Bayesian Networks as Generative Models for Face Recognition

IDIAP RESEARCH INSTITUTE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Guillaume Heusch

supervised by:

Dr. S. Marcel
Prof. H. Bourlard

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

2009

Acknowledgements

Since I'm engaged in a PhD thesis, I have been, from time to time, thinking about this part, trying to mentally list all the people that were with me along the way... So, apologies if I forgot someone: I assume that if you read this, you were certainly in my list at some point. First of all, I would like to thank Prof. Hervé Bourlard, my thesis director, and the member of my jury: Prof. Stan Li, Prof. Massimo Tistarelli, Prof. Jean-Philippe Thiran and the president Prof. Anja Skrivervik. This work would not have been possible without the ideas, the help and the guidance of Dr. Sébastien Marcel, my day-to-day supervisor. I would also thank him for having introduced me to the fascinating world of artificial intelligence quite a long time ago now, it was a real pleasure to work with him during these years, with a good balance of valuable inputs and also the freedom he gave me to pursue my own ideas and research. I'm also grateful to colleagues with whom I had discussions at the beginning of this work: they certainly help paved the way.

Many thanks goes to the people at the Idiap Research Institute: Nadine & Sylvie for their constant availability whenever administrative duties showed up, all the system guys for providing everything you need to build a nice work environment and also for their efficiency to solve technical problems. I also thank the colleagues for interesting discussions and the closer friends there, for all the fun (and the fondues) we had outside of the working place (you know who you are ...).

I'm grateful to the friends who accompanied me along this long way, it was always nice to spend time with you when I needed a break from research, while feeling your support at the same time. And I would also like to thank you for preventing me to be too much in my work, and keeping me aware of the surrounding world. There are way too many names to list here, but I still would like to mention Alex, Thomas and Christian, who were always there when I needed to escape a little bit. Of course, I cannot forget all the guys playing in the LUC Rugby team: it was (and it's still is) a real pleasure being by your side during all these years. Many thanks for providing me with the nicest way to fill my brain with oxygen, and all the special moments we shared together.

Of course, such work would have been much more difficult and more painful without the support and the presence of special persons. This include the entire Charrière family, who always makes me feel proud of what I was achieving. I could certainly not thank my family as much as they deserve for their everlasting support, help and love. Big big thanks goes to my aunt and my uncle, my grandfather for his wisdom, my three little sisters for their constant smiles, my older sister and her boyfriend (especially during this past year). Finally, this would have been impossible without my parents, who always backed me up: thanks for their unfailing support and unconditional love ... You actually made this !

Abstract

In this thesis, we address the problem of face modelling by using dedicated statistical generative models, with an application to the face authentication task. Face authentication consists in either accepting or rejecting a user's claim supported by its face image. Classical generative models, such as Gaussian Mixture Models (GMM), Hidden Markov Models (HMM) and their variants have been proved to be successful to tackle this problem. However, these models are not appropriate to the structure of the observed data. In particular, these models implicitly assume independence between features extracted from the image, which is obviously not true in the case of a face.

We thus propose new generative models, based on Bayesian Networks, and especially tailored to deal with the object we have to handle: the face. Actually, we would like to exploit as much as a priori knowledge as possible. For this purpose, Bayesian Networks provide an intuitive framework: they allow to encode causal relationships between different kind of random variables, thus enabling to express correlations between different source of information. As a first step, we thus propose a model acting on local observations extracted around salient facial features, which is designed to capture relationships among these pieces of information. The proposed model is shown to be competitive with state-of-the-art approaches based on generative models when applied to the authentication task. At the same level of performance, it is also less complex and thus less time consuming than previous approaches. Besides, and as opposed to classical models, meaningful information could be retrieved from the proposed model.

We then extend this model using other sources of information as complementary clues to local features extracted from grayscale face images. Indeed, cognitive studies in face recognition showed that human beings are using various information such as shape, low-resolution representation of the face and also skin color to recognize an individual. We thus proposed new models taking these information into account, and apply them to our authentication problem.

Key Words: Face Recognition, Statistical Models, Bayesian Networks.

VERSION ABRÉGÉE

Dans cette thèse, nous étudions le problème de l'authentification biométrique basée sur le visage à l'aide de modèles génératifs dédiés. L'authentification de visages consiste à accepter ou rejeter le postulat d'identité d'un utilisateur en se basant sur l'image de son visage. Bien que les modèles classiques, tels que les Mixtures de Gaussiennes (GMM) ou les modèles de Markov Cachés (HMM) résolvent cette tâche avec succès, ils ne sont pas appropriés à la nature des données observées. En particulier, ces modèles posent de manière implicite une hypothèse d'indépendence entre les différentes caractéristiques extraites de l'image, ce qui ne se vérifie pas forcément dans le cas d'un visage.

Nous proposons donc de nouveaux modèles, basés sur les Réseaux Bayésiens, et spécialement conçus pour traiter l'objet en question: le visage. Nous désirons exploiter le plus de connaissances possibles a priori. Dans cette optique, les Réseaux Bayésiens fournissent une approche intuitive: en effet, ils permettent d'encoder des relations causales dans un ensemble de variables aléatoires, pouvant ainsi exprimer des corrélations entre différentes informations. Dans un premier temps, nous proposons une approche traitant les différentes parties d'un visage et modélisant leurs correlations. Le modèle proposé est compétitif en comparaison des modèles génératifs existants appliqués à l'authentification de visage. Au même niveau de performance, il est également moins complexe et de ce fait plus rapide que les approches précédentes. De plus, et au contraire des modèles classiques, le modèle proposé peut aussi être utilisé pour visualiser différentes informations sur les visages.

Par la suite, ce modèle est augmenté en utilisant d'autres informations en complément des parties du visage extraites d'une image en niveau de gris. En effet, des recherches cognitives ont montré que les êtres humains utilisent par exemple la forme globale du visage, ses caracteristiques géométriques ou encore la couleur pour reconnaitre un individu. Nous proposons donc de nouveaux modèles intégrant ces informations et les appliquons à l'authentification de visage.

Mots Clés: Reconnaissance de visages, Modèles statistiques, Réseaux Bayésiens.

Contents

List of Figures

List of Tables

Chapter 1

OVERVIEW

1.1 Aims & Motivation

In this thesis, the problem of face authentication is addressed. Face authentication consists in either accepting or rejecting the identity claim of an individual supported by its face image, hence it falls into the more generic problem of Face Recognition. Although current algorithms are quite successful on controlled conditions, performance decreases rapidly in case of unconstrained viewing conditions, such as head pose and illumination for instance [Zhao 03]. Authentication systems based on local features and statistical models were shown to perform well for this task [Sanderson 03] [Lucey 04] [Cardinaux 05]. However, currently used models, including Gaussian Mixture Models (GMM), Hidden Markov Models (HMM) or their variants, such as embedded HMM [Nefian 00] or coupled HMM [Nefian 02] are not appropriate to the structure of the observed data. Furthermore, these models implicitly assume independence between the observed local features, which is obviously not true in the case of the face image.

Several computer vision tasks could be solved using the more generic framework of probabilistic graphical models. Graphical models comes in two flavors: directed (Bayesian Networks) and undirected (Markov Random Fields). Such models have already been used for various face processing related task. An example of Markov Random Field applied to face recognition can be found in [Huang 04]. Regarding Bayesian Networks, they have been applied in the context of face detection [Yow 97] or facial expression recognition [Cohen 02], but were never applied to face recognition.

