A Symmetric Transformation for LDA-based Face Verification

Sébastien Marcel
Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP)
rue du Simplon 4, 1920 Martigny, Switzerland
marcel@idiap.ch

Abstract

One of the major problem in face verification is to deal with a few number of images per person to train the system. A solution to that problem is to generate virtual samples from an unique image by doing simple geometric transformations such as translation, scale, rotation and vertical mirroring. In this paper, we propose to use a symmetric transformation to generate a new virtual sample. This symmetric virtual sample is obtained by computing the average between the original image and the vertical mirrored image. The face verification system is based on LDA feature extraction, successfully used in previous studies, and MLP for classification. Experiments were carried out on a difficult multi-modal database, namely BANCA. Results on this database show that our face verification system performs better that the state-of-the-art and also that the addition of the symmetric virtual sample improves the performance.

1. Introduction

Identity verification is a general task that has many real-life applications such as access control, transaction authentication (in telephone banking or remote credit card purchases for instance), voice mail, or secure teleworking.

The goal of an *automatic identity verification system* is to either accept or reject the identity claim made by a given person. Biometric identity verification systems are based on the characteristics of a person, such as its face, fingerprint or signature. A good introduction to identity verification can be found in [14]. Identity verification using face information is a challenging research area that was very active recently, mainly because of its natural and non-intrusive interaction with the authentication system.

The paper is structured as follows. In section 2 we introduce the reader to the problem of identity verification. Then, in section 3 we present the proposed symmetric transformation within the framework of a state-of-the-art face verification system based on a linear discriminant feature extraction technique, successfully applied to face verification [8], and on a Multi-Layer Perceptron classifier. In section 4, we provide experimental results on the multi-modal benchmark database BANCA using its associated protocol. Finally, we analyze the results and conclude.

2. Face verification

An identity verification system has to deal with two kinds of events: either the person claiming a given identity is the one who he claims to be (in which case, he is called a *client*), or he is not (in which case, he is called an *impostor*). Moreover, the system may generally take two decisions: either *accept* the *client* or *reject* him and decide he is an *impostor*.

The classical face verification process can be decomposed into several steps, namely *image acquisition* (grab the images, from a camera or a VCR, in color or gray levels), *image processing* (apply filtering algorithms in order to enhance important features and to reduce the noise), *face detection* (detect and localize an eventual face in a given image) and finally *face verification* itself, which consists in verifying if the given face corresponds to the claimed identity of the client.

One of the major problem in face verification is to deal with a few number of images per person to train the system. A solution to that problem is to generate virtual samples from an unique image by doing simple geometric transformations [12] such as translation, scale, rotation and vertical mirroring. In this paper, we propose to use a symmetric transformation to generate a new virtual sample. It is obtained by computing the average between the original image and the vertical mirrored image. This symmetric transformation has also the effect to normalize the face by smoothing local deformations due to small out-of-plane rotations.

In this paper, we assume (as it is often done in comparable studies, but nonetheless incorrectly) that the detection step has been performed perfectly and we thus concentrate on the last step, namely the face verification step.

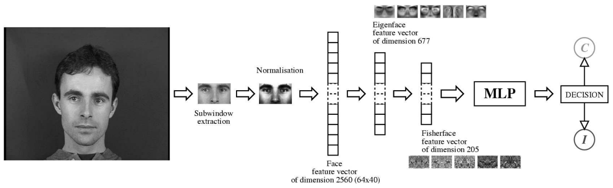

Figure 1. Face Verification using LDA and MLP.

3. The proposed approach

In face verification, we are interested in particular objects, namely faces. The representation used to code input images in most state-of-the-art methods are often based on gray-scale face image [10, 9] or its projection into principal component subspace or linear discriminant subspace [8].

Principal Component Analysis (PCA) identifies the subspace defined by the eigenvectors of the covariance matrix of the training data.

The projection of face images into the coordinate system of eigenvectors (Eigenfaces) associated with nonzero eigenvalues achieves information compression, decorrelation and dimensionality reduction to facilitate decision making. The linear discriminant analysis (LDA) subspace holds more discriminant features for classification [1] than the PCA subspace. A linear discriminant is a simple linear projection $\hat{y} = b + \mathbf{w} \cdot \mathbf{x}$ of the input vector onto an output dimension:

$$\hat{y} = b + \mathbf{w} \cdot \mathbf{x} \ . \tag{1}$$

where the estimated output $\hat{y}$ is a function of the input vector $\mathbf{x}$, and the parameters $\{b, \mathbf{w}\}$ are chosen according to a given criterion. Depending on the criterion (Fisher criterion [7] for instance) chosen to select the optimal parameters, one could obtain a different solution.

