shows the results for the 2D DCT and local PCA techniques. For the latter two techniques we also evaluated the effects of removing coefficients which are deemed to be the most affected by illumination changes (i.e. we are removing lower order coefficients).

dim.	2D DCT					local PCA				
	clean			linear	non-lin.	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER	best N_G	EER	HTER	HTER	HTER
21	256	4.83	4.91	8.61	9.86	512	5.68	5.00	13.68	11.29
21 - 1 + 2	256	5.33	4.68	7.34	17.98	256	5.67	5.15	7.83	16.63
21 - 3 + 6	128	4.51	4.56	5.08	6.01	256	**4.83**	5.08	5.14	7.62
21 - 6 + 12	256	**4.50**	4.75	5.11	6.62	256	5.00	4.90	5.01	6.75
21 - 10 + 20	256	4.67	**4.17**	**4.49**	**5.93**	256	5.67	**4.81**	**4.87**	**5.96**

Table 4: Performance for 2D DCT and local PCA based feature extraction techniques on clean faces and faces corrupted with the linear and non-linear illumination changes. The baseline techniques were modified by *replacing* the elements from the *start* of the 21 dimensional baseline vectors with their corresponding horizontal and vertical deltas.

dim.	2D DCT deltas					local PCA deltas				
	clean			linear	non-lin.	clean			linear	non-lin.
	best N_G	EER	HTER	HTER	HTER	best N_G	EER	HTER	HTER	HTER
2 (1+1)	32	14.02	12.72	27.11	46.20	32	13.61	12.40	27.37	46.28
6 (3+3)	128	5.33	5.90	9.43	30.44	256	4.87	5.57	8.57	32.00
12 (6+6)	512	**3.83**	4.23	5.66	14.43	512	**3.32**	4.24	5.58	13.84
20 (10+10)	512	4.16	**4.01**	**4.78**	**7.18**	512	4.04	**3.92**	**4.32**	**7.92**

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

As can be seen, the raw pixel technique quickly falls apart; mean removal significantly helps, as removing the mean can be interpreted as a form of illumination normalization (this is somewhat akin the throwing out the 0-th coefficient from a 2D DCT feature vector); however, even with mean removal the performance still degrades considerably. The 2D DCT and local PCA techniques are more robust, with the local PCA technique being somewhat more affected by illumination changes than the 2D DCT method. For both 2D DCT and local PCA, 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 no further gains in robustness.

In the third experiment we evaluated the effects of replacing coefficients (as opposed to throwing them out, as it was done in the second experiment) with their corresponding horizontal and vertical deltas. By comparing Tables 3 and 4 it can be observed that the use of deltas tends to ameliorate the performance loss which occurs when coefficients are thrown out, while in most cases keeping the robustness to illumination changes. Compared to just throwing out the first coefficient from the 21 dimensional baseline vectors (resulting in 20 dimensional vectors), the use of deltas results in a small performance and robustness increase at the larger dimensionality of 31 (where $21 - 10 + 20 = 31$).

In the fourth experiment we appraised the performance and robustness of feature vectors which contain only horizontal and vertical deltas. As can be seen in Table 5, horizontal and vertical deltas of the first element from 2D DCT and local PCA vectors are considerably affected by illumination changes. The more deltas are utilized, the higher the performance and robustness is, suggesting that only deltas of higher order elements from the baseline vectors are useful. It is interesting to see that the performance of feature vectors comprised of 20 deltas (i.e. 10 horizontal and 10 vertical deltas), is very similar to the performance of baseline vectors with the first coefficient thrown out (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 for 16×16 2D DCT based feature extraction on clean faces.

dim.	16×16 **2D DCT**				
	clean			**linear**	**non-lin.**
	best N_G	**EER**	**HTER**	**HTER**	**HTER**
21	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 for 16×16 2D DCT based feature extraction techniques 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.

As mentioned in Section 5.4, one of the effects of using deltas is an increase in the effective spatial area used when obtaining each feature vector. Instead of using the indirect method of deltas to increase the spatial 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 neighboring 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 linear and non-linear illumination changes on the performance of the 16×16 2D DCT based feature extraction technique; we also evaluated the effects of removing coefficients which are deemed to be the most affected by illumination changes. As can be observed in Table 7, removing the first three coefficients considerably enhances robustness to illumination changes, at the cost of a small performance degradation on clean images. Removing more coefficients causes a noticeable reduction in performance on clean images with no further gains in robustness. Lastly, by comparing Tables 3 and 7 it can be observed that the performance and robustness of 16×16 2D DCT vectors with the first three coefficients removed (resulting in 18 dimensional vectors) is similar to the performance of 8×8 2D DCT vectors with the first coefficient removed (i.e. 20 dimensional vectors).

