IDIAP RESEARCH REPORT

# Face Authentication with Salient Local Features and Static Bayesian Network

Guillaume Heusch [a]    Sébastien Marcel [a]

IDIAP–RR 07-04

February 2007

[a] IDIAP Research Institute

**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 07-04

# Face Authentication with Salient Local Features and Static Bayesian Network

Guillaume Heusch Sébastien Marcel

February 2007

**Abstract.** In this paper, the problem of face authentication using salient facial features together with statistical generative models is adressed. Actually, classical generative models, and Gaussian Mixture Models in particular make strong assumptions on the way observations derived from face images are generated. Indeed, systems proposed so far consider that local observations are independent, which is obviously not the case in a face. Hence, we propose a new generative model based on Bayesian Networks using only salient facial features. We compare it to Gaussian Mixture Models using the same set of observations. Conducted experiments on the BANCA database show that our model is suitable for the face authentication task, since it outperforms not only Gaussian Mixture Models, but also classical appearance-based methods, such as Eigenfaces and Fisherfaces.

# 1 Introduction

Face recognition has been an active research area since three decades, and a huge variety of different systems are now capable to recognize people based on their face image, at least in so-called controlled conditions (good illumination, fixed pose). Existing algorithms are often divided into two categories, depending on the information they use to perform the classification: appearance-based methods (also called *holistic*) are typically using the whole face as input to the recognition system. On the other hand, feature-based methods are considering a set of local observations derived from a face image. Such observations may be geometric measurements (distance between the eyes, etc.), particular blocks of pixels, or local responses to a set of filters for instance.

Examples of appearance-based systems include the well-known Principal Component Analysis (PCA) [19], Linear Discriminant Analysis (LDA) [4] as well as Independent Component Analysis (ICA) [3] to name a few. These projection techniques are used to represent face images as a lower-dimensional vector, and the classification itself is actually performed by comparing these vectors according to a metric in the subspace domain (or using a more sophisticated classification technique, such as Multi-Layer Perceptron or Support Vector Machines). On the other hand, feature-based approaches are trying to derive a model of an individual's face based on local observations. Examples of such systems include the Elastic Bunch Graph Matching (EBMG) [20], recent systems using Local Binary Patterns (LBP) [1] [17], and also statistical generative models: Gaussian Mixture Models (GMM) [6] [13], Hidden Markov Models (HMM) [15] [18] or its variant [5] [14].

Face recognition systems using local features were empirically shown to perform better as compared to holistic methods [5] [12] [13]. Moreover, they also have several other advantages: first, face images are not required to be precisely aligned. This is an important property, since it increases robustness against imprecisely located faces, which is a desirable behaviour in real-world scenarios. Second, local features are also less sensitive to little variations in pose and in illumination conditions.

In this paper we will focus on statistical generative models and propose a new model, based on static Bayesian Networks, especially dedicated to the data we have to deal with, that is the human face. Actually, we think that classical statistical models (GMM and HMM), although successful, are not really appropriate to properly describe the set of local observations extracted from a face image. Indeed, GMM as applied in [5] are modelling the distribution of overlapping blocks among the whole face image, thus considering each block to be independent with respect to the others. Furthermore, it was shown in [13] that better results are obtained by modelling each part of the face using a different GMM. However, the model likelihood is computed as the product of the the GMM likelihood, hence again considering the different face parts independently.

Obviously, this is not the case due to the nature of the "face object". Consider the two eyes for instance: the block containing one eye is likely to be related somehow to the block containing the other eye. Going one step further, HMM-based approaches, as well as its variant (2D-HMM, coupled HMM) are able to add structure to the observations and therefore usually perform better. Examples of embbeded *dynamic* Bayesian Networks (which are nothing else but an extension of the HMM framework) applied to face recognition can be found in [14]. However, such systems cannot introduce *causal relationships* between observations themselves, they mainly act on their ordering. By using *static* Bayesian Networks, it is then possible to model causal relationships between a set of different observations represented by different variables. Hence, in this contribution we propose a first attempt, to our knowledge, to derive a statistical generative model based on this paradigm and especially dedicated to the particular nature of the human face. Conducted experiments on the BANCA database show a performance improvement over a GMM-based system making the independence assumption between different facial feautres.

