Combining experts for improved face verification performance*

Vitomir Štruc, France Mihelič, Nikola Pavešić

Fakulteta za elektrotehniko, Univerza v Ljubljani
Tržaška 25, 1000 Ljubljana, Slovenija
vitomir.struc@fe.uni-lj.si, france.mihelic@fe.uni-lj.si, nikola.pavesic@fe.uni-lj.si

Večmodalni pristop k razpoznavanju obrazov

Samodejno razpoznavanje (avtentikacija/identifikacija) obrazov predstavlja eno najaktivnejših raziskovalnih področij biometrije. Avtentikacija oz. identifikacija oseb z razpoznavanjem obrazov ponuja možen način povečanja varnosti pri različnih dejavnostih, (npr. pri elektronskem poslovanju na medmrežju, pri bančnih storitvah ali pri vstopu v določene prostore, stavbe in države). Ponuja univerzalen in nevsiljiv način razpoznavanja oseb, ki pa trenutno še ni dovolj zanesljiv. Kot možna rešitev problema zanesljivosti razpoznavanja se v literaturi vse pogosteje pojavljajo večmodalni pristopi, v katerih se razpoznavanje izvede na podlagi večjega števila postopkov razpoznavanja obrazov. V skladu z opisanim trendom, bomo v članku ovrednotili zanesljivost delovanja različnih postopkov razpoznavanja obrazov, ki jih bomo na koncu združili še v večmodalni pristop. S pomočjo eksperimentov na podatkovni zbirki XM2VTS bomo preverili zanesljivost delovanja večmodalnega pristopa in jo primerjali z zanesljivostjo uveljavljenih postopkov razpoznavanja.

1 Introduction

Amongst the numerous biometrics (e.g., fingerprints, palmprints, iris scans, voice recordings, etc.) that can be used for personal recognition the human face has a privileged role for several reasons: *(i)* people use faces as the primary means of identification in daily interactions and have, therefore, no objections using their faces as means of identification in other scenarios such as access or border control, *(ii)* unlike iris or fingerprint recognition, face recognition requires no cooperation from the user and is considered one of the most unintrusive recognition techniques, and *(iii)* face recognition requires only a low- (medium-) resolution video-camera and some data storage- and processing-unit which results in low-cost recognition systems.

Due to the listed reasons a considerable research effort has been directed towards face recognition. Various methods and algorithms have been proposed in the literature; however, challenging problems related to recognition from images captured under varying illumination conditions, under partial occlusion of the face, with different pose or facial expression still remain. A major issue of face recognition is also how to improve the overall performance of the employed recognition techniques. The current trend in solving the above-mentioned problem is to combine different recognition experts into a multi-modal, i.e., intra-modal, face recognition approach and hopefully improve the final recognition performance.

In this paper we will assess the performance of several face recognition techniques and combine them into an intra-modal face recognition approach. The feasibility of the proposed approach will be demonstrated in a series of face verification experiments performed on the XM2VTS database.

The rest of the paper is organized as follows: in Sections 2 to 5 the tested face recognition techniques are briefly reviewed. The verification experiments are presented in Section 7 and the paper concludes with some final comments in Section 8.

2 Subspace projection techniques

Subspace projection techniques are amongst the most widely used feature-extraction techniques in the field of face recognition. They range from linear techniques, such as principal component analysis (PCA) and linear discriminant analysis (LDA), to nonlinear (or kernel) techniques like kernel principal component analysis (KPCA) and kernel Fisher analysis (KFA). Some of these techniques will be presented in the remainder of this section.

2.1 Linear subspace projection techniques

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ represent a matrix containing in its columns n d-dimensional training images (in vector form). Linear subspace projection techniques aim at constructing a $d \times d'$ transformation matrix $\mathbf{W}$ which can be used to project an arbitrary face image $\mathbf{x}$ into a lower-dimensional subspace, i.e., $\mathbf{y} = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu})$, where $\boldsymbol{\mu}$ denotes the mean vector of the training images and $\mathbf{y}$ represents the d'-dimensional feature vector.