In this thesis, we thus intend to investigate the usage of the generic Bayesian Network framework [Pearl 88] as statistical generative models for face authentication. Indeed, Bayesian Networks (also known as Belief Networks or Probabilistic Experts Systems) 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. In this framework, it is then possible to model correlations between the different observations derived from a face image, and hence to relax the independence assumptions when modelling a face with GMMs or HMMs. Besides, and within this framework, it is particularly easy to introduce prior knowledge on the problem at hand: one can freely choose which variables to consider, and how to relate them to each other through the specified graph. In our case, we will consider that certain observations are generated by the same underlying process. Imagine for instance that several local features are extracted from the same face image: these features should be somehow related to each other, because they have been generated by the same identity. Such a property cannot be expressed using classical statistical models, but Bayesian Networks are able to model such relationships through the network structure. However, finding a set of pertinent observations together with an appropriate corresponding network structure is an open research issue.

Another property of Bayesian Networks is their ability to handle different kinds of random variables. Hence, we intend to add auxiliary information to grayscale image patches in the generative model. It is done to improve its modelling power in face description. As evidenced by psychological and cognitive studies, where key results can be found in [Sinha 06], human beings use several and different visual clues to recognize a face. Examples of such information include the global face shape and the skin color for instance. Although it may seem obvious that adding such information may help the recognition process, most of the current algorithms usually act on part of such information, or treat them independently. Moreover, they usually do not exploit the prior knowledge we may have on the object to recognize, but rather apply standard pattern recognition algorithms. We believe that Bayesian Networks is a well suited tool to design generative models mainly for two reasons. First, explicit relationships between facial features can be encoded through the network structure, hence relaxing independence assumptions. Second, we believe that incorporating auxiliary information in a principled way could be beneficial to describe faces more accurately.

1.2 Contributions

The aim of this thesis is thus to design, implement and apply new statistical models, based on Bayesian Networks, to tackle the face authentication task. Actually, we would like to design models that are able to meaningfully describe the object to process, unlike classical statistical models used so far.

The main contribution of this thesis is hence the design and the development of novel generative models dedicated to the face. Inspired by previous work on face recognition based on salient facial features, by the usage of statistical models in the context of face authentication and by results obtained in cognitive research on human face recognition, we propose new models aiming at encoding as much of prior knowledge as possible on the object to process. More precisely:

- The first proposed generative model explicitly correlates observations extracted around salient facial features, as opposed to traditionally used statistical models, such as GMMs and HMMs.
- This model is then extended by incorporating various other, auxiliary information in a principled way. In particular, the usage of holistic representations of the face, geometrical measurements and also color information are investigated. Several network architectures are proposed to incorporate these pieces of information.
- For this purpose, the development of an efficient implementation of the inference (Junction Tree) and learning (Expectation Maximization) algorithms was performed. These algorithms may be applied in directed Bayesian Networks with both discrete and continuous hidden variables. This development has been made within a well-known machine learning library[1].

Moreover, a systematic and unbiased experimental comparison with existing generative models (GMM, HMM and Pseudo 2D-HMM), but also with popular appearance-based face authentication systems is provided. This is done using well-defined protocols on databases comprising several and challenging authentication scenarios. Experiments are conducted using both manual and automatic face detection, since robustness to imperfectly located faces is a desirable property in real world applications. The proposed models are shown to achieve comparable results to state-of-the-art generative approaches, while being less complex, less computationally demanding, and exhibiting meaningful properties of the different faces.

[1] www.torch.ch

1.3 Organisation of the Thesis

The following document is organized as follows:

- **Chapter 1** introduces the problem and briefly presents the limitations of current systems based on local features and generative models. It then states the motivation and contribution of this work.
- **Chapter 2** provides a brief review of the state-of-the-art in face recognition. The face authentication problem is presented, together with corresponding databases and performance measures. An emphasis is put on face authentication using generative models: a discussion on the advantages and drawbacks of existing approaches is made, and the motivation for new models is then explained in more details.
- **Chapter 3** gives the necessary theoretical background to effectively design new generative model introducing prior knowledge through correlations between observations. Bayesian Networks and their key properties are introduced, before detailing two algorithms related to our problem and used in our work: inference and parameters learning.
- **Chapter 4** presents a simple yet efficient generative model for the face. It empirically shows that the model is actually able to learn different facial feature types and their relationships. Experiments are provided on the face authentication task and a systematic comparison with different approaches is made, using both manual and automatic registration of the face.
- **Chapter 5** extends the first approach and presents new models incorporating auxiliary information. In particular, models embedding information used by human beings are investigated. We developed new models using holistic representation of the face, configural information and color, and applied them to the face authentication task.
- **Chapter 6** finally concludes this thesis by providing a summary of the performed work and obtained results. Possible directions for future research are also suggested.

Chapter 2

FACE RECOGNITION

Face recognition falls into the more generic topic of biometric recognition, which consists in recovering the identity of people based on their intrinsic physical properties. Biometrics have a certain advantage over classical identification methods, since they do not require any information that people usually detain, such as passports, driver's licence or PIN codes for instance. Indeed, such means of identification could be easily forgotten, stolen or even forged. Rather, biometric recognition is based on the individual's own characteristics, such as the fingerprint, the iris, the gait and, of course, the face.

As compared to other biometric modalities, the face has the main advantage of being non-invasive for the user, and it could also work at a distance and without explicit user cooperation (i.e. there is no need to swipe your finger on a specific device for instance). On the other hand, biometric recognition based on faces remains a challenging problem since it has to deal with a wide range of challenges, such as pose and illumination [Zhao 03]. Furthermore, long-term instability of such a biometric sample is also an issue. Face appearance may greatly vary over time due to external artifacts (glasses, beard) and aging for example. Nevertheless and in comparison to other modalities, the face as a biometric trait seems to be a good tradeoff between its ease of use, the little need for user cooperation and the required accuracy. Face recognition technology could be applied in an enormous range of applications : The first example which comes to mind is certainly security: border control, law enforcement, building access are example where face recognition is already deployed. But it is also worth mentioning other applications who could benefit from this technology, such as human-computer interaction, automatic photo/video indexing or image retrieval to name a few.

Basically, any face recognition system consists in comparing previously known face images (often referred to as the *gallery*) to other provided face images (termed as *probes*). But prior to the recognition task itself, there are necessary steps that need to be addressed: the first one is face detection, whose goal is to detect if a face is present or not in the image given as input to the system. Once the face has been detected, it has to be registered (i.e. salient facial features have to be located) and is usually geometrically normalized. The recognition task could then take place, and may also be broken down into three distinct tasks: the optional preprocessing of the registered face image, the feature extraction step, and the final classification, which depends on the extracted features.

Face recognition could be operated in three different modes : identification, authentication and watch-list. Although the same algorithms could be applied to all of these modes, they all have specific properties, and special care on the application should be made when designing a particular system. The main difference between these applications is whether they operate on so-called *closed-set* or *open-set* scenarios.

Identification The task of identification consists in retrieving, based on a face image, the identity of an individual out of a pool of people previously enrolled into the system. Since in this case all identities are known a priori, and that every people to be identified are present in the database, this problem is termed as closed-set.

Authentication Face authentication (also known as face verification) is addressing the following scenario: an individual claims an identity and support his/her claim by providing its face image. In that case, the system has to decide on whether the claim should be accepted or rejected. Authentication has hence to deal with *impostors*, who are individuals trying to fool the system. Since impostors are not present in the database, authentication is therefore referred to as an open-set scenario.

