IDIAP RESEARCH REPORT

FACE VERIFICATION USING LDA AND MLP ON THE BANCA DATABASE

Sébastien Marcel [a]

IDIAP–RR 03-66

DECEMBER 2003

SUBMITTED FOR PUBLICATION

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] IDIAP, CP 592, 1920 Martigny, Switzerland

IDIAP Research Report 03-66

FACE VERIFICATION USING LDA AND MLP ON THE BANCA DATABASE

Sébastien Marcel

DECEMBER 2003

SUBMITTED FOR PUBLICATION

Abstract. In this paper, we propose a system for face verification. It describes in detail each stage of the system: the modeling of the face, the extraction of relevant features and the classification of the input face as a client or an impostor. This 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 show that our approach perform better than the state-of-the-art on the same database. Experiments show also contradictory results in the state-of-the-art literature.

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 [21]. 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 and we present the current state-of-the-art approach. Then, in section 3 we present the proposed approach, a LDA (Linear Discriminant Analysis) feature extraction technique, successfully applied to face verification [13], together with a MLP (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

2.1 Problem Description

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.

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. A good survey on the different methods used in face verification can be found in [7, 22].

2.2 State-of-the-art approach

This section, briefly introduces one of the best method [12, 13]. In this method, faces are represented in both Principal Component and Linear Discriminant subspaces and the main decision tool is Support Vector Machines (SVMs) [5].

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 than the PCA subspace. The LDA based features for personal identity verification is theoretically superior to that achievable with the features computed using PCA [20] and many others [2, 8].

2.2.1 Linear Discriminant.

A linear discriminant is a simple linear projection of the input vector onto an output dimension:

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

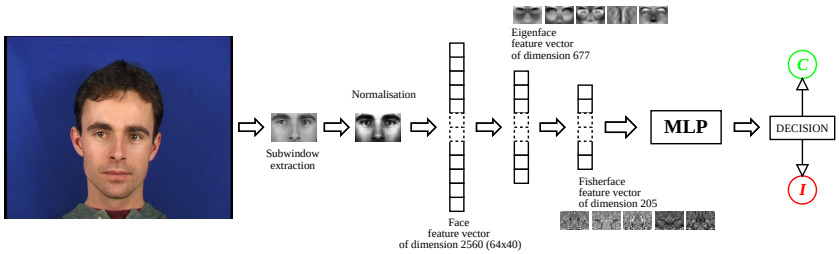

Figure 1: Face Verification using LDA and MLP

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

2.2.2 Fisher Linear Discriminant.

The Fisher criterion [10] 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

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

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 or its projection into PCA or LDA subspace [13, 14, 1]. In this section, we describe our approach a MLP classifier trained on a gray-scale face image projected into LDA subspace (Fig. 1).

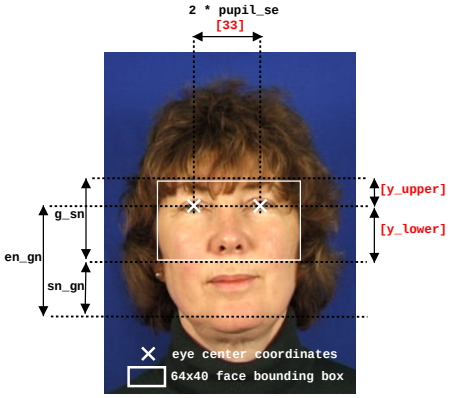

Figure 2: Face modeling using eyes center coordinates and facial anthropometry measures.

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 [18, 17], 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 [9] according to a face model (Fig. 2).

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` = 2x`pupil_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 [9].

3.1.2 Face Pre-Processing.

First, 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.

Figure 3: Face pre-processing. From left to right: the original 64x40 image, the histogram normalized image and the smoothed image.

3.1.3 Face Representation.

After enhancement and smoothing, the face image becomes a feature vector of dimension 2560. We have decided to choose the state-of-the-art face representation describe in the previous section, namely LDA.

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 describe in the previous section 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. A good introduction to machine learning algorithms can be found in [4, 11].