Principal component analysis - PCA

PCA, first introduced to face recognition by Turk and Pentland in [1], identifies a subspace whose basis vectors correspond to the maximum variance directions present in the training data. Each training image can be projected into this subspace and again reconstructed with minimum error. From the mathematical point of view, the PCA transformation matrix $\mathbf{W}$ corresponds to the leading eigenvectors of the covariance matrix of the training data.

* Partially supported by the national research program P2-0250(C) Metrology and Biometric Systems, the bilateral project with the People's Republic of China Bi-CN/07-09-019, the bilateral project with the Bulgarian Academy of Sciences - Face and Signature Biometrics, the national project AvID M2-0210 and the EU-FP7 project 217762 Homeland security, biometric Identification and personal Detection Ethics (HIDE)

Linear discriminant analysis – LDA

LDA or the Fisherface feature-extraction approach, proposed in [2], derives the transformation matrix $\mathbf{W}$ by maximizing Fisher's discriminant criterion in form of the ratio of the between-class to the within-class scatter matrix. Thus, LDA seeks a subspace in which the discriminant information contained in the training data is emphasized.

2.2 Kernel subspace projection techniques

Let us again consider the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ containing in its columns vector forms of the training images. Furthermore, let Φ represent a nonlinear mapping of the d-dimensional input variable $\mathbf{x}_i$ from the original input space $\mathbb{R}^d$ to a high-dimensional feature space $\mathcal{F}$, i.e., $\Phi: \mathbf{x}_i \in \mathbb{R}^d \rightarrow \Phi(\mathbf{x}_i) \in \mathcal{F}$. Kernel methods try to identify a linear subspace in the high-dimensional feature space rather than the original input space since (according to Cover's theorem) data which is nonlinearly separable in the input space is with high probability linearly separable if the input space is nonlinearly transformed to a high-dimensional feature space [3]. Kernel methods commonly avoid direct computation of the nonlinear mapping Φ, but rather use the so-called "kernel trick" and derive the kernel transformation matrices based on the kernel matrices of the training data. Similar to the linear case, the most commonly used kernel methods are kernel principal component analysis (KPCA) [4] and kernel Fisher analysis (KFA) [5].

3 Gabor wavelet based techniques

Unlike the subspace projection techniques which are typical representatives of the appearance-based face recognition techniques, Gabor wavelet based methods represent another class of recognition approaches – the so called feature-based approaches. These methods extract features at specific facial landmarks (also called fiducial points) and are, therefore, more robust in terms of illumination, pose and facial expression than the appearance-based methods.

The most successful techniques based on Gabor wavelets use a two-stage approach for feature extraction: *(i)* in the first step each face image is convolved with a set of forty Gabor wavelets and the results (in an appropriate form) are concatenated into a high-dimensional vector, *(ii)* in the second step the dimensionality of the high-dimensional vector is reduced with the help of a subspace projection technique. Popular methods which employ the presented two-stage approach are GaborPCA [6], GaborLDA[7], GaborKPCA[3] and GaborKFA[5].

4 Correlation filters

Correlation filters have only recently been applied to face recognition. They exhibit some desirable properties such as shift-invariance or occlusion-insensitiveness while simultaneously achieving "good" recognition performance [8].

During the training stage one correlation filter is constructed for each subject and stored in the database. When a "new" image needs to be recognized it is filtered with the correlation filter corresponding to the claimed identity and classified according to the result. A block diagram of the described approach is presented in Fig. 1.