The remaining of this paper is organized as follows. Section 2 describes the general framework to perform face authentication with statistical models. Then, Bayesian Networks are briefly introduced before presenting the proposed generative model used to represent a face. The BANCA database, the experimental framework and obtained results are discussed in Sec. 4. Finally, the conclusion is drawn in Sec. 5, and some future possible research directions are outlined.

# 2 Face Authentication using Generative Models

In the framework of face authentication, a client claims its identity and supports the claim by providing an image of its face to the system. There are then two different possibilities: either the client is claiming its real identity, in which case it is referred to as a *true client*, either the client is trying to fool the system, and is referred as an *impostor*. In this open-set scenario, subjects to be authenticated may or may not be present in the database. Therefore, the authentication system is required to give an opinion on whether the claimant is the true client or an impostor. Since modelling all possible impostors is obviously not feasible, a so-called *world-model* (or universal background model) [5] [10] is trained using data coming from different identities, and will be used to simulate impostors.

More formally, let us denote $\lambda_{\bar{C}}$ as the parameter set defining the world-model whereas $\lambda_C$ represents the client-specific parameters. Given a client claim and its face representation $X$, an opinion on the claim is given by the following log-likelihood ratio:

$$\Lambda(X) = \log P(X|\lambda_C) - \log P(X|\lambda_{\overline{C}}) \tag{1}$$

where $P(X|\lambda_C)$ is the likelihood of the claim coming from the true client and $P(X|\lambda_{\overline{C}})$ is an approximation of the likelihood of the claim coming from an impostor. Based on a threshold $\tau$, the claim is accepted if $\Lambda(X) \geq \tau$ and rejected otherwise.

In order to find the parameters $\lambda_{\overline{C}}$ of the world model, and since we are dealing with model containing unobserved (or hidden) variables, the well-known Expectation-Maximisation (EM) algorithm [9] in the Maximum Likelihood (ML) learning framework is used. However, when it comes to client parameter estimation, ML learning cannot be reliably used due to the small amount of available training data for each client, instead the Maximum A Posteriori (MAP) criterion is used [10] [5]. In this case, client-specific parameters are *adapted* from the world-model parameters (i.e. the prior) using client data in the following manner:

$$\lambda_C^{MAP} = \alpha \cdot \lambda_C^{ML} + (1 - \alpha) \cdot \lambda_{\overline{C}} \tag{2}$$

where $\lambda_C^{ML}$ denotes the client parameters obtained from a Maximum Likelihood estimation. The adaptation parameter $\alpha$ is used to weight the relative importance of the obtained ML statistics with respect to the prior.

# 3 Proposed Model

## 3.1 Bayesian Networks

In this section, we will briefly describe the framework used to build the statistical generative model to represent a face. Bayesian networks (also known as belief networks) provide an intuitive way to represent the joint probability distribution over a set of variables: random variables are represented as nodes in a directed acyclic graph, and links express *causality relationships* between these variables.

More precisely, defining $Pa(X_i)$ as the parents of the variable $X_i$, the joint probability encoded by such a network over the set of variables $\mathbf{X} = (X_1, ..., X_n)$ is given by the following chain rule:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{3}$$

Hence, a Bayesian Network is fully defined by the *structure* of the graph and by its *parameters*, which consists in the conditional probability distributions of each variable given its parents. Note however that a variable may have no parents, in which case its probability distribution is a prior distribution.

**Inference.**

The task of inference in Bayesian Networks consists in computing probabilities of interest, once evidence has been entered into the network (i.e. when one or more variables has been observed). In other words, entering evidence consists in either fixing the state of a discrete variable to one of its possible value or to assign a value in the case of a continuous variable. We are then interested in finding the effect this evidence has on the distribution of the others unobserved variables.

There are many different algorithm allowing to perform inference, the most renowned is certainly the belief propagation due to Pearl [16], which is a generalisation of the forward-backward procedure for HMM. However, it becomes problematic when applied to multiply-connected networks. Another more generic method is the Junction Tree algorithm [8], which allows to compute such posterior probabilities in any kind of networks and is also the most efficient algorithm to perform exact inference.

**Learning.**

Learning in Bayesian Networks refers either to structure learning, parameters learning or both [11]. In our case, we are considering networks of fixed structure. Hence, parameters are learned using the classical EM algorithm [9] with the either the ML or the MAP criterion described previously (Sec. 2).