3.2.1 Multi-Layer Perceptrons.

We will assume that we have access to a training dataset of l pairs $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i$ is a vector containing the pattern, while y_i is the class of the corresponding pattern often coded respectively as 1 and -1.

A MLP is a particular architecture of artificial neural networks composed of layers of non-linear but differentiable parametric functions. For instance, the output $\hat{y}$ of a 1-hidden-layer MLP can be written mathematically as follows

$$\hat{y} = b + \mathbf{w} \cdot \tanh\left(\mathbf{a} + \mathbf{x} \cdot \mathbf{V}\right) \tag{7}$$

where the estimated output $\hat{y}$ is a function of the input vector $\mathbf{x}$, and the parameters $\{b, \mathbf{w}, \mathbf{a}, \mathbf{V}\}$. In this notation, the non-linear function $\tanh()$ returns a vector which size is equal to the number of hidden units of the MLP, which controls its capacity and should thus be chosen carefully, by cross-validation for instance.

An MLP can be trained by gradient descent using the backpropagation algorithm [19] to optimize any derivable criterion, such as the *mean squared error* (MSE):

$$\text{MSE} = \frac{1}{l} \sum_{i=1}^{l} (y_i - \hat{y}_i)^2 \ . \tag{8}$$

3.2.2 MLP for Face Verification.

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 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 [3].

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* (c) for sessions 1-4,
- *degraded* (d) for sessions 5-8,
- *adverse* (a) for sessions 9-12.

Two cameras were used, a cheap one and an expensive one. The cheap camera was used in the degraded scenario, while the expensive camera was used for controlled and adverse scenarios. Two microphones, a cheap one and an expensive one, were used simultaneously in each of the three scenarios. During the recordings, the camera was placed on the top of the screen and the two microphones were placed in front of the monitor and below the subject chin.

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 7 distinct training-test configurations, depending on the actual conditions corresponding to the training and to the testing conditions.

- Matched controlled (Mc):
    - client training from 1 controlled session
    - client and impostor testing from the other controlled sessions (within the same group)
- Matched degraded (Md):
    - client training from 1 degraded session
    - client and impostor testing from the other degraded sessions (within the same group)
- Matched adverse (Ma):
    - client training from 1 adverse session
    - client and impostor testing from the other adverse sessions (within the same group)

[1] English, French, German, Italian and Spanish

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

- Unmatched degraded (Ud):
    - client training from 1 controlled session
    - client and impostor testing from degraded sessions (within the same group)
- Unmatched adverse (Ua):
    - client training from 1 controlled session
    - client and impostor testing from adverse sessions (within the same group)
- Pooled test (P):
    - client training from 1 controlled session
    - client and impostor testing from all conditions sessions (within the same group)
- Grand test (G):
    - client training from 1 controlled, 1 degraded and 1 adverse sessions
    - client and impostor testing from all conditions sessions (within the same group)

From the comparison of these various performances, it is possible to measure: the intrinsic performance in a given condition, the degradation from a mismatch between controlled training and uncontrolled test, the performance in varied conditions with only one (controlled) training session, and the potential gain that can be expected from more representative training conditions.

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 [15], which plots on a log-deviate scale the *False Rejection Rate* P_{FR} as a function of the *False Acceptance Rate* P_{FA}. Traditionally, the point on the DET curve corresponding to $P_{FR} = P_{FA}$ is called EER (Equal Error Rate) and is used to measure the closeness of the DET curve to the origin. The EER value of an experiment is reported on the DET curve, to comply with this tradition.

We also 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 $R = 0.1$, $R = 1$, $R = 10$. Assuming equal *a priori* probabilities of genuine clients and impostor, these situations correspond to 3 quite distinct cases:
$R = 0.1$ $\rightarrow$ a FA is an order of magnitude less harmful than a FR
$R = 1$ $\rightarrow$ a FA and a FR are equally harmful
$R = 10$ $\rightarrow$ a FA is an order of magnitude more harmful than a FR.