8 Conclusions

In this report we have compared four local feature extraction techniques for the task of face verification. As opposed to holistic feature extraction techniques, local features describe only a small part of the face. The four evaluated techniques are based on (ordered in terms of complexity): raw pixels, raw pixels with mean removal, 2D Discrete Cosine Transform (DCT) and *local* Principal Component Analysis (PCA). The comparison was performed in terms of discrimination ability and robustness to illumination changes. We have also evaluated the effectiveness of several feature extraction *modification* approaches 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 and using only deltas.

Results on the XM2VTS database suggest that when using a Gaussian Mixture Model (GMM) based classifier, the raw pixel technique provides poor discrimination and is easily affected by illumination changes;

the mean removed raw pixel technique provides performance that is fairly close to 2D DCT and local PCA, but is considerably affected by illumination changes. The performance of 2D DCT and local PCA techniques is quite similar, suggesting that the 2D DCT technique is to be preferred over the local PCA technique, due to the lower complexity of the 2D DCT (i.e. the basis functions in 2D DCT are fixed while in local PCA they first have to be learned). Both 2D DCT and local PCA techniques are considerably more robust to illumination changes compared to the raw pixel techniques.

Modifying the 2D DCT and local PCA techniques by removing the first coefficient, which is deemed to be the most affected by illumination changes, clearly enhances robustness; when utilizing analysis blocks of size 8×8, removing more than the first coefficient causes a noticeable reduction in performance on clean images and provides no further gains in robustness. Further modification through the use of deltas tends to ameliorate the performance loss which occurs when coefficients are thrown out, while in most cases keeping the robustness to illumination changes. Compared to just throwing out the first coefficient, the use of deltas can achieve a small performance and robustness increase. The results also show that systems using only the deltas of the first element from 2D DCT and local PCA vectors are considerably affected by illumination changes; however, the more deltas are utilized, the higher the performance and robustness is, suggesting that only deltas of higher order coefficients are useful.

For 2D DCT based feature extraction, increasing the analysis block size from 8×8 to 16×16 results in slightly better performance on clean faces; however, the first three coefficients need to be removed, rather than just the first one, in order to achieve similar robustness.

Results in [6] suggest that feature vectors derived from a larger spatial area are more affected by out-of-plane rotations of the face; combined with the results presented in this report, we thus conclude that out of the feature extraction techniques presented here, it is more appropriate to use analysis blocks of size 8×8 (as opposed to 16×16) with 2D DCT decomposition; out of the 64 resulting coefficients, the second through to 21-st (resulting in 20 dimensional feature vectors) are the most robust to illumination changes while providing good discriminatory information.

9 Acknowledgments

The authors thanks S. Bengio and J. Mariethoz for their useful suggestions. The authors also thank 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). The implementation of the experiments was aided by the Newmat C++ matrix library [13] and the Torch machine learning library [12].

References

[1] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J.P. Thiran. The BANCA database and evaluation protocol. In *4th Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pages 625–638, Guildford, UK, 2003.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[3] S. Bengio, M. Keller, and J. Mariethoz. The Expected Performance Curve. IDIAP Research Institute, 2003. Research Report RR 03-85.

[4] S. Bengio, J. Mariethoz, and S. Marcel. Evaluation of Biometric Technology on XM2VTS. IDIAP Research Institute, 2001. Research Report RR 01-21.

[5] R. M. Bolle, J. H. Connell, and N. K. Ratha. Biometrics perils and patches. *Pattern Recognition*, 35(12):2727–2738, 2002.

[6] C. Sanderson and S. Bengio. Augmenting Frontal Face Models for Non-Frontal Verification. In *Proc. Workshop on Multi-Modal User Authentication*, pages 165–172, Santa-Barbara, 2003.

[7] C. Sanderson and S. Bengio. Statistical Transformations of Frontal Models for Non-Frontal Face Verification. In *Proc. IEEE Int. Conf. Image Processing*, Singapore, 2004.

[8] F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS. In *4th Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pages 911–920, Guildford, UK, 2003.

[9] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and Machine Recognition of Faces: a Survey. *Proc. IEEE*, 83(5):705–740, 1995.

[10] L-F. Chen, H-Y. Liao, J-C. Lin, and 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):1393–1403, 2001.

[11] L. Chengjun. A Bayesian discriminating features method for face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.

[12] R. Collobert, S. Bengio, and J. Mariéthoz. Torch: a modular machine learning software library. IDIAP Research Institute, 2002. Research Report RR 02-46.