## 3.2 Face Representation

Figure 1 depicts the proposed model to represent a face using salient facial features. Shaded nodes are representing visible observations (eyebrows, eyes, nose and mouth) derived from the face image, whereas white nodes are representing the hidden *causes* that generated these observations. This model can be understood as follows: a face is described by a set of unknown dependencies between eyebrows and eyes (node $BE$), eyes and nose (node $EN$) and nose and mouth (node $NM$). These combinations then generate a certain type of facial features (such as a small nose, or broad lips for instance) which are represented by the nodes at the second level. And finally, these types of facial features then generate the corresponding observations.

In this network, hidden nodes are discrete-valued and observed nodes are multivariate gaussians. The likelihood of the face representation defined by $X = (O_{lb}, O_{rb}, O_{le}, O_{re}, O_n, O_m)$ is obtained by first inferring the distribution of the hidden variables once observations has been entered in the network, and then by summing out over the state of the hidden variables. Note that our model introduce relationships between observations: if the node $O_{le}$ is observed, information about the node $O_{re}$ can be inferred through the node $E$ node for instance.

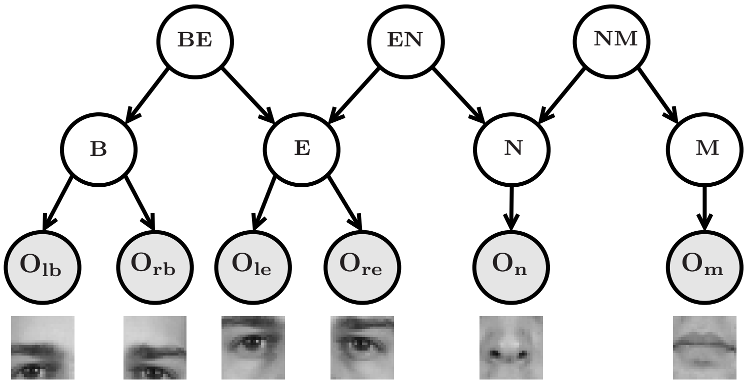
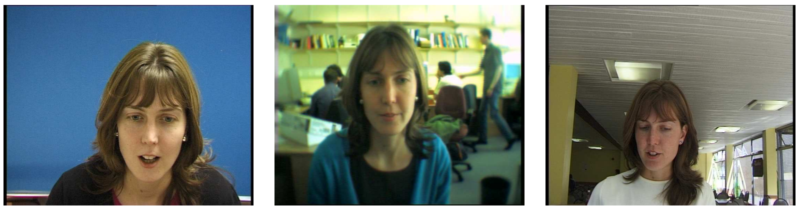



Figure 1: Static Bayesian Network for Face Representation

# 4 Experiments and Results

## 4.1 The BANCA Database

The BANCA database [2] was especially meant for multi-modal biometric authentication and contains 52 clients (English corpus), equally divided into two groups g1 and g2 used for development and evaluation respectively. Each corpus is extended with an additional set of 30 other subjects and is referred as the *world model*. Image acquisition was performed with two different cameras: a cheap analogue webcam, and a high-quality digital camera, under several realistic scenarios: controlled (high-quality camera, uniform background, controlled lighting), degraded (webcam, non-uniform background) and adverse (high-quality camera, arbitrary conditions). Figure 2 shows examples of the different acquisition scenarios.



(a) controlled (b) degraded (c) adverse

Figure 2: Example of the different scenarios in the BANCA database

In the BANCA protocol, seven distinct configurations for the training and testing policy have been defined. In our experiments, the configuration referred as Match Controlled (Mc) has been used. Basically, it consists in training the system with five images per client acquired during the first session. Then, the testing phase is performed with images acquired during the remaining sessions of the controlled scenario.

## 4.2 Experimental Framework

Each image was first converted to grayscale and processed by an Active Shape Model (ASM) in order to locate the facial features, defined by sixty-height landmark points [7]. Then, histogram equalization was applied on the whole image so as to enhance its contrast. Blocks centered on a subset of facial features were extracted (Fig. 1), and in order to increase the amount of training data, shifted versions were also considered. Hence, in our experiments we use the original extracted block as well as twenty-four other neighbouring blocks, resulting from extractions with shifts of two, three and four pixels in each directions. Each block is finally decomposed in terms of 2D Discrete Cosine Transform (DCT) in order to build the final observation vectors.