The Fisher criterion aims at maximizing the ratio of between-class scatter to within-class scatter. Given a set of l_i points belonging to class $\mathcal{C}_i$, we can define the mean of each class $i = 1...c$, where c is the number of classes, as

$$\mu_i = \frac{1}{l_i} \sum_{k \in \mathcal{C}_i} \mathbf{x}_k \ . \tag{2}$$

The within-class scatter matrix is then defined as

$$\mathbf{S}_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in \mathcal{C}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^t \ . \tag{3}$$

where N is the total number of image sample $N = \sum_{i=1}^{c} l_i$. The between-class scatter matrix is defined as

$$\mathbf{S}_b = \frac{1}{c} \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^t \ . \tag{4}$$

where μ is the grand mean, i.e the mean of the means μ_i. Fisher's criterion can then be defined as maximizing

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_b \mathbf{w}}{\mathbf{w}^t \mathbf{S}_w \mathbf{w}} \ . \tag{5}$$

and a solution can be found by computing the eigenvectors of

$$\mathbf{w} = \mathbf{S}_w^{-1} \mathbf{S}_b \ . \tag{6}$$

In the following sections, we describe our face verification system: an MLP classifier trained on a gray-scale face image projected into LDA subspace (Fig. 1) as described in [8].

3.1. Feature extraction

3.1.1. Face modeling In a real application, the face bounding box will be provided by an accurate face detector [13, 6] but here the bounding box is computed using manually located eyes coordinates, assuming a perfect face detection. In this paper, the face bounding box is determined using face/head anthropometry measures [5] according to a face model (Fig. 2).

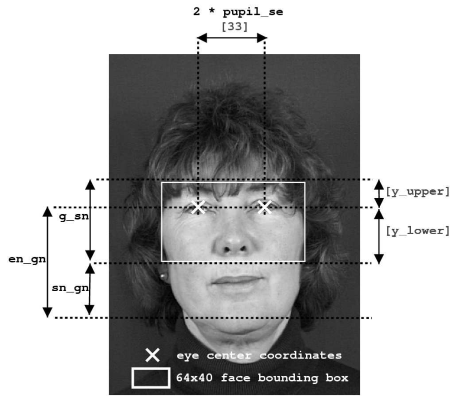

Figure 2. Face modeling using eyes center coordinates and facial anthropometry measures.

The face bounding box `w/h` crops the face from the glabella to the subnasale and do not includes the ears in order to minimize the influence of the hair-cut and of the lip movement.

The height `h` of the face is given by `y_upper+y_lower` where `y_lower = (en_gn - sn_gn)/s` and `y_upper = ((g_sn + sn_gn) - en_gn)/s`. In this model, the ratio `w/h` is equal to the ratio 64/40 and we force the eyes distance to be 33 pixels. Thus, the scale is `s = 2xpupil_se/33`.

The constants `pupil_se` (pupil-facial middle distance), `en_gn` (lower half of the craniofacial height), sn_gn (height of the lower face), and g_sn (distance between the glabella and the subnasale) can be found in [5].

3.1.2. Face pre-processing The extracted face is downsized to a 64x40 image. Then, we perform histogram normalization to modify the contrast of the image in order to enhance important features. Finally, we smooth the enhanced image by convolving a 3x3 Gaussian ($\sigma = 0.25$) in order to reduce the noise. After enhancement and smoothing (Fig. 3), the face image becomes a feature vector of dimension 2560.

Figure 3. Face pre-processing and symmetric transformation. From left to right: the original 64x40 pre-processed image, the mirrored image and the symmetric image.

3.1.3. Symmetric transformation The symmetric transformation is obtained simply by computing the average between the original image and the vertical mirrored image (Fig. 3). It generates a new virtual sample to enlarge the training and testing dataset. This transformation also normalizes the face by smoothing local deformations due to small out-of-plane rotations for instance.

3.1.4. Face representation It was chosen to represent the pre-processed input face into the LDA subspace, as described in [8]. The direct computation of the *LDA*-transform matrix is impractical because of the huge size of the face data in the original space (2560 dimensions). Therefore, a dimensionality reduction must be applied before solving the eigenproblem. This reduction is usually achieved by PCA.

PCA and LDA projection matrices have been computed on all images from XM2VTS database (295 identities and 8 images per identity). In the PCA space, the components accounting for $>= 4\%$ of the total variation are selected, reducing the dimensionality to 677. Then, the *LDA*-projection matrix is computed as described in [8] using all images of each identity projected into PCA subspace.