[13] R. B. Davies. Newmat, a matrix library in C++. available at http://www.robertnz.net/.

[14] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal Royal Statistical Soc. Ser. B*, 39(1):1–38, 1977.

[15] G.R. Doddington, M.A. Przybocki, A.F. Martin, and D.A. Reynolds. The NIST speaker recognition evaluation - Overview, methodology, systems, results, perspective. *Speech Communication*, 31(2-3):225–254, 2000.

[16] B. Duc, S. Fischer, and J. Bigun. Face Authentication with Gabor Information on Deformable Graphs. *IEEE Trans. Image Processing*, 8(4):504–516, 1999.

[17] R.O. Duda, P.E. Hart, and D.G Stork. *Pattern Classification*. Wiley, 2001.

[18] J.-L. Dugelay, J.-C. Junqua, C. Kotropoulos, F. Perronnin, and I. Pitas. Recent Advances in Biometric Person Authentication. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, volume IV, pages 4060–4062, Orlando, 2002.

[19] S. Eickeler, S. Muller, and G. Rigoll. Recognition of JPEG Compressed Face Images Based on Statistical Methods. *Image and Vision Computing*, 18(4):279–287, 2000.

[20] F. Cardinaux and C. Sanderson and S. Bengio. Face Verification Using Adapted Generative Models. In *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pages 825–830, Seoul, 2004.

[21] J.-L. Gauvain and C.-H. Lee. Maximum *a Posteriori* Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. *Proc. IEEE Trans. Speech and Audio Processing*, 2(2):291–298, 1994.

[22] R. C. Gonzales and R.E. Woods. *Digital Image Processing*. Addison-Wesley, 1993.

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

[24] L. Guevin. Is the iris the gateway to our true identities? available at http://www.biometritech.com/features/laura3.htm.

[25] S.G. Kong, J.Heo, B.R. Abidi, J. Paik, and M.A. Abidi. Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding*, in press.

[26] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Wurtz, and W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Trans. Computers*, 42(3):300–311, 1993.

[27] T.S. Lee. Image Representation Using 2D Gabor Wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996.

[28] J. Luettin and G. Maitre. Evaluation Protocol for the XM2FDB Database (Lausanne Protocol). IDIAP Research Institute, 1998. Communication COM 98-05.

[29] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET curve in assessment of detection task performance. In *Proceedings of Eurospeech'97*, pages 1895–1898, Rhodes, Greece, 1997.

[30] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The Extended M2VTS Database. In *2nd Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pages 72–77, Washington, D.C., 1999.

[31] A. Nefian and M. Hayes. Face Recognition Using an Embedded HMM. In *Proc. Audio and Video-based Biometric Person Authentication*, pages 19–24, Washington D.C., 1999.

[32] J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. Gonzales-Rodriguez. Authentication gets personal with biometrics. *IEEE Signal Processing Magazine*, 21(2):50–62, 2004.

[33] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[34] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn. Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing*, 10(1–3), 2000.

[35] F.S. Samaria. *Face Recognition Using Hidden Markov Models*. PhD Thesis, Cambridge University, 1994.

[36] C. Sanderson. Face Processing & Frontal Face Verification. IDIAP Research Institute, 2003. Research Report RR 03-20.

[37] C. Sanderson and S. Bengio. Robust Features for Frontal Face Authentication in Difficult Image Conditions. In *4th Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pages 495–504, Guildford, UK, 2003.

[38] C. Sanderson and K. K. Paliwal. On the Use of Speech and Face Information for Identity Verification. IDIAP Research Institute, 2004. Research Report RR 04-10.

[39] C. Sanderson and K.K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.

[40] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression Database. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[41] F.K. Soong and A.E. Rosenberg. On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition. *IEEE Trans. Acoustics, Speech and Signal Processing*, 36(6):871–879, 1988.

[42] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991.

[43] H. L. Van Trees. *Detection, Estimation and Modulation Theory, vol. 1.* Wiley, New York, 1968.

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

[45] J.L. Wayman. Digital signal processing in biometric identification: a review. In *Proc. IEEE Int. Conf. Image Processing*, volume 1, pages 37–40, Rochester, 2002.

[46] K.W. Wong, K.M. Lam, and W.C. Siu. An efficient algorithm for human face detection and facial feature extraction under different conditions. *Pattern Recognition*, 34(10):1993–2004, 2001.

[47] M-H. Yang, D.J. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[48] J. Zhang, Y. Yan, and M. Lades. Face Recognition: Eigenface, Elastic Matching, and Neural Nets. *Proc. IEEE*, 85(9):1423–1435, 1997.