Watch-list The watch-list problem consists in performing identification on an open-set scenario: the individual to be identified may or may not be previously known to the system. In this case, special care has to be made on the decision: the best match between the gallery and the probe does not necessarily correspond to the right identity.

Face recognition is inherently a three-dimensional pattern recognition problem, due to the nature of the observations. The main advantage for 3D recognition is its ability to cope with variations in lighting conditions and pose. However, deploying a 3D-based system is quite complicated and special capturing devices (i.e. laser-scan) are needed. 3D face recognition is usually used when the security level is of great concern. On the other hand, acquiring two-dimensional face images is a rather straightforward task. Indeed, there is a huge amount of digital cameras, webcams and mobile phones available and widely used nowadays. Although 2D face images are more convenient to capture in everyday life, retrieving identities from such input is much more difficult. Nevertheless, picture-based face recognition remains the primary choice for many applications, mainly because the tradeoff between its simplicity (as compared to 3D-based systems) and its accuracy is good enough in most cases. Hence, in this thesis, the focus is made on the face authentication task based on two-dimensional face images. In the next section, we thus present a brief review of some of the most popular algorithms developed so far to address this task.

2.1 State-of-the-art

For more than 30 years, face recognition has been an active research area and different systems are now capable of correctly recognizing people's faces under controlled environment (i.e. when the probe and the gallery images are acquired under the same global viewing conditions) [Zhao 03]. There is a huge amount of face recognition algorithms, and developed approaches can be subdivided into appearance-based methods (also called *holistic*) and feature-based methods (also called *local*). The major difference between these methods is that holistic approaches use the whole face region as a raw input to a classifier, whereas in local feature-based methods the face is considered as a set of observations, which will be used for subsequent analysis. The next subsections are devoted to present these approaches and their usage as face recognition algorithms, before discussing their respective advantages and drawbacks.

2.1.1 Appearance-based algorithms

In this section, we will briefly list some of the appearance-based algorithms developed for 2D face recognition. An emphasis will be made on two particular algorithms: Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), since they are used in this thesis for comparison purposes.

Principal Component Analysis One of the first face recognition systems is the Eigenface method developed by Turk & Pentland [Turk 91]. Principal Component Analysis (PCA) is applied to the face images in the database, and each face image is then projected onto a lower-dimensional subspace. A face image is thus described by a vector of weights, which represents the linear combination of the so-called *eigenfaces* (face images composing the orthonormal basis of the subspace). Considering the holistic face representation as a vector denoted by x, it is linearly projected to obtain the feature vector y:

$$y = W_{pca}^T x \tag{2.1}$$

where W_{pca}^T is a matrix containing the first m eigenvectors (corresponding to the m largest eigenvalues) of the training data covariance matrix. Classification is then performed directly in this subspace, using either a simple metric, such as the Euclidean distance [Turk 91], or more sophisticated techniques like Multi-Layer Perceptron (MLP) [Marcel 02] or Support Vector Machines (SVM) [Jonsson 00] for instance.

Linear Discriminant Analysis Whereas PCA projects face images into the maximum variance subspace, Linear Discriminant Analysis (LDA) seeks to find the projection that optimally separates the different classes. Unlike PCA, LDA is a supervised system and was first applied in face recognition in [Belhumeur 97]. It has also been successfully applied to the problem of face authentication [Kittler 00] [Marcel 04]. The LDA algorithm is based on Fisher's linear discriminant formula to find the optimal discriminative subspace where data will be projected. Let us consider a training set composed of N examples x_i where each example is labeled and belongs to one of C classes. The training data are used to derive the between-class scatter matrix S_b and the within-class scatter matrix S_w:

$$S_b = \sum_{i=1}^{C} N_i\,(\mu_i - \mu)(\mu_i - \mu)^T \tag{2.2}$$

$$S_w = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i^T) \tag{2.3}$$

where N_i are the number of examples in class i, μ_i is the mean of the examples in class i and μ is the mean of all training data. Finally, the optimal projection W_{lda} is the orthogonal matrix satisfying

Equation (2.4):

$$W_{lda} = \underset{W}{\operatorname{argmax}} \frac{|W^T S_b W|}{|W^T S_w W|} \tag{2.4}$$

In practice, Principal Component Analysis (PCA) is first applied on the data so as to achieve decorrelation and dimensionality reduction. The projected face representations into the coordinate system of eigenvectors are then used as training examples to derive the optimal projection W_{lda}. The face vector x is then projected to get the final feature vector y according to Equation 2.5:

$$y = W_{lda}^T W_{pca}^T x \tag{2.5}$$

Like in the PCA case, classification is then performed on the feature vector y using either a metric, or more sophisticated approaches.

Other Popular Algorithms In this paragraph, we simply lists some of the existing algorithms based on the idea of subspace projection. For instance, another successful approach is to use the Independent Component Analysis (ICA) [Bartlett 02], which is nothing else but a generalization of PCA, which decorrelates the high-order moments in addition to the second order moment. It is also worth to mention the kernelised version of PCA and LDA [Yang 00] [Kim 02] [Yang 02] [Shen 07], which first project face data into a higher-dimensional space before computing projection matrices. There also exists non-linear projections, such as the so-called Laplacianfaces [He 05], where faces are projected onto a manifold which preserves the local structure of the data by using Locality Preserving Projections. Another non-linear projection-based approach is the Non-negative Matrix Factorization (NMF), which has also been applied to face recognition [Rajapakse 04] [Zhang D. 05].

It is also worth to mention some other holistic approaches, that are not based on a subspace projection paradigm. Such algorithms include Neural Networks [Lawrence 06] [Cardinaux 03], Support Vector Machines [Phillips 99] [Jonsson 00] [Gates 05] but also Active Appearance Models [Edwards 98].

2.1.2 Feature-based algorithms

Feature-based algorithms, as opposed to appearance-based approaches, are usually considering local features instead of the whole face image. In this section, we list some of the most prominent approaches belonging to this family. In this case, some prior knowledge is generally introduced in the system. Indeed, these algorithms are usually dedicated to handle faces. Actually, early face recognition algorithms were purely feature-based, since they tried to recognize people based on geometrical properties of the face [Kanade 73] [Nixon 85].

Local PCA and LDA In this case, the face is divided into rectangular regions located around salient facial features. The algorithm then acts directly in each region. The projection subspace is built for each region using either PCA [Pentland 94] or LDA [Short 06]. Recognition is first performed independently on each component, and results are combined to reach a final decision. Usually, recognition results are better using local information than the ones obtained by their holistic counterpart.

Elastic Bunch Graph Matching The Elastic Bunch Graph Matching (EBGM) [Wiskott 99] is certainly the most renowned local feature-based face recognition algorithm. In this algorithm, a face image is represented by a set of wavelets coefficients, called jets, arranged in a graph, whose nodes corresponds to fiducial points (eyes, tip of the nose, corner of the mouth, etc.). During the recognition process, the lattice is allowed to be deformable so as to maximize the correlation between corresponding wavelet coefficients of the gallery and of the probe image. The final decision is made based on a distance between jets extracted from both the gallery and the probe images.