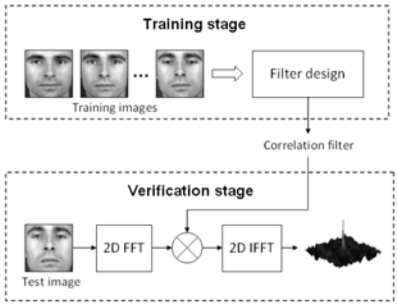

Figure 1. Using correlation filters for face recognition

In the remainder we will present two popular correlation filters, namely, the minimum average correlation energy (MACE) filter and the synthetic discriminant function (SDF) filter.

4.1 The MACE filter

Given a set of N training images $\mathcal{X} = \{\mathbf{x}_i, i=1,2,\dots,N\}$ of a given subject, the MACE correlation filter $\mathbf{h}$ is constructed as $\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}$, where $\mathbf{X}$ denotes a matrix containing in its columns vector forms of the 2D Fourier transforms of the training images, $\mathbf{D}$ represents a diagonal matrix containing the average power spectrum of the training images along its diagonal, $\mathbf{u}$ denotes a N-dimensional vector whose values are commonly set to one and $^{+}$ stands for the complex conjugate transpose operator. Like with all correlation filters the equation for the MACE filter is obtained by solving a constrained optimization problem – a detailed description of the filter can be found in [8].

4.2 The SDF filter

Using the same notation as in Section 4.1 the SDF filter is constructed as $\mathbf{h} = \mathbf{X}(\mathbf{X}^{+}\mathbf{X})^{-1}\mathbf{u}$. The filter equation is derived under the assumption that the filter itself is a linear combination of the training images, while the optimization criterion is to achieve a pre-defined value at the origin of the correlation plane (determined by the vector $\mathbf{u}$) [8].

5 Four-directional features

The last type of face recognition techniques considered in this paper are the four-directional-feature-based (FDF-based) methods introduced in [9]. Similar to Gabor wavelet based feature-extraction techniques, these methods use a filter bank which is, however, comprised of only four filters. The bank contains four directional filters, e.g., Sobel filters, which are used to produce edge images of four directions, i.e., horizontal, vertical and both diagonals. The results are then down-sampled, smoothed with a Gaussian kernel and finally concatenated into a vector. Combining the FDF vector with different subspace projection techniques results in the face recognition techniques tested in Section 6.5, namely, FDF+PCA, FDF+LDA, FDF+KPCA and FDF+KFA.

6 Experiments

6.1 Database and experimental protocol

The performance of the presented face recognition techniques was assessed on the publicly available XM2VTS database [4] which is the benchmark database for assessing face verification technology. The database contains 2360 colour face images that correspond to 295 distinct subjects. Thus, each subject in the database is represented with a total of 8 facial images. As the images were captured in five recording sessions which were distributed over a period of five months, different images of the same subject exhibit variations in terms of pose, facial expression, hair-style, absence or presence of glasses and so forth.

Similar to other studies on face verification (e.g., [3],[5],[6]) all images from the database were subjected to a pre-processing procedure which included: *(i)* a conversion of the colour face images to 8-bit gray scale images, *(ii)* a geometric normalization procedure which, based on the *ground truth* (eye coordinates) provided with the database, rotated the images in such a way that the eyes were located at pre-defined positions, scaled the faces to a standard inter-ocular distance and finally cropped the face regions to a standard size of 128×128 pixels, *(iii)* a photometric normalization procedure which featured histogram equalization followed by a conversion of the pixel intensity distribution to N(0,1).

The experiments presented in the remainder of the paper were performed in accordance with the first configuration of the well-established experimental protocol associated with the XM2VTS database, i.e., the Lausanne protocol. The protocol defines which images should be used for training (i.e., constructing client models), evaluation (i.e., setting the decision threshold that ensures a certain operational point on the receiver operating characteristic curve – ROC curve) and testing (i.e., determining the false acceptance (FA) and false rejection (FR) rates of the system). A detailed description of the protocol can be found in [10].