Face authentication is subject to two type of errors, either the true client is rejected (false rejection) or an impostor is accepted (false acceptance). In order to measure the performance of authentication systems, we use the Half Total Error Rate, which combines the False Rejection Rate and the False Acceptance Rate and is defined as:

$$HTER = \frac{(FAR + FRR)}{2} \quad [\%] \tag{4}$$

Hyperparameters, such as the threshold $\tau$, the dimension of the DCT feature vectors, the cardinality of the hidden nodes and the adaptation parameter $\alpha$ were selected using the validation set along a DET curve at the point corresponding to the Equal Error Rate (EER), where the false acceptance rate equals the false rejection rate. HTER performance is then obtained on the evaluation set with these selected hyperparameters.

### 4.3 Results

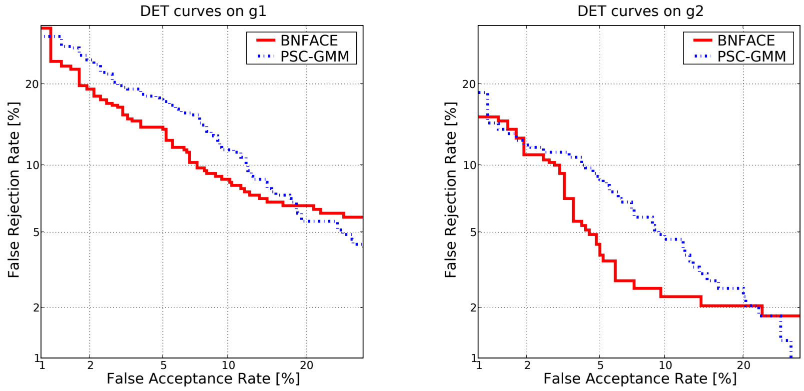
Here we present face authentication results obtained with the proposed model based on Bayesian Networks (BNFACE), and also with a baseline GMM system. Since the main assumption that drove us towards our approach was to state that blocks containing facial features should not be treated independently, we reproduced the experiment with the so-called *Partial Shape Collapse* GMM (PSC-GMM) first presented in [13]. However, and in order to yield a fair comparison, we use exactly the same set of features produced for our model, and hence did not take into account the nose bridge and both cheek regions used in [13]. In our experiments, DCT feature vectors of dimensions 64 were used, and the cardinality of the discrete variables was set to 3 at the first level and to 8 at the second level. Regarding the PSC-GMM model, we used 512 gaussians for each of the six GMM corresponding to the six extracted facial features, as suggested in [13].

In Tab. 1, we report the results obtained by our approach (BNFACE), by our implementation of the PSC-GMM and by the GMM approach as published in [5]. Note that in [5] only the results on g2 are available. The proposed BNFACE model outperforms the corresponding PSC-GMM approach on both sets of the BANCA database. Moreover, obtained results on the test set g2 are better than those obtained with a single GMM [5]. This comparison is interesting since this GMM-based system uses much more features extracted from the whole face image. Note also that our model contains the less client-specific parameters to be learned.

| FA system | HTER on g1 [%] | HTER on g2 [%] | number of parameters |
|---|---|---|---|
| BNFACE | 9.01 | 5.41 | 5225 |
| PSC-GMM | 11.31 | 11.34 | $6 \cdot 33280$ |
| GMM [5] | not available | 8.9 | 9216 |

Table 1: HTER Performance on the Mc protocol of the BANCA database.

Since results presented in [13] are reported in terms of EER on a graph, we also compare EER performance on both development and test sets in Tab. 2. Note however that the numeric results from [13] are estimated from the graph, and are thus subject to little imprecisions. Once again, we noticed that the proposed approach performs better than using the PSC-GMM with the same features, as can also be seen on DET curves (Fig. 3). However, results of the original PSC-GMM are better. It can




(a) (b)

Figure 3: DET curves obtained on the Mc protocol of the BANCA database for the BNFACE (solid line) and the PSC-GMM (dashed) models.

be explained by the fact that it uses more features than our model to perform the face authentication task. Note also that our model provide better performance than classical appareance-based models such as Eigenfaces and Fisherfaces as provided in [13].

| FA system | EER on g1 [%] | EER on g2 [%] |
|---|---|---|
| BNFACE | 9.01 | 4.84 |
| PSC-GMM | 11.31 | 6.92 |
| PSC-GMM [13] | 3.9 | 4.1 |
| Fisherfaces [13] | 10.2 | 11.5 |
| Eigenfaces [13] | 13.8 | 14.0 |

Table 2: EER Performance on the Mc protocol of the BANCA database.