Local Binary Patterns Recently, Ahonen et al. [Ahonen 04] proposed to apply to face images a simple operator, called Local Binary Patterns (LBP) and originally meant for texture analysis. Here the face is represented by a set of concatenated LBP histograms, each one being computed in a different block of pixels along the image. Recognition is made by comparing χ^2 distances between histograms. This approach yield good results when applied to face recognition, and a lot of extension have been proposed since then: see [Huang 05], [Rodriguez 06b], [Liao 07] or [Ahonen 09] for examples.

Statistical Models Another popular approach to local feature-based face recognition are the usage of statistical generative models [Samaria 94b] [Nefian 98] [Eickeler 00] [Sanderson 03] [Lucey 04]

[Cardinaux 05]. The aim of such systems is to model the distribution of a feature set extracted from the face image. Since the work presented in this thesis is based on this framework, it will be explained in more details in the next section.

2.2 Face Authentication using Generative Models

In the framework of face authentication, a user 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 to 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* is used to simulate impostors. It is trained using data coming from different identities and thus represents the model for an "average", or general individual [Reynolds 00].

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

$$\Lambda(\mathbf{x}) = \log p(\mathbf{x}|\theta_{client}) - \log p(\mathbf{x}|\theta_{world}) \tag{2.6}$$

where $p(\mathbf{x}|\theta_{client})$ is the likelihood of the claim coming from the true client and $p(\mathbf{x}|\theta_{world})$ is an approximation of the likelihood of the claim coming from an impostor. Based on a threshold τ, the claim is accepted if $\Lambda(\mathbf{x}) \geq \tau$ and rejected otherwise.

The probability $p(\mathbf{x}|\theta)$ can be computed using various statistical models. The first work using local features together with a generative model for face representation is due to Samaria & Young [Samaria 94b]. In their work, the authors are considering blocks of pixels as Gaussian observations, and a sequence of such observations is obtained by scanning line strips from the top to the bottom of the face image. An HMM then allows to compute the likelihood of a sequence, which consists in the probability that the HMM has generated the observed sequence. Hence, one HMM is built for each identity and the likelihood of the probe image is computed for every stored model. The recognized individual is then the one whose model gives the highest data likelihood. Later, Nefian [Nefian 98] improves upon Samaria's work by using Gaussian Mixtures as a distribution for the observations, but also by reducing the dimensionality of the observation vector by applying a two-dimensional Discrete Cosine Transform (2D-DCT) on the pixels intensity values. The same approach was used in [Martinez 99] and showed good robustness against variations present in face images, such as expressions or occlusions. However, a major drawback of these HMM approaches is that the observations are related to each other in only one dimension (as opposed to the 2D nature of the images). Kohir et al. also applied an HMM-based system for face recognition [Kohir 98], but in this case, the sequence is obtained by scanning the face image in a zig-zag fashion, and the HMM is ergodic.

To be in accordance with the 2D nature of face images, an embedded HMM architecture was first proposed in [Nefian 00]: here, the emission probability of a main, vertical HMM is itself modelled by an horizontal HMM. An image is hence analyzed on a block-by-block basis rather than line-by-line. This system was shown to outperform not only PCA, but also neural networks based algorithms on a rather easy database. Other experiments using the pseudo-2D-HMM were reported in [Eickeler 00]: here the model of each client takes a so-called common initial model as starting point for the training procedure, and learning is thus more efficient even with few training samples. Recently, Maximum A Posteriori (MAP) adaptation of statistical models [Gauvain 94] has been applied with success to face authentication [Cardinaux 03] [Lucey 04] [Cardinaux 05]. In this case, each client model is adapted from the world model during training, using available client data. Note also that a true low complexity 2D-HMM has been proposed in [Othman 03], but this model needs a large amount of training data for its parameters to be properly estimated.

2.3 Discussion

Holistic approaches are today's best performers for the face authentication task: they have some advantages, such as being able to handle low-resolution face images, but they also have some drawbacks: a large amount of data is needed to compute the projection matrix for instance. Another limitation of such methods is their poor generalization ability: variations such as pose or illumination could be accounted for only if present in the training set [Martinez 02] (i.e the one used to build the projection matrix), but usually, a specific subspace is built for each possible condition: Georghiades et al. [Georghiades 01] construct low-dimensional face models taking illumination variations into account for each possible pose for instance. Furthermore, holistic-based systems do not exploit any prior knowledge on the object to recognize, since image data is usually represented as a single feature vector, and no specific hypothesis are assumed regarding the content of the image: face recognition is thus treated as a standard pattern recognition problem.

On the other hand, feature-based methods seem to better handle little variations in appearance. Using local information helps when the face is not frontal [Heisele 03] [Lucey 06], but also if an occlusion or a change in expression occurs [Martinez 02]. Moreover, it has been shown [Cardinaux 05] [Rodriguez 06a] that systems based on local features are more robust to imperfect face localization. Such approaches are also able to encode more prior knowledge as compared to holistic approaches: one can freely choose which features to use and how to relate them to each other. However, local feature methods also have some drawbacks: they usually need higher resolution images than holistic approaches, and in case of statistical models, learning parameters form data with few samples per class is an important issue.

2.4 Databases, Protocols and Performance Measures

Performance evaluation of face recognition algorithms is an important step and should be carried with well-defined protocols on the database of choice. There exists a wide variety of face databases, which usually have different purposes. Indeed, a face database is generally designed and built to test face recognition systems in particular, and sometimes difficult, conditions. For instance, the AR database [Martinez 98] is especially meant to evaluate the robustness of algorithms against variations in illumination conditions, facial expressions and occlusions. Other well-known examples are the CMU-PIE database [Sim 01], where pose variation is also present or the Yale B Face Database [Georghiades 01], which contains extreme illumination conditions. However in these cases, there are no strict protocols associated to the databases, preventing from a fair comparison between face recognition systems.

To reach the goal of a standardized assessment procedure, two databases and their associated detailed protocols, are provided by the National Institute of Standard in Technology (NIST) : the FERET [Phillips 00] and the FRGC [Phillips 05] databases. These two databases are meant to be used in a closed-set face identification setting with different scenarios. For example, some of the protocols propose to evaluate the behavior of face identification systems in challenging conditions: the performance over time (i.e. when the probe and gallery image are taken one year apart) and in strongly mismatched conditions: in this case, the gallery image is usually taken in indoor controlled conditions whereas the probe image is acquired outdoor under arbitrary pose and illumination.

In our case however, we focus on the authentication scenario and thus present in more details two databases especially meant to address this particular problem. Evaluation procedures and performance measures are also introduced.

2.4.1 The XM2VTS Database

The XM2VTS database [Messer 99] was especially meant for multimodal (voice and face) identity verification, where the user (i.e. the client) is assumed to be cooperative with the system. The data corpus consists in synchronized video and speech data from 295 subjects, acquired in 4 sessions of 2 shots, and over a 5 month period. All shots were captured under controlled conditions: the camera settings, the illumination and the background remained the same across all the sessions. Figure 2.1 shows some examples of face images in this database.

There are two well-defined experimental protocols provided with the database. First, the database was randomly divided into 3 disjoints set : the client set, which contains 200 subjects, and two other impostors set, which contains 25 and 70 subjects for evaluation and test respectively. The Lausanne Protocol (LP) then define which images should be used for the client models, and is used in two different configurations. Figure 2.2 shows the partioning in both cases. Note that the difference between the two protocols resides only in the client training and evaluation sets: in the first configuration, only 3 images per subject are available, whereas the second configuration allows 4 images to be used.

Fig. 2.1: Some examples of subjects present in the XM2VTS database. Images in each column are taken from a different session.