In the LDA space, the components accounting for $>= 1\%$ of the total variation are selected, reducing the dimensionality to 205.

3.2. Classification

Our face verification method is based on Multi-Layer Perceptrons (MLPs). MLPs are learning machines used in many classification problems [3].

For each client, an MLP is trained to classify an input to be either the given client or not. The input of the MLP is a feature vector corresponding to the projection of the face image into the LDA subspace. The output of the MLP is either 1 (if the input corresponds to a client) or -1 (if the input corresponds to an impostor). The MLP is trained using both client images and impostor images, often taken to be the images corresponding to other available clients. In the present study, we used the 300 client images from the Spanish part of the BANCA database (see next section).

Finally, the decision to accept or reject a client access depends on the score obtained by the corresponding MLP which could be either above (accept) or under (reject) a given threshold, chosen on a separate validation set to optimize a given criterion.

4. Experimental results

4.1. The BANCA database and protocol

This section gives an overview of the BANCA database and protocol, but a detailed description can be found in [2].

4.1.1. The database The BANCA database was designed in order to test multi-modal identity verification with various acquisition devices (2 cameras and 2 microphones) and under several scenarios (controlled, degraded and adverse).

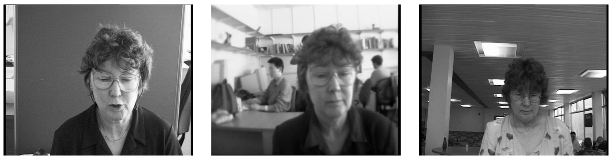

Figure 4. Examples of images from the BANCA database for each scenario. From left to right: controlled, degraded and adverse.

For 5 different languages[1], video and speech data were collected for 52 subjects (26 males and 26 females), i.e. a total of 260 subjects. Each language - and gender - specific population was itself subdivided into 2 groups of 13 subjects (denoted $g1$ and $g2$).

Each subject participated to 12 recording sessions, each of these sessions containing 2 records: 1 true *client access* (T) and 1 informed[2] *impostor attack* (I). For the image part of the database, there is 5 shots per record. The 12 sessions were separated into 3 different scenarios (Fig. 4): *controlled* (for sessions 1-4), *degraded* (for sessions 5-8), and *adverse* (for sessions 9-12).

4.1.2. The protocol In the BANCA protocol, we consider that the true client records for the first session of each condition is reserved as training material, i.e. record T from sessions 1, 5 and 9. In all our experiments, the client model training (or template learning) is done on at most these 3 records. We then consider four distinct training-test configurations, depending on the actual conditions corresponding to the training and to the testing conditions. The configurations are Matched Controlled (Mc), Unmatched Degraded (Ud), Unmatched Adverse (Ua)and Pooled test (P).

4.1.3. Performance measures In order to visualize the performance of the system, irrespectively of its operating condition, we use the conventional DET curve [11], which plots on a log-deviate scale the *False Rejection Rate* FRR as a function of the *False Acceptance Rate* FAR. Traditionally, the point on the DET curve corresponding to $FRR = FAR$ is called EER (Equal Error Rate) and is used to measure the closeness of the DET curve to the origin.

Thus, we measure the performance of the system for 3 specific operating conditions, corresponding to 3 different values of the Cost Ratio $R = C_{FA}/C_{FR}$, namely $R = 0.1$ (FA is an order of magnitude less harmful than a FR), $R = 1$ (FA and FR are equally harmful) and $R = 10$ (FA is an order of magnitude more harmful than FR). Assuming equal *a priori* probabilities of genuine clients and impostor, these situations correspond to 3 quite distinct cases:

When R is fixed and when FRR and FAR are given, we define the Weighted Error Rate (WER) as:

$$WER(R) = \frac{FRR + R\,FAR}{1 + R} \qquad (7)$$

FRR and FAR (and thus $WER(R)$) vary with the value of the decision threshold Θ, and Θ is usually optimized so as to minimize the $WER(R)$ on the development set. The *a priori threshold* thus obtained is always less efficient than the *a posteriori threshold* that optimizes the WER on the evaluation set itself.

1 English, French, German, Italian and Spanish

2 The actual speaker knew the text that the claimed identity speaker was supposed to utter.

4.2. Results

In this section, we provide experimental[3] results obtained by our approach, namely LDA/MLP, that we compare to state-of-the-art results [9] published on the BANCA database.

4.2.1. Experiments 1 First, we compare results obtained with the symmetric virtual sample (LDA/MLP^{+}) and without the symmetric virtual sample (LDA/MLP^{-}). We report in Table 1 the average (on groups g1 and g2) FAR/FRR and WER(1) of the above methods on the evaluation set when the *a posteriori* threshold was chosen at the EER on the development set.