## 5 Conclusion and Future Directions

In this paper, we proposed a new statistical generative model to represent human faces, and applied it to the face authentication task. The main novelty of our approach consists in introducing dependencies between observations derived from salient facial features. As shown by the conducted experiments on a benchmark database, our main hypothesis seems to be verified, since our model performs better than systems relying on the independence assumptions between facial features. Moreover, obtained results also compares favourably against classical holistic methods such as Eigenfaces and Fisherfaces.

However, this work is a preliminary attempt to use static bayesian networks in face processing and many issues are still open. Indeed, future research directions are manifold. First, causal relationships between facial features are not known (at least to our knowledge) and finding the right structure for the network is not straightfoward. Second, it will be interesting to use other facial features possibly carrying more discriminative information, such as skin texture for instance, and incorporate it into a network.

## 6 Acknowledgments

This work has been performed in the framework of the GMFace project of the Swiss National Science Foundation (SNSF). Softwares were implemented using the TorchVision library (http://torch3vision.idiap.ch/) and experiments were carried out using the PyVerif framework (http://pyverif.idiap.ch/).

## References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face Recognition With Local Binary Patterns. In *European Conference on Computer Vision (ECCV)*, pages 469–481. Springer, 2004.

[2] E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J-P. Thiran. The Banca Database and Evaluation Protocol. In *4th Intl. Conf. Audio- and Video-based Biometric Person Authentication, AVBPA'03*. Springer, 2003.

[3] M. Bartlett, J. Movellan, and T. Sejnowski. Face Recognition by Independent Component Analysis. *IEEE Trans. on Neural Networks*, 13(6):1450–1464, 2002.

[4] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[5] F. Cardinaux, C. Sanderson, and S. Bengio. User Authentication via Adapted Statistical Models of Face Images. *IEEE Trans. on Signal Processing*, 54(1):361–373, 2005.

[6] F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of MLP and GMM classifiers for face verification on XM2VTS. In *4th Intl. Conf. Audio- and Video-based Biometric Person Authentication, AVBPA'03*. Springer, 2003.

[7] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active Shape Models: Their Training and Applications. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[8] G. Cowell, P. Dawid, L. Lauritzen, and J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer Verlag, 1999.

[9] A. Dempster, N. Laird, and D. Rubin. Maximum Likelihood From Incomplete Data via the EM Algorithm. *The Journal of Royal Statistical Society*, 39:1–37, 1977.

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

[11] D. Heckerman. *Learning in Graphical Models*, chapter A Tutorial on Learning With Bayesian Networks, pages 301–354. MIT Press, 1999.

[12] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face Recognition: Component-based versus Global Approaches. *Computer Vision and Image Understanding*, 91(1):6–21, 2003.

[13] S. Lucey and T. Chen. A GMM Parts Based Face Representation for Improved Verification through Relevance Adaptation. In *IEEE Intl. Conf on Computer Vision and Pattern Recognition (CVPR)*, pages 855–861, 2004.

[14] A. Nefian. Embedded Bayesian Networks for Face Recognition. In *IEEE Intl. Conference on Multimedia and Expo (ICME)*, August 2002.

[15] A. Nefian and M. Hayes. Hidden Markov Models for Face Recognition. In *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 2721–2724, 1998.

[16] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[17] Y. Rodriguez and S. Marcel. Face Authentication Using Adapted Local Binary Pattern Histograms. In *European Conference on Computer Vision (ECCV)*, pages 321–332, 2006.

[18] F. Samaria and S. Young. HMM-based Architecture for Face Identification. *Image and Vision Computing*, 12(8):537–543, October 1994.

[19] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. In *IEEE Intl. Conf on Computer Vision and Pattern Recognition (CVPR)*, pages 586–591, 1991.

[20] L. Wiskott, J-M. Fellous, N. Krüger, and C. Von Der Malsburg. Face Recognition By Elastic Bunch Graph Matching. In *7th Intl. Conf. on Computer Analysis of Images and Patterns (CAIP)*, pages 456–463, 1997.