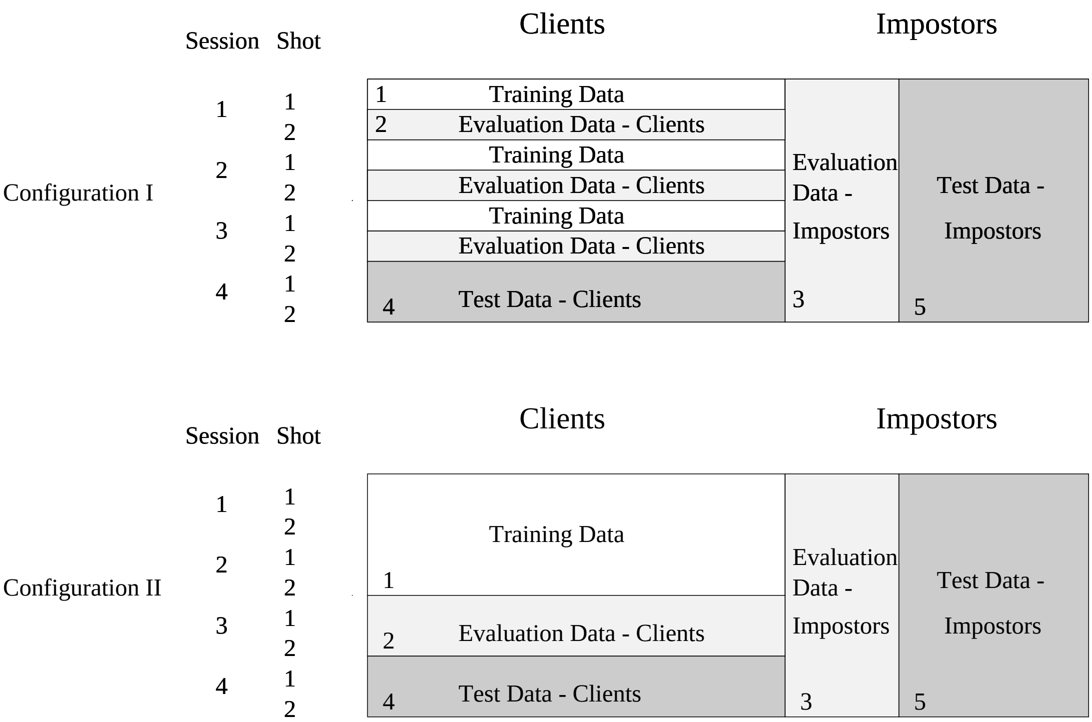


Fig. 2.2: The different protocols associated to the XM2VTS database

2.4.2 The BANCA Database

The BANCA database [Bailly-Baillière 03] was also designed to perform multimodal identity verification. There exists two main differences with the XM2VTS database though. First, the corpus is smaller and contains only 52 clients and second, the acquisition conditions are not the same across the twelve different sessions. Sessions 1 to 4 were acquired in *controlled* conditions (uniform background and illumination, high-quality camera), sessions 5 to 8 are referred to as *degraded* (captured by a webcam in an office environment) and finally, sessions 9 to 12 are obtained in so-called *adverse* conditions (public environment with arbitrary conditions, but with the same high-quality device than sessions 1-4). Figure 2.3 shows examples of these different scenarios. The 52 subjects are divided into two disjoints sets $g1$ and $g2$ used for evaluation and test respectively. Unlike the XM2VTS database, the evaluation and test sets do not contain the same subjects, hence avoiding any face recognition system to be tuned to specific identities. Note also that the BANCA database contains an additional set of 30 people, acquired in the same conditions and referred to as the world data.

There are seven distinct protocols defined with this database. Basically, they consists in specifying which conditions (hence sessions) should be used for training and testing, and are summarized in Table 2.1. In this thesis, we will only focus on 5 out of the 7 protocols: Match controlled (Mc), Unmatched adverse (Ua), Unmatched degraded (Ud), Pooled test (P) and Grand test (G).

Protocol	Training	Testing
Mc	controlled	controlled
Ma	adverse	adverse
Md	degraded	degraded
Ua	controlled	adverse
Ud	controlled	degraded
P	controlled	all
G	all	all

Tab. 2.1: Conditions used in the training and testing phases of the BANCA database.

2.4.3 Performance Measures

In a face authentication scenario, two errors can occur: the false acceptance, when an impostor is recognized as a true client to the system, and the false rejection, when a legitimate user is wrongly rejected. The performance of a face authentication system is thus usually characterized by the false acceptance

Fig. 2.3: Some examples of subjects present in the BANCA database. Images in each column are taken from a different acquisition scenario: the first column corresponds to controlled conditions, the second column depicts the degraded conditions and the third one shows the adverse conditions.

rate (FAR) and the false rejection rate (FRR):

$$\text{FAR}(\tau, \mathcal{D}) = \frac{\text{\# of FA}}{\text{\# of impostor accesses}} \quad [\%] \tag{2.7}$$

$$\text{FRR}(\tau, \mathcal{D}) = \frac{\text{\# of FR}}{\text{\# of client accesses}} \quad [\%] \tag{2.8}$$

where $\mathcal{D}$ denotes the used dataset and τ is the decision threshold. Another widely used measure to summarize the performance of a system is the Half Total Error Rate (HTER), which summarizes the FAR and the FRR in the following way:

$$\text{HTER}(\tau, \mathcal{D}) = \frac{\text{FAR}(\tau, \mathcal{D}) + \text{FRR}(\tau, \mathcal{D})}{2} \quad [\%] \tag{2.9}$$

Note that the FAR and the FRR are highly related, since they depend on the threshold τ: increasing the FAR will reduce the FRR and vice-versa. For this reason, authentication results are often presented using either Receiver Operating Characteristic (ROC) or Detection-Error Tradeoff (DET) curves [Martin 97]. The ROC curve plots the True Positive Rate (i.e. when a client is legitimately accepted, this is the complement of the FAR) versus the False Positive Rate (when a client is wrongly recognized, which corresponds exactly to the FAR) for different values of the threshold. Note that the DET curve directly plots the FAR versus the FRR. Examples of ROC and DET curves can be found in Figure 2.4 and 2.5 respectively. Another widely used quantity to summarize the performance of a system is the Equal Error Rate (EER), defined as the point along the ROC or the DET curve where the FAR equals the FRR.