LDA/MLP^{-}			LDA/MLP^{+}		
FAR	FRR	WER(1)	FAR	FRR	WER(1)
15.38	15.81	15.59	14.26	14.74	**14.5**

Table 1. Comparative results between LDA/MLP^{-} and LDA/MLP^{+} on protocol P.

Table 1 shows that the use of symmetric virtual samples improves the performance of the LDA/MLP based system. The symmetric transformation brings more variability to the training and testing datasets, but also normalizes small out-of-plane rotations. We provide also the corresponding DET curves (Fig. 5).

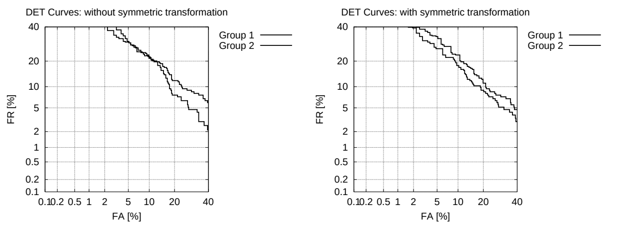

Figure 5. DET curves for experiments using LDA/MLP. From left to right: without symmetric transformation (LDA/MLP^{-}) and with symmetric transformation (LDA/MLP^{+}).

3 The machine learning library used for all experiments is Torch *http://www.torch.ch*.

4.2.2. Experiments 2 Second, we compare LDA/MLP$^+$ to the methods describe in [9], namely ORG/SVM and LDA/SVM respectively. ORG/SVM is using the original face image of size 61x57 as input of a Support Vector Machine (SVM) [4] and LDA/SVM is using the projection of the same face image into LDA subspace also as input of a SVM.

ORG/SVM

Protocol	FAR	FRR	WER(1)
Mc	2.18	6.92	**4.55**
Ua	6.79	41.35	24.07
Ud	5.77	34.9	20.34
P	4.91	27.72	16.32

LDA/SVM

Protocol	FAR	FRR	WER(1)
Mc	0.58	11.03	5.8
Ua	2.69	66.46	34.55
Ud	1.92	62.44	32.17
P	1.73	46.62	24.17

LDA/MLP$^+$

Protocol	FAR	FRR	WER(1)
Mc	4.8	5.13	4.96
Ua	15.86	18.58	**17.22**
Ud	12.02	12.82	**12.42**
P	14.26	14.74	**14.5**

Table 2. Comparative results between ORG/SVM, LDA/SVM and LDA/MLP$^+$.

We report in Table 2 the average (on groups g1 and g2) FAR/FRR and WER(1) of the above methods on the evaluation set when the *a posteriori* threshold was chosen at the EER on the development set. We provide also the corresponding DET curves (Fig. 6) of the LDA/MLP$^+$ method only.

Table 2 shows that LDA/MLP$^+$ performs much better than the two other methods on the difficult unmatched protocols Ua and Ud. LDA/MLP$^+$ is close to ORG/SVM on the easiest protocol Mc and globally performs better on the pooled test protocol P. We provide also, for future comparisons, results obtained by LDA/MLP$^+$ according to configurations Mc, Ua, Ud, P of the BANCA protocol (Table 3) when minimizing the WER for each cost ratio. These results show that an average WER of 1.9 can be reached with our method when choosing a cost ratio equal to 10.

Group g1

	R=0.1		
Protocol	**FAR**	**FRR**	**WER**
Mc	10.577	3.846	4.458
Ua	78.846	1.282	8.333
Ud	25	8.974	10.431
P	53.205	1.282	6.002
	R=1		
Protocol	**FAR**	**FRR**	**WER**
Mc	3.846	8.974	6.41
Ua	15.385	11.538	13.462
Ud	14.423	11.538	12.981
P	12.179	14.103	13.141
	R=10		
Protocol	**FAR**	**FRR**	**WER**
Mc	1.923	12.821	**2.914**
Ua	1.923	44.872	5.828
Ud	3.846	28.205	6.061
P	0.321	48.718	4.72

Group g2

	R=0.1		
Protocol	**FAR**	**FRR**	**WER**
Mc	16.346	1.282	2.652
Ua	42.308	8.974	12.005
Ud	50	1.282	5.711
P	52.244	2.991	7.469
	R=1		
Protocol	**FAR**	**FRR**	**WER**
Mc	2.885	2.564	2.724
Ua	13.462	28.205	20.833
Ud	9.615	19.231	14.423
P	14.103	17.094	15.598
	R=10		
Protocol	**FAR**	**FRR**	**WER**
Mc	0.0	10.256	**0.932**
Ua	0.962	64.103	6.702
Ud	3.846	48.718	7.925
P	0.962	52.991	5.692

Table 3. FAR, FRR and WER for each cost ratio on the evaluation set using LDA/MLP$^+$.