6.2 Applying subspace projection technique to grey-scale images

Our first series of verification experiments assessed the performance of the linear (PCA and LDA) and nonlinear (KPCA and KFA) subspace projection techniques when applied to the pre-processed grey-scale images from the XM2VTS database. The parameters of the employed techniques, such as the number of features or the kernel function used, were chosen in such a way that the verification errors, when using the nearest neighbour classifier and the cosine similarity measure, would be as low as possible. The ROC curves (showing the dependencies of the false acceptance and the false rejection rates at various decision thresholds) generated during the experiments are presented in Fig. 2.

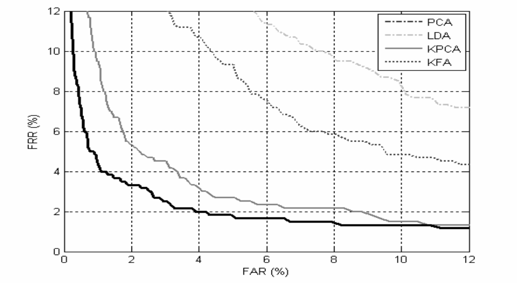

Figure 2. The ROC curves of the experiments

We can see that the nonlinear (kernel) version of Fisher's discriminant analysis, i.e., KFA, performed the best, followed in order by the LDA, KPCA and PCA feature-extraction techniques. As expected, both kernel techniques performed better than their linear counterparts.

6.3 Applying subspace projection technique to Gabor-filtered images

Our second series of experiments assessed the performance of the subspace projection techniques in connection with Gabor-filtered face images. Similar to the experiments presented in Section 6.2, the following techniques were implemented for the comparison: Gabor+PCA [6], Gabor+LDA[7], Gabor+KPCA[3] and Gabor+KFA [5], which are denoted as GPCA, GLDA, GKPCA and GKFA in Fig. 3 respectively. For all the listed methods the cosine similarity measure was used for matching score calculation.

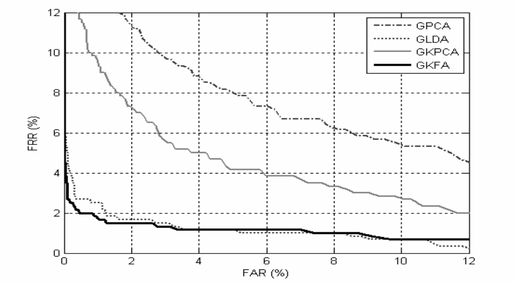

Figure 3. The ROC curves of the experiments

From the presented graphs we can find that the performance of all techniques improved significantly when Gabor filtered images were used instead of the original grey-scale ones, while the relative ranking of the individual methods remained unchanged.

6.4 Assessing the performance of correlation filters

The goal of the third series of experiments was to assess the performance of two popular correlation filters, i.e., the MACE and SDF correlation filter, in the verification scenario. As we can see from Fig. 4 where the ROC curves of the experiments are presented, the MACE

	Grey-scale images				*Gabor-filtered images*				*Four-directional features*				*Corr. filters*		*MM*
	PCA	LDA	KPCA	KFA	PCA	LDA	KPCA	KFA	PCA	LDA	KPCA	KFA	MACE	SDF	
FAR	9.4	3.9	5.8	2.8	7.3	1.9	4.9	1.3	7.4	2.6	2.9	2.0	8.4	12.4	0.3
FRR	8.8	3.0	5.3	2.8	7.3	1.5	5.0	0.8	8.0	2.5	2.5	2.5	9.0	10.3	0.6
TER	18.2	6.9	11.1	5.6	14.6	3.4	9.9	2.1	15.4	5.1	5.4	4.5	17.4	22.7	**0.9**

Table 2.Error rates in (%) obtained on the test set with the decision threshold that ensured equal error rates on the evaluation set

filter performed better than the SDF filter; however, the error rates at the equal error operating point (where the FAR and FRR are equal) for both filters are somewhere in the range of the PCA method, which performed worst among the feature-extraction techniques tested in the previous two sections. While the results are not the best in terms of error rates, the correlation filters still exhibit some desirable characteristics, such as shift-invariance or occlusion-insensitiveness. Thus, they are suitable for inclusion into a multi-modal face verification approach.