When R is fixed and when P_{FR} and P_{FA} are given, we define the Weighted Error Rate (WER) as:

$$WER(R) = \frac{P_{FR} + R\, P_{FA}}{1 + R} \tag{9}$$

P_{FR} and P_{FA} (and thus WER) vary with the value of the decision threshold Θ, and Θ is usually optimized so as to minimize WER on the development set D:

$$\hat{\Theta}_R = \arg\min_{D} \; WER(R) \tag{10}$$

The *a priori threshold* thus obtained is always less efficient than the *a posteriori threshold* that optimizes the WER on the evaluation set E itself:

$$\Theta^*_R = \arg\min_{E} \; WER(R) \tag{11}$$

Table 1: FAR, FRR and WER for each cost ratio on the test set using LDA/MLP.

Group g1

	R=0.1			R=1			R=10		
Protocol	**FAR**	**FRR**	**WER**	**FAR**	**FRR**	**WER**	**FAR**	**FRR**	**WER**
Mc	13.462	7.692	8.217	4.808	12.821	8.814	0.000	29.487	**2.681**
Ua	65.385	1.282	7.110	17.308	6.410	11.859	4.808	41.026	8.100
Ud	42.308	1.282	5.012	14.423	12.821	13.622	2.885	34.615	5.769
P	48.397	1.282	5.565	18.269	8.974	13.622	3.205	39.316	6.488

Group g2

	R=0.1			R=1			R=10		
Protocol	**FAR**	**FRR**	**WER**	**FAR**	**FRR**	**WER**	**FAR**	**FRR**	**WER**
Mc	41.346	0.000	3.759	2.885	6.410	4.647	0.962	10.256	**1.807**
Ua	38.462	12.821	15.152	18.269	17.949	18.109	0.962	53.846	5.769
Ud	49.038	2.564	6.789	9.615	24.359	16.987	0.962	52.564	5.653
P	50.962	3.419	7.741	19.551	11.966	15.759	0.321	52.991	5.109

Table 2: Comparative results between ORG/SVM, LDA/SVM and LDA/MLP.

	ORG/SVM			**LDA/SVM**			**LDA/MLP**		
Protocol	**FAR**	**FRR**	**HTER**	**FAR**	**FRR**	**HTER**	**FAR**	**FRR**	**HTER**
Mc	2.18	6.92	**4.55**	0.58	11.03	5.8	5.77	7.05	6.41
Ua	6.79	41.35	24.07	2.69	66.46	34.55	15.86	15.38	**15.62**
Ud	5.77	34.9	20.34	1.92	62.44	32.17	12.5	14.10	**13.3**
P	4.91	27.72	16.32	1.73	46.62	24.17	15.38	15.81	**15.59**

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 [1] published on the BANCA database.

First, we provide for future comparison results obtained by LDA/MLP according to configurations Mc, Ua, Ud, P of the BANCA protocol (Table 1). These results show that an average WER of 2.24 can be reached with our method when chosing a cost ratio equal to 10.

Second, we compare LDA/MLP to the methods describe in [1], namely ORG/SVM and LDA/SVM respectively.

ORG/SVM is using the original face image of size 61x57 as input of a SVM and LDA/SVM is using the projection of the same face image into LDA subspace also as input of a SVM. We report in Table 2 the average (on groups g1 and g2) FAR/FRR and HTER of the above methods on the test set. We provide also the corresponding DET curves (Fig. 5) 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 not as good as ORG/SVM on the easiest protocol Mc but globally performs slightly better on the pooled test protocol P.

Furthermore, it appears a contradiction about the results using LDA. In [1], it was shown that ORG/SVM was better than LDA/SVM. It was also conclude that “projecting the image into PCA and LDA spaces does not improve the performance of the system using SVM”. This conclusion is contradictory with previous results