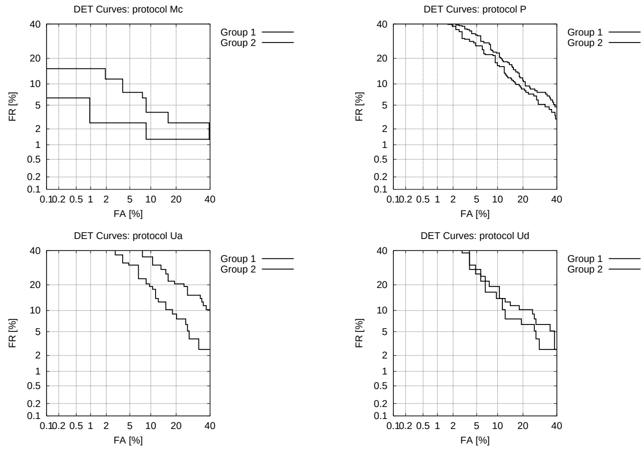

Figure 6. DET curves for experiments using LDA/MLP$^+$. From left to right on the first row: protocols Mc and P. From left to right on the second row: protocols Ua and Ud.

5. Conclusion

In this paper, a detailed system for face verification was presented. It was describing in detail each stage of the system: the modeling of the face (a 64x40 face image), the extraction of relevant features (Fisher Linear Discriminant) and the classification of the input face as a client or an impostor using a MLP. We also proposed to use a transformation based on symmetry that generate a new virtual sample in order to enlarge the training and testing dataset of a face verification system.

Experiments were carried out on the BANCA benchmark multi-modal database using its experimental protocol. The BANCA database was designed in order to test multimodal identity verification with various acquisition devices and under several scenarios (controlled, degraded and adverse). The BANCA protocol allows to measure the performance in varied conditions with only one (controlled) training session and the degradation from a mismatch between controlled training and uncontrolled test,

Results have shown that the proposed symmetric virtual sample improves the performance on the pooled test protocol. It has been shown also that this approach performs better than the state-of-the-art on unmatched protocols and globally on the pooled test protocol.

Acknowledgments

The author wants to thank the Swiss National Science Foundation for supporting this work through the National Center of Competence in Research (NCCR) on "Interactive Multimodal Information Management (IM2)".

References

[1] P. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In *ECCV'96*, pages 45–58, 1996. Cambridge, United Kingdom.

[2] S. Bengio, F. Bimbot, J. Mariéthoz, V. Popovici, F. Porée, E. Bailly-Baillière, G. Matas, and B. Ruiz. Experimental Protocol on the BANCA database. Technical Report IDIAP-RR 02-05, IDIAP, 2002.

[3] C. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[4] C. J. C. Burges. A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):1–47, 1998.

[5] L. Farkas. *Anthropometry of the Head and Face*. Raven Press, 1994.

[6] R. Féraud, O. Bernier, J.-E. Viallet, and M. Collobert. A fast and accurate face detector based on Neural Networks. *Transactions on Pattern Analysis and Machine Intelligence*, 23(1), 2001.

[7] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(II):179–188, 1936.

[8] K. Jonsson, J. Matas, J. Kittler, and Y. Li. Learning support vectors for face verification and recognition. In *4th International Conference on Automatic Face and Gesture Recognition*, pages 208–213, 2000.

[9] A. Kostin, M. Sadeghi, J. Kittler, and K. Messer. On representation spaces for SVM based face verification. In *Proceedings of the COST275 Workshop on The Advent of Biometrics on the Internet*, Rome, Italy, 2002.

[10] S. Marcel and S. Bengio. Improving face verification using skin color information. In *Proceedings of the 16th ICPR*. IEEE Computer Society Press, 2002.

[11] 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, Rhodes, Greece*, pages 1895–1898, 1997.

[12] N. Poh, S. Marcel, and S. Bengio. Improving face authentication using virtual samples. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.

[13] H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-based face detection. *Transactions on Pattern Analysis and Machine Intelligence*, 20(1), 1998.

[14] P. Verlinde, G. Chollet, and M. Acheroy. Multi-modal identity verification using expert fusion. *Information Fusion*, 1:17–33, 2000.