IDIAP RESEARCH REPORT

Face Authentication using Client-Specific Matching Pursuit

Sébastien Marcel [a] Philippe Jost [b]
Pierre Vandergheynst [b]
Jean-Philippe Thiran [b]

IDIAP–RR 04-78

December 2004

submitted for publication

[a] IDIAP Research Institute, Martigny, Switzerland 1920
[b] Signal Processing Institute (EPFL-ITS), Lausanne, Switzerland 1015

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 04-78

Face Authentication using Client-Specific Matching Pursuit

Sébastien Marcel Philippe Jost Pierre Vandergheynst
Jean-Philippe Thiran

December 2004

submitted for publication

Abstract. In this paper, we address the problem of finding image decompositions that allow good compression performance, and that are also efficient for face authentication. We propose to decompose the face image using Matching Pursuit and to perform the face authentication in the compressed domain using a MLP (Multi-Layer Perceptron) classifier. We provide experimental results and comparisons with PCA and LDA systems on the multi-modal benchmark database BANCA using its associated protocol.

1 Introduction

Security is a major issue in our modern society. Specifically, identity authentication is one of the most important aspects of security management. Biometrics identification technology represents an extraordinary automatic mean to positively identify a person, and thus to complement the current authentication and access control protocols. In contrary to other biometric solutions, due to their absence of contact and their non-invasiveness, face recognition, as well as speech recognition, are viewed as excellent solutions for biometrics authentication for widely spread applications such as authentication for banking, security system access, advanced video surveillance, video annotation. In addition, for these two modalities the acquisition systems are very simple, cheap, and virtually free for future multimedia applications.

To enable the design of efficient identification/authentication applications, it is important to avoid unnecessary transcoding operations, or moving between different data representations. In the same time, the size of biometrics information databases imposes drastic compression requirements on storage and transmission of identification data. Classically, the compression of these data involves one kind of representation (e.g. DCT, wavelets,.) whilst identification/authentication generally involves a different one (e.g. eigenfaces, fisherfaces). It is then of major importance to find decompositions that allow good compression performance and that are also efficient for identification/authentication processes. The recognition can therefore be performed directly in the compressed domain, thus heavily reducing the computation overhead.

The paper is structured as follows. In section 2 we introduce the reader to the problem of identity authentication and we present the current state-of-the-art approaches. In section 3 we provide a description of image compression using the Matching Pursuit (MP) algorithm. Then, in section 4, we present the proposed approach, a feature extraction technique based on Matching Pursuit, together with a MLP (Multi-Layer Perceptron) classifier. In section 5, we provide experimental results and comparisons with PCA and LDA systems on the multi-modal benchmark database BANCA using its associated protocol. Finally, we analyze the results and conclude.

2 Face Authentication

The goal of an *automatic identity authentication system* is to either accept or reject the identity claim made by a given person. Biometric identity authentication systems are based on the characteristics of a person, such as its face, fingerprint or signature. Identity authentication 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.

2.1 Problem Description

An identity authentication 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*.

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

Many approaches have been used for face recognition using *holistic* approaches such as Eigenfaces [20], Fisherfaces [2], Multi-Layer Perceptrons [13], Support Vector Machines (SVMs) [10] or *local* approaches such as Elastic Graph Matching [21], Hidden Markov Models [15] and Gaussian Mixture Models [4].

2.2 State-of-the-art approach

This section, briefly introduces one of the best method [14]. In this method, faces are represented in both Principal Component and Linear Discriminant subspaces.

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 [2].

A linear discriminant is a simple linear projection of the input vector onto an output dimension. Depending on the criterion chosen to select the optimal parameters, one could obtain a different solution. The Fisher criterion [7] aims at maximizing the ratio of between-class scatter to within-class scatter.

3 Matching Pursuit

One of the ultimate goals in image representation is to find an efficient and natural way to manipulate data. A strong emphasis has been put on the search for *sparse approximations*, i.e. techniques yielding good approximations of images with very few terms. Wavelets for example do not yield good sparse approximations of images because they fail at efficiently capturing edges. These limitations can be overcome by techniques using redundant basis of functions to represent the images. A function belonging to this basis is called *atom*. The dictionary $\mathcal{D}$ is the overcomplete set of all atoms, and can be written as $\mathcal{D} = \{g_{\vec{\gamma}}\}_{\vec{\gamma}\in\Gamma}$ with $\|g_{\vec{\gamma}}\| = 1$. In the case of redundant expansions for images, the atoms are bi-dimensional functions. They are often chosen to match features contained in the scene as edges for example. The design of a dictionary depends on the application and on the purpose to fulfill. In this paper, we used the dictionary described in [16]. The atoms of the dictionary are built from a generating function that is scaled, rotated, translated and bended. The generating function (1) is made of a Gaussian in one direction and its second derivative in the other direction. It has a good ability to capture edges in the images.