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

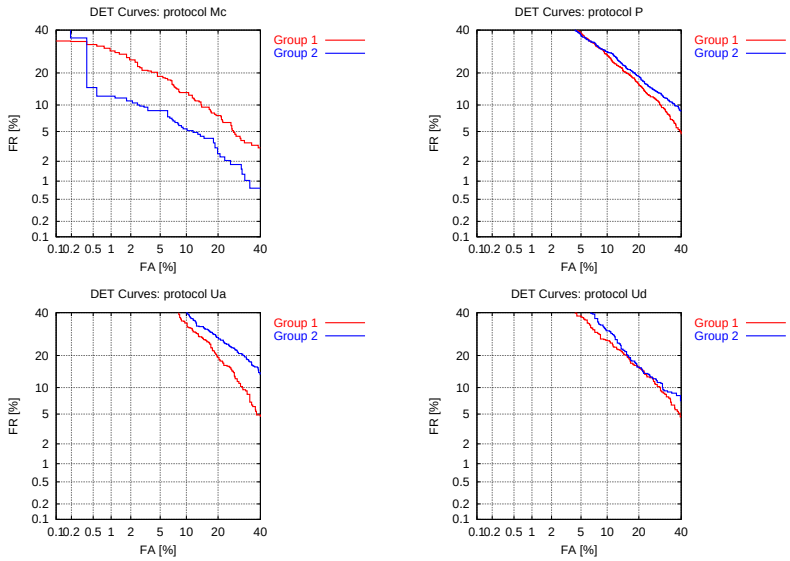

Figure 5: DET curves for experiments using LDA/MLP. From left to right on the first row: protocols Mc and P. From left to right on the second row: protocols Ua and Ud

on the XM2VTS database [12, 13, 16]. MLP and SVM provide comparative results in face verification [6]. Thus, the main difference between the work presented in this paper and [1] is not the feature extraction method, which is the same, nor the choice of the classifier (MLP or SVM) but certainly in the choice of the face model. In this study, we chose to crop the face from the glabella to the subnasale and not to include the ears in order to minimize the influence of the hair-cut and of the lip movement.

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.

Experiments were carried out on the BANCA benchmark multi-modal database using its experimental protocol. The BANCA database was designed in order to test multi-modal 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 our approach performs better than the state-of-the-art on unmatched protocols and globally on the pooled test protocol. It has been shown also that this performance improvement may be due to the choice of the face model.

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)”. This work was also funded by the European project ”BANCA”, through the Swiss Federal Office for Education and Science (OFES).

References

[1] J. Kittler A. Kostin, M. Sadeghi 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.

[2] 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.

[3] 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.

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

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

[6] F. Cardinaux and S. Marcel. Face verification using MLP and SVM. In *Proceedings of the COST275 Workshop on The Advent of Biometrics on the Internet*, Rome, Italy, 2002.

[7] R. Chellappa, C.L Wilson, and C.S Barnes. Human and machine recognition of faces: A survey. Technical Report CAR-TR-731, University of Maryland, 1994.

[8] Pierre A. Devijver and Josef Kittler. *Pattern Recognition: A Statistical Approach*. Prentice-Hall, Englewood Cliffs, N.J., 1982.

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

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

[11] S. Haykin. *Neural Networks, a Comprehensive Foundation, second edition*. Prentice Hall, 1999.

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

[13] Y. Li, J. Kittler, and J. Matas. On matching scores of LDA-based face verification. In T. Pridmore and D. Elliman, editors, *Proceedings of the British Machine Vision Conference BMVC2000*. British Machine Vision Association, 2000.

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

[15] 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.

[16] J. Matas, M. Hamouz, K. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, W. Gerstner, S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Comparison of face verification results on the XM2VTS database. In A. Sanfeliu, J. J. Villanueva, M. Vanrell, R. Alqueraz, J. Crowley, and Y. Shirai, editors, *Proceedings of the 15th ICPR*, volume 4, pages 858–863. IEEE Computer Society Press, 2000.

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

[18] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural Network-based face detection. *Transactions on Pattern Analysis and Machine Intelligence*, 20(1), 1998.

[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and James L. McClelland, editors, *Parallel Distributed Processing*, volume 1. MIT Press, Cambridge, MA., 1986.

[20] M. Turk and A. Pentland. Eigenface for recognition. *Journal of Cognitive Neuro-science*, 3(1):70–86, 1991.

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

[22] J. Zhang, Y. Yan, and M. Lades. Face recognition: Eigenfaces, Elastic Matching, and Neural Nets. In *Proceedings of IEEE*, volume 85, pages 1422–1435, 1997.