However, it was noted in [Bengio 05] that ROC and DET curves may be misleading when comparing models. Hence, the so-called Expected Performance Curve (EPC) was proposed, and consists in an unbiased estimate of the reachable performance of a model at various operating points [Bengio 05]. Indeed, in a real-world scenario, the threshold τ has to be set a priori: this is typically done using an evaluation set. Nevertheless, the optimal threshold can be different depending on the relative importance

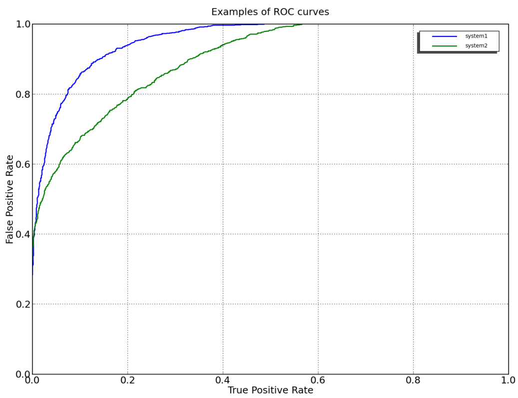


Fig. 2.4: Example of ROC curves

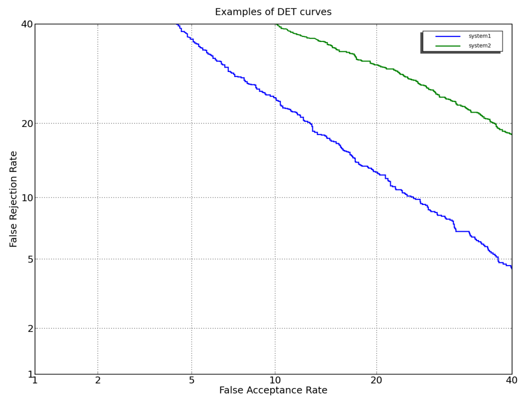


Fig. 2.5: Example of DET curves

given to the FAR and the FRR. Hence, in the EPC framework, $\beta \in [0;1]$ is defined as the tradeoff between FAR and FRR. The optimal threshold τ^* is then computed using different values of β, corresponding to different operating points:

$$\tau^* = \underset{\tau}{\operatorname{argmin}} \quad \beta \cdot \text{FAR}(\tau, \mathcal{D}_e) + (1-\beta) \cdot \text{FRR}(\tau, \mathcal{D}_e) \tag{2.10}$$

where $\mathcal{D}_e$ denotes the evaluation set. Performance for different values of β is then computed on the test set $\mathcal{D}_t$ using the previously found threshold. Note that setting β to 0.5 yields the Half Total Error Rate (HTER) as defined in Equation (2.9). An example of EPC can be found in Figure 2.6.

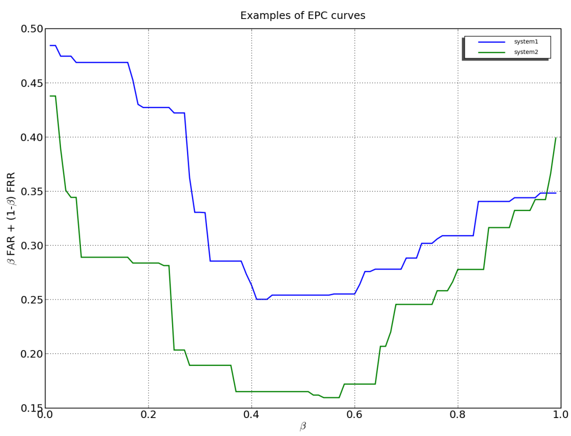


Fig. 2.6: Example of Expected Performance Curves.

2.5 Summary

In this chapter, we first defined the Face Recognition task and stressed its usefulness with respect to other biometric modalities: ease-of-use and non-invasive. The different scenario in which a face recognition system could be deployed were introduced, with an emphasis made on the authentication task, which is the focus of this thesis. For this purpose, two dedicated databases, XM2VTS and BANCA, along with their experimental protocols have been presented.

The current state-of-the art in Face Recognition was briefly reviewed, defining and presenting examples of algorithms for two main families: appearance-based and feature-based. The advantages and drawbacks of the different approaches were discussed, and a deeper overview of face authentication using local features and statistical models was made, since this framework will serve as a basis for this thesis.

Chapter 3

BAYESIAN NETWORKS

3.1 Definition and Properties

Bayesian Networks (also known as Belief Networks or Probabilistic Expert Systems) where first introduced by Pearl [Pearl 88] and could be seen as a marriage between graph theory and probability theory [Jordan 99a]. Bayesian 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. This is thus a representation particularly well-suited to look for correlations among random variables [Heckerman 99]. Conversely, the absence of link between two variables represents a conditional independence assumption. More precisely, a node representing a variable X and directly linked to a node representing a variable Y (i.e. X is the parent of Y) defines a relation having the form of a conditional probability $P(Y|X)$. Usually, X is said to *causes* Y, since fixing the value of X provides information on the value of Y. Note that knowing the value of Y also provides information on X (through the Bayes rule), in this case Y is said to *explain away* X.

More generally, let us define $Pa(X_i)$ as the set of parents of the variable X_i in the directed acyclic graph. The joint probability encoded by a Bayesian Network over the set of variables $\mathbf{X} = (X_1, ..., X_n)$ is then given by the following chain rule:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{3.1}$$

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 X_i may have no parents, in which case its probability distribution is simply given by its prior $P(X_i)$.

In order to better understand the notion of causal relationships and conditional independence assumptions, an example of a Bayesian Network is provided on Figure 3.1. This simple example is referred to as the Asia network [Lauritzen 88] and is interpreted as follow:

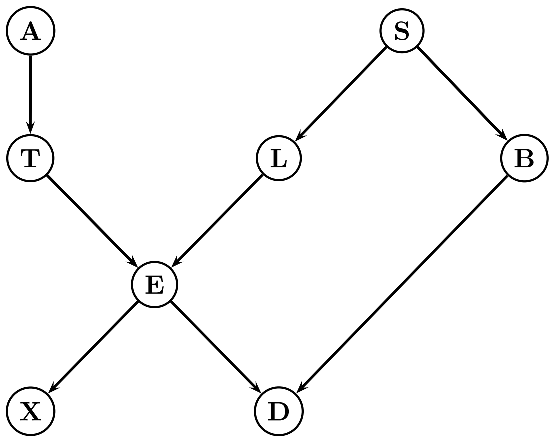


Fig. 3.1: The Asia network.

> *Shortness-of-breath (dyspnoea, node D) may be due to tuberculosis (node T), lung-cancer (node L) or bronchitis (node B), or none of them, or more than one of them. A recent visit to Asia (node A) increases the chance of tuberculosis, while smoking (node S) is known to be a risk factor for both lung cancer and bronchitis. The result of a single chest X-ray (node X) do not discriminate between lung cancer and tuberculosis (either lung cancer or tuberculosis, node E), as neither does the presence or absence of dyspnoea.*

We can see that this network encodes dependencies between these different variables. For instance, dyspnoea is clearly correlated to both tuberculosis and smoking activity. On the other hand, one can also figure out conditional independence assumptions: the smoking activity is not related to a visit to Asia for instance. The parameters of the network are the conditional probability distributions associated to each link, and are usually set according to expert knowledge, and so is the network designed. As an example, it is well-known that smoking increases the chance of developing lung cancer. Hence, we would certainly have $P(L = 1|S = 1) > P(L = 1|S = 0)$.

Using this kind of networks, beliefs on the variables could then be updated once some of the other variables have been observed. For instance, if we know that a person with dyspnoea has indeed been visiting Asia recently, the probability of having tuberculosis will augment. The task of computing such probabilities in a Bayesian Network is referred to as *inference* and will be discussed in the next section.

3.2 Inference

3.2.1 Introduction

Inference is an important task in Bayesian Networks. It consists in computing probabilities of interest, once *evidence* has been entered into the network (i.e. when one or more variables have been observed). Entering evidence consists either in 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. Inference algorithms are then interested in finding the effect this evidence has on the distribution of the other unobserved (or hidden) variables.