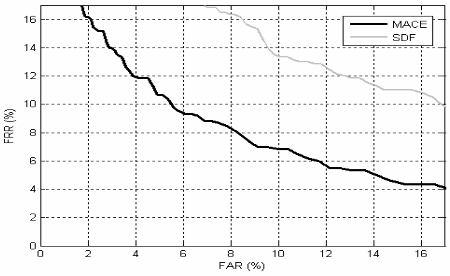

Figure 4. The ROC curves of the experiments

6.5 Assessing the performance of four-directional features (FDFs)

In our fourth series of verification experiments we aimed at determining the performance of the four-directional features in conjunction with the four subspace projection techniques already used in Sections 6.2 and 6.3., i.e., PCA, LDA, KPCA and KFA. The results presented in Fig. 5 again show that regardless of the data (either original grey-scale or somehow pre-processed face images) that the four tested subspace projection techniques are applied to, their relative ranking remains the same, i.e., KFA performs the best, followed by the LDA, KPCA and PCA techniques.

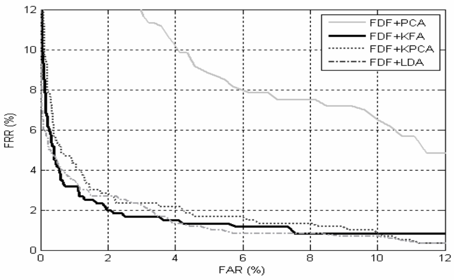

Figure 5. The ROC curves of the experiments

6.6 Testing the multi-modal verification approach

In our last series of experiments we combined all the implemented techniques at the matching score level and tested the resulting multi-modal approach for its face verification performance. Unlike in the experiments presented in Sections 6.2 to 6.5, this series of tests featured only images from the test set. The error rates FAR, FRR and TER, where TER=FAR+FRR, of the multi-modal (MM) and all other methods for the decision threshold that ensured equal error rates on the evaluation image set are presented in Table. 1. We can see that the multi-modal approach achieved lower error rates than any of the remaining methods on its own.

7 Conclusion

In this paper we have presented an empirical assessment of the verification performance of several popular face recognition (or feature-extraction) techniques as well as a multi-modal verification approach which was shown to outperform all individual methods. The presented results clearly show that the combination of verification experts, i.e., intra-modality, offers a simple and reliable way to improve the performance of automated biometric verification systems.

References

[1] M. Turk, A. Pentland, Eigenfaces for Recognition. *Journal of Cognitive Neurosicence* 3, 1991, 71-86.

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *Proceedings of ECCV*, 1996, pp. 45-58.

[3] C. Liu, Gabor-based kernel PCA with fractional power polynomial models for face recognition, *IEEE TPAMI* 26(5), 2004, 572-581.

[4] B. Schölkopf, A.J. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation* 10, 1998,1299-1319.

[5] C. Liu, Capitalize on dimensionality increasing techniques for improving FRGC performance, *IEEE TPAMI* 28(5), 2006, 725-737.

[6] K.C. Chung, S.C. Kee, S.R. Kim, Face recognition using PCA of Gabor filter responses, *Proceedings of the workshop RATFG*, 1999, pp. 53-57.

[7] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced FLD model for face recognition, *Transactions on Image Processing* 11(4), 2002, 467-476.

[8] N. Muller, B. Herbst, On the use of SDF-type filters for distortion invariant image location, *Proceedings of ICPR*, 2000, pp. 526-529.

[9] Y. Ishii, H. Hongo, K. Yamamoto, Y. Niwa, Real-time head and face detection using four directional features, *Proceedings of AFGR*, 2004, pp. 403-408.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, *XM2VTSDB: The extended M2VTS database, Proceedings of AVBPA*, 1999, pp. 72-77.