$$g(x,y) = \frac{2}{\sqrt{3\pi}}(4x^2 - 2)\exp -(x^2 + y^2). \tag{1}$$

Greedy algorithms iteratively construct an approximate by selecting the element of a dictionary of waveforms that best matches the signal at each iteration. The pure greedy algorithm is known as *Matching Pursuit* [12]. Assuming that all atoms in the dictionary $\mathcal{D}$ have unit norm, we initialize the algorithm by setting the initial residual $R_0 = s$ where s is the signal to approximate. Initially, the signal is decomposed as

$$R_0 = \langle g_{\gamma_0}, R_0\rangle g_{\gamma_0} + R_1\,.$$

Clearly g_{γ_0} is orthogonal to R_1 and we thus have

$$\|R_0\|^2 = |\langle g_{\gamma_0}, R_0\rangle|^2 + \|R_1\|^2\,.$$

If we want to minimize the energy of the residual R_1 we must maximize the projection $|\langle g_{\gamma_0}, R_0\rangle|$. At the next step, we simply apply the same procedure to R_1, which yields

$$R_1 = \langle g_{\gamma_1}, R_1\rangle g_{\gamma_1} + R_2\,,$$

where g_{γ_1} maximizes $|\langle g_{\gamma_1}, R_1\rangle|$. Iterating this procedure, we thus obtain an approximate after M steps:

$$s = \sum_{m=0}^{M-1} \langle g_{\gamma_m}, R_m\rangle g_{\gamma_m} + R_M\,, \tag{2}$$

where the norm of the residual (approximation error) satisfies

$$\|R_M\|^2 = \|s\|^2 - \sum_{m=0}^{M-1} |\langle g_{\gamma_m}, R_m \rangle|^2 .$$

One can easily show that Matching Pursuit (MP) converges [9] and even converges exponentially in the strong topology in finite dimension (see [12] for a proof). Very recently, a new wealth of constructive results added to the interest of greedy algorithms, or more generally to sparse approximations in redundant libraries [19].

4 The Proposed Approach

In face authentication, 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 [13, 1] or its projection into PCA or LDA subspace [11, 1]. In most of these studies, MLP or SVM classifiers have already been used.

In this section, we describe our approach a MLP classifier trained on a gray-scale face image projected into MP subspace (Figure 1).

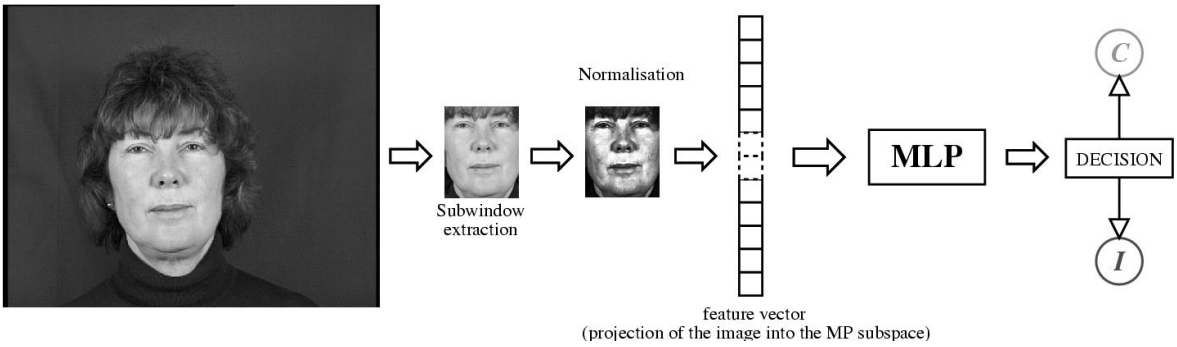

Figure 1: Face Authentication using Matching Pursuit and MLP

4.1 Feature Extraction

Face Modeling. In a real application, the face bounding box will be provided by an accurate face detector [17], but here the bounding box is computed using manually located eyes coordinates, assuming a perfect face detection. The face modeling is a critical stage which is, unfortunately, rarely described in most of the papers. It is thus almost impossible to reproduce the experiments. In this paper, the face bounding box is determined using face/head anthropometry measures [6] according to a face model (Figure 2).

The face bounding box `w/h` crops the face approximately from the glabella (in order to minimize the influence of the hair-cut) to the chin and do not includes the ears.