Consider the Asia Network depicted in Figure 3.1, and assume that we want to compute the probability of dyspnoea $P(D)$. The joint probability over all the variables in the network $\mathbf{X} = (A, S, B, L, T, E, D, X)$ is given by the chain rule (Equation 3.1). In our case, we have:

$$P(\mathbf{X}) = P(S)P(A)P(B|S)P(L|S)P(T|A)P(E|L,T)P(D|B,E)P(X|E) \tag{3.2}$$

A straightforward (and naive) approach consists in integrating (summing, in the discrete case) over the irrelevant variable:

$$P(D) = \sum_{s,a,b,l,t,e,b,x} P(S)P(A)P(B|S)P(L|S)P(T|A)P(E|L,T)P(D|B,E)P(X|E) \tag{3.3}$$

This summation will hence take exponential time in the number of nodes. Indeed, this procedure does not take advantage of the conditional independence assumptions encoded in the network structure. A more clever way hence consists in distributing the summation among the variables, and to start by summing over variables that corresponds to end nodes. Doing this way is like eliminating end nodes from the graph, and this procedure can be repeated until every node is eliminated, except the one that we are interested in. This procedure is called bucket elimination and is due to Dechter [Dechter 96]: it has the advantage to scale linearly with the number of nodes. However, one must run the algorithm for every marginal calculation. There are actually many different algorithms allowing to perform inference. The most renowned one is certainly belief propagation, originally due to Pearl [Pearl 88]. Here, messages are passed between all the nodes until convergence, and thus multiple queries are answered in a more efficient way. However, this algorithm only works on singly connected graphs, although "loopy" belief propagation could be used as an approximation scheme [Murphy 99b]. Another more generic method to perform exact inference, and which is both able to deal with multiple queries and multiply-connected networks is the Junction Tree algorithm [Lauritzen 88] [Huang 96] [Cowell 99], and will be used in our case. It is presented in more details in the next subsection.

3.2.2 Junction Tree Algorithm

The Junction Tree algorithm is a general method to perform exact inference in both directed and undirected graphical models. Actually, doing inference in directed acyclic graphs could be problematic in the presence of *undirected* cycles (i.e. if there are cycles when arrow's direction are ignored): belief propagation is not guaranteed to converge anymore in this case. Hence, the Junction Tree algorithm first transform the original graph into another representation, called the junction tree (or equivalently, the cluster graph), in which messages are propagated to calculate or update the different probabilities. This algorithm consists in five steps that will be presented in the next paragraphs. Note that this algorithm is introduced with the discrete variable case using the Asia network as example, but it could also be applied to Bayesian Networks containing continuous gaussians distributions, or both [Murphy 99b].

Step 1: Moralization This operation is required only for directed distributions and consists basically in "marrying parents" and removing directions. This is done by adding a link between any pair of variables with a common child and then by replacing directed links with undirected ones. As an example,

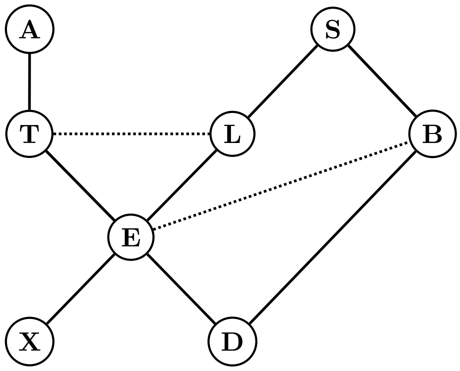


Fig. 3.2: Moral graph.

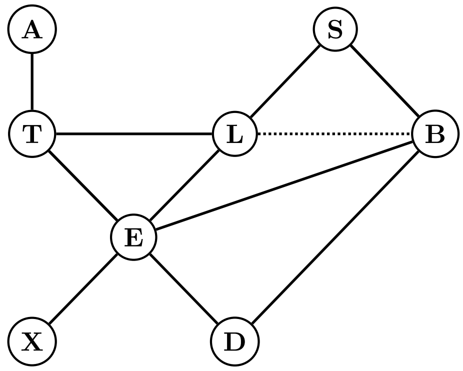


Fig. 3.3: Triangulated graph.

the moral graph of the Asia network is depicted in figure 3.2.

Step 2: Triangulation A graph is said to be triangulated if every undirected loop of length 4 or more has a chord, where a chord is defined as an undirected link between two non-consecutive nodes in a loop. Note that this operation is not unique, and that finding the best triangulation in an arbitrary graph (i.e. the one which adds the less possible links) is NP-hard. The algorithm we used to perform triangulation on any undirected graph is not presented here, but the interested reader could refer to [Cowell 99] or [Stephenson 00] for triangulation algorithm examples. A possible triangulated graph of the Asia network is depicted in Figure 3.3.

Step 3: Junction Tree construction In order to build the junction tree, cliques in the triangulated graph have first to be identified. A *clique* is defined as a complete subgraph where all the nodes are connected [Cowell 99]. Variables in each clique are then grouped into *clusters* to form the nodes of the cluster graph. Note that variables may be present in several cliques: in this case, clusters sharing the same variables must be linked through a *separator* containing these variables to ensure the *running intersection property*. If such a property is satisfied, the cluster graph is said to be globally consistent, which is a necessary condition to pass messages between nodes. Finally, the junction tree is found by removing any link in the cluster graph that causes a loop. Note finally that transforming any Bayesian Network into a Junction Tree can lead to different solutions, which yield different complexities during the propagation of the messages. Building optimal junction trees depends on both the triangulation and the final junction tree construction, and these issues are discussed in [Jensen 94]. Figure 3.4 depicts a possible junction tree obtained from the Asia network.

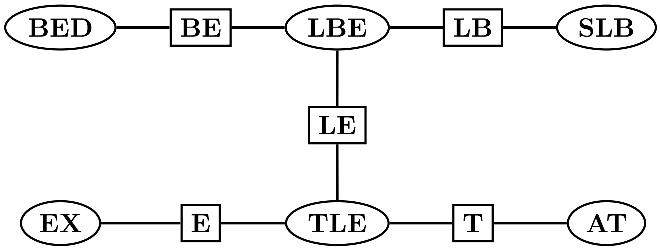


Fig. 3.4: The Junction Tree derived from the Asia network.

Step 4: Potential Assignment A potential $\phi(.)$ is a real-valued function on the configuration of the variables in a cluster. The potential assignment consists in matching the original joint probability distribution defined by the Bayesian Network into its cluster graph representation. In a cluster graph (such as the junction tree), the joint probability distribution is given by the product of the cliques potentials divided by the product of the separators potentials. The probability distribution encoded by the junction tree depicted Figure 3.4 is thus expressed as:

$$P(\mathbf{X}) = \frac{\phi(A,T)\phi(T,L,E)\phi(B,L,E)\phi(S,L,B)\phi(D,B,E)\phi(X,E)}{\phi(T)\phi(L,E)\phi(L,B)\phi(B,E)\phi(E)} \tag{3.4}$$

The initial potential assignment consists in matching cliques potentials to the factors in the original distribution (i.e. the one encoded in the corresponding Bayesian Network), and to set the separators potentials to one. In our examples, a possible assignment for the cliques potentials could be:

$$\begin{aligned}
\phi(A,T) &= P(A)P(A|T) \\
\phi(T,L,E) &= P(E|T,L) \\
\phi(B,L,E) &= 1 \\
\phi(S,L,B) &= P(S)P(L|S)P(B|S) \\
\phi(D,B,E) &= P(D|B,E) \\
\phi(X,E) &= P(X|E)
\end{aligned}$$

and separator potentials are simply given by:

$$\phi(T) = \phi(L,E) = \phi(L,B) = \phi(B,E) = \phi(E) = 1. \tag{3.5}$$

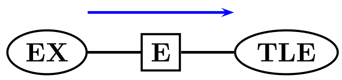


Fig. 3.5: Message passing between two nodes in the Junction Tree.