The height `h` of the face is given by `y_upper+y_lower` where `y_lower` = 16 pixels and `y_upper` = 64 pixels. In this model, the ratio `w/h` is equal to the ratio 64/80 and we force the eyes distance to be 33 pixels. The constant `pupil_se` (pupil-facial middle distance) can be found in [6].

Face Pre-Processing. First, the extracted face is downsized to a 64x80 image. Then, we performed histogram normalization to modify the contrast of the image in order to enhance important features. Finally, we smoothed the enhanced image by convolving a 3x3 Gaussian ($\sigma = 0.25$) in order to reduce the noise.

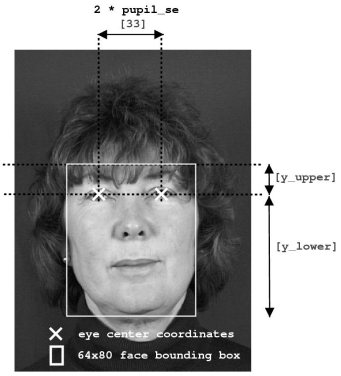

Figure 2: Face modeling using eyes center coordinates and facial anthropometry measures.

Face Representation. After pre-processing, the face image becomes a feature vector of dimension 5120. In the proposed approach, the face image is projected into the MP subspace spawn by a set of atoms.

Our objective is to find the minimum number of atoms that allows good compression of the face and also that is appropriate for face authentication. Figure 3 illustrates the compression of a face image using MP with different number of atoms.

Figure 3: Face compression using Matching Pursuit. From top-left to bottom-right: the original image, the reconstruction of the image with 10, 25, 50, 100, 150, 250, 500 and 1000 atoms.

We decided to adopt a client-specific approach where a face image will be decomposed by Matching Pursuit into a weighted sum of atoms from the dictionary (2) representing the identity claimed. Thus, every client will have its own atoms-based representation.

Let assume that we have a training set of T images for a specific identity noted as $\{s_t\}_{t\in[1\ T]}$. A modified version of the Matching Pursuit algorithm is used to find the atomic decomposition that best matches the training set. The atom at iteration i is such that

$$g_{\gamma_i} = \max_{g_k \in \mathcal{D}} \sum_{t=1}^{T} \langle g_k, R_i^t \rangle.$$

where R_i^t is the residual associated to s_t at iteration i. The weight w_i associated to the previously found atom is

$$w_i = \frac{1}{T} \sum_{t=1}^{T} \langle g_{\gamma_i}, R_i^t \rangle.$$

The residuals are updated.

$$R_{i+1}^t = R_i^t - w_i g_{\gamma_i}.$$

The initial residuals are $R_0^t = s_t, \forall t \in [1\ T]$. One can easily see that the previously described method leads to the same results as shown in section 3 if the training set contains only one element i.e. $T = 1$.

4.2 Classification

Our face authentication 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 [3, 8].

4.2.1 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

$$\hat{y} = b + \mathbf{w} \cdot \tanh\left(\mathbf{a} + \mathbf{x} \cdot \mathbf{V}\right) \tag{3}$$

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 back-propagation algorithm [18] to optimize any derivable criterion, such as the *mean squared error* (MSE):

$$\text{MSE} = \frac{1}{l} \sum_{i=1}^{l} (y_i - \hat{y}_i)^2 \ . \tag{4}$$

or more efficiently using an optimal criterion [5] designed for classification.

4.2.2 MP and MLP for Face Authentication

For each client, an MLP is trained to classify an input to be either the given client or not. The input $\mathbf{x}$ of the MLP is a feature vector corresponding to the projection of the face image X into the client-specific MP subspace $\mathbf{x} = [\langle g_{\gamma_0}, X \rangle, ...\langle g_{\gamma_i}, X \rangle, ...\langle g_{\gamma_N}, X \rangle]$ where N is the number of atoms.

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 images from the world model 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.

5 Experimental Results

In this section, we provide experimental results obtained by our approach, namely Matching Pursuit and Multi-Layer Perceptrons (MP/MLP), that we compare to two baseline systems, PCA and LDA both using Multi-Layer Perceptrons, respectively PCA/MLP and LDA/MLP.

5.1 The Database

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

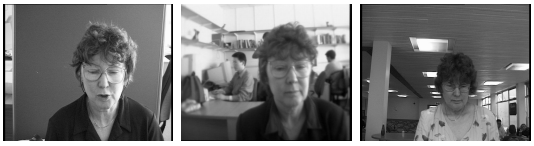

Figure 4: Examples of images from the BANCA database for each scenario. From left to right: controlled, degraded and adverse.

Video and speech data were collected for 52 subjects (26 males and 26 females). Each 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 [1] *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).

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.