Step 5: Message Passing In order to derive the message passing scheme in the junction tree, we first need to define the so-called *absorption* procedure. Consider two neighboring nodes $a = (EX)$ and $b = (TLE)$ in a junction tree, with separator $c = (E)$. As can be seen on Figure 3.5, the separator c contains the variables that are present in both cliques a and b. Absorption consists in two operations that update the potentials $\phi(c)$ and $\phi(b)$:

$$\phi^*(c) = \phi(c) \downarrow \phi(a) \tag{3.6}$$

$$\phi^*(b) = \phi(b) \cdot \left(\frac{\phi^*(c)}{\phi(c)} \right) \tag{3.7}$$

where $\downarrow$ denotes the marginalization operator over the variables present in a but not in c. Hence absorption "passes a message" between two clusters. The exact formulas for the absorption procedure are given in Appendix A. A key property of the junction tree is that the product of all cluster potentials divided by the product of all separators potentials remains invariant under absorption. The message-passing scheme hence perform a serie of absorptions so that, at the end of the procedure, clusters and separators potentials contain the exact joint probability of the variables in the corresponding node. The convergence is achieved by performing absorption in both direction along every link in the cluster graph. Figure 3.6 illustrates the complete message passing scheme in the case of the Asia network.

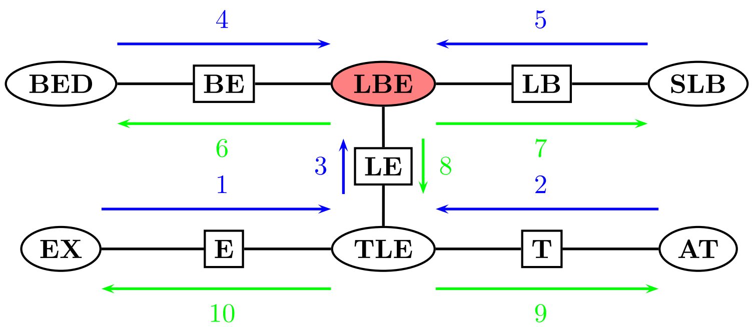


Fig. 3.6: Full message-passing scheme in the Junction Tree. The red node **LBE** is defined as the root, blue arrows denotes the *collect-evidence* procedure, whereas green arrows represents the *distribute-evidence* procedure.

Entering Evidence The Junction Tree algorithm, as previously presented, is meant to compute probabilities without having observed any of the variables. It is thus only able to describe the prior knowledge we have on a given problem, since the distribution $P(X)$ of any variable X can be computed.

Entering evidence consists in making a statement on some of the variables $\mathbf{v}$. Now we would like to compute the effect this observation has on the distribution $P(\mathbf{h}|\mathbf{v})$ of the other, unobserved variables $\mathbf{h}$. To do so, we only have to modify the potential initialization procedure. Consider for instance the SLB cluster (see Figure 3.4) and suppose that the variable S is now fixed to 1. Then, the potential of the cluster should be modified: $\phi(S = 1, L, B) = P(S = 1)P(L|S = 1)P(B|S = 1)$.

3.3 Learning: EM Algorithm

Learning in Bayesian Networks refers either to structure learning, parameters learning or both [Heckerman 99]. In our case, we are considering networks of fixed structure, so the task of learning reduces to learning parameters from data. This is done by maximizing the data likelihood. However, and since hidden variables are present in the proposed models, the log-likelihood of the data cannot be decomposed according to the network topology (note that it would have been possible if every variable would have been observed). In this case, the log-likelihood is given by:

$$\mathcal{L}(\theta, \mathbf{v}) = \log \sum_{\mathbf{h}} p(\mathbf{v}, \mathbf{h}|\theta) \tag{3.8}$$

where θ denotes the parameters of the model (i.e. the conditional probability distributions), $\mathbf{v}$ represents the set of variables corresponding to visible observations and $\mathbf{h}$ is the set of hidden variables. Since maximizing directly Equation (3.8) may be difficult, we simplify the problem using the variational approximation to the Expectation-Maximization (EM) algorithm [Jordan 99b]:

$$\begin{aligned} \mathcal{L}(\theta, \mathbf{v}) &= \log \sum_{\mathbf{h}} p(\mathbf{v}, \mathbf{h}|\theta) \\ &= \log \sum_{\mathbf{h}} q(\mathbf{h}) \frac{p(\mathbf{v}, \mathbf{h}|\theta)}{q(\mathbf{h})} \\ &\geq \sum_{\mathbf{h}} q(\mathbf{h}) \log \frac{p(\mathbf{v}, \mathbf{h}|\theta)}{q(\mathbf{h})} \\ &= \sum_{\mathbf{h}} q(\mathbf{h}) \log p(\mathbf{v}, \mathbf{h}|\theta) - \sum_{\mathbf{h}} q(\mathbf{h}) \log q(\mathbf{h}) \end{aligned} \tag{3.9}$$

where $q(\mathbf{h})$ is the variational parameter and is a distribution over the hidden variables. Furthermore, it can be shown [Jordan 99b] that the optimal setting (i.e. when the bound corresponds to equality) for the variational distribution $q(\mathbf{h})$ is nothing else but $p(\mathbf{h}|\mathbf{v}, \theta)$. Moreover, the second term in Equation (3.9) can be neglected (since it does not depend on θ). This formulation thus becomes equivalent to the classical formulation of the EM algorithm [Dempster 77]. Note that now, the first term in Equation (3.9) can be decomposed according to the network topology. The maximization can thus be done independently for each local conditional distribution.

Hence, starting with initial parameters θ_0, the iteration t of the EM algorithm is performed by first inferring the distribution of the hidden variables $\mathbf{h}$ given the data $\mathbf{v}$ and the current settings of the parameter θ_t using the Junction Tree algorithm (E-step), and then by finding the parameters that maximize the log-likelihood as defined by the first term in Equation (3.9) (M-step).

$$\textbf{E-step:} \qquad \text{compute} \quad p(\mathbf{h}|\mathbf{v}, \theta_t) \tag{3.10}$$

$$\textbf{M-step:} \qquad \theta_{(t+1)} \leftarrow \underset{\theta}{\operatorname{argmax}} \sum_{\mathbf{h}} p(\mathbf{h}|\mathbf{v}, \theta_t) \log p(\mathbf{v}, \mathbf{h}|\theta) \tag{3.11}$$

Maximum A Posteriori (MAP) learning The Maximum A Posteriori (MAP) parameters estimation framework provides a way of incorporating prior information on the training process [Gauvain 94]. It is particularly useful when the number of training examples per class is relatively small with respect to the number of parameters to learn. In this case, the Maximum Likelihood (ML) criterion could even give inaccurate estimates. The idea of MAP learning was first used in speaker verification [Gauvain 94] [Reynolds 00] and was also successfully applied in face authentication [Cardinaux 03] [Lucey 04].

In the case of MAP learning, a prior distribution $g(\theta)$ is set on the parameters, hence the MAP estimation of the model's parameters is given by [Gauvain 94]:

$$\theta_{\mathtt{MAP}} = \underset{\theta}{\operatorname{argmax}}\, g(\theta|\mathbf{v}, \mathbf{h}) \tag{3.12}$$

$$= \underset{\theta}{\operatorname{argmax}}\, p(\mathbf{v}, \mathbf{h}|\theta) g(\