5.2 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 consider the following protocol, namely Pooled test (P) protocol. One controlled session is used for client training. There is, thus, only 5 images per client for training. All conditions sessions (within the same group) are used for client and impostor testing.

5.3 Performance Measures

We measure the performance of the system using the Half Total Error Rate ($HTER$) defined as:

$$HTER = (FR + FA)/2 \tag{5}$$

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

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

5.4 Results and Discussions

We report in Table 1 and Table 2 the average (on groups g1 and g2) FAR/FRR and HTER of the above methods on the evaluation set.

Table 1 indicates the HTER in function of the number of atoms. Experiments were performed only with 50, 100, 150, 250 and 500 atoms because of the expensive computational cost required to train all the client models (client-specific MP and MLP). It shows clearly that the authentication error goes down, to a certain point, with the number of atoms. We note that the best result is reached for 250 atoms and that the error increases for 500 atoms. Note that the number of atoms have a direct impact on the number of parameters of the MLP. Note also that the number of hidden units is chosen on the development set and not on the evaluation set. As a consequence, at a certain point the accuracy of the MLP will decrease when the input dimension will increase because there are too many parameters to estimate.

# atoms	FAR	FRR	HTER
50	20.19	21.36	20.77
100	15.38	16.66	16.02
150	13.78	16.88	15.33
250	12.98	15.60	**14.29**
500	14.42	16.23	15.32

Table 1: Results using MP/MLP with 50, 100, 150, 250 and 500 atoms.

Table 2 provides for comparison results obtained by PCA and LDA. The PCA matrix has been computed on 17′800 faces (377 different identities) from the XM2VTS database and world models of the BANCA database. We keep the eigenvectors corresponding to the biggest non-zero eigenvalues that account for 95% of the total variance. This lead to 448 eigenvectors and thus to 448 input dimensions for the MLP based on PCA.

The LDA matrix was not computed directly on face images, but as usual on the projection of the face image into the PCA subspace (more precisely the above PCA matrix). The number of eigenvectors is 104 corresponding to 90% of the total variance.

PCA/MLP			LDA/MLP		
FAR	FRR	HTER	FAR	FRR	HTER
12.5	14.32	13.41	10.74	13.67	**12.2**

Table 2: Comparative results with baseline PCA_{448}/MLP and LDA_{104}/MLP.

From the results, we observe that the best results are obtained first by LDA and then second by PCA. However, the difference between LDA and PCA is 1.21% and the difference between MP and PCA is less 0.88%. It shows that Matching Pursuit performs nearly as well as PCA. Furthermore, we computed that the reconstruction error, in terms of MSE, is lower for MP (0.0013) than for PCA 0.0024 with 448 atoms/components[2].

6 Conclusion

In this paper, we addressed the problem of face authentication. We proposed to decompose the face image using the Matching Pursuit (MP) algorithm and to perform the face authentication in the

[2]The MSE have been computed on 178 faces randomly chosen from the large dataset of 17′800.

compressed domain using a Multi-Layer Perceptron (MLP) classifier. For each client, a client-specific image decomposition is found using MP and a MLP is trained.

We provide experimental results and comparisons with PCA and LDA systems on the multi-modal benchmark database BANCA using its associated protocol. Results show that Matching Pursuit performs well compared to PCA or LDA. LDA is still performing slightly better. However, we think that our approach based on Matching Pursuit is promising and that it can be improved.

Acknowledgments

The authors wish 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] 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] C. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

[4] F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of MLP and GMM classifiers for face verification on XM2VTS. In *Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication.* Springer-Verlag, 2003.

[5] R. Collobert and S. Bengio. A gentle hessian for efficient gradient descent. In *IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP*, 2004.

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

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

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

[9] L. K. Jones. On a conjecture of huber concerning the convergence of projection pursuit regression. *Annals of Statistics*, 15(2):880–882, June 1987.

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

[11] 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.

[12] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, Dec 1993.

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

[14] Kieron Messer, Josef Kittler, Mohammad Sadeghi, Miroslav Hamouz, Alexey Kostyn, Sebastien Marcel, Samy Bengio, Fabien Cardinaux, Conrad Sanderson, Norman Poh, Yann Rodriguez, Jacek Czyz, and al. Face authentication test on the BANCA database. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Cambridge, August 23-26 2004.

[15] A. Nefian and M. Hayes. Face recognition using an embedded HMM. In *Proceedings of the IEEE Conference on Audio and Video-based Biometric Person Authentication (AVBPA)*, pages 19–24, 1999.

[16] L. Peotta, L. Granai, and P. Vandergheynst. Very low bit rate image coding using redundant dictionaries. In *48th annual meeting*. SPIE, SPIE, August 2003.

[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] 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.

[19] Joel A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, October 2004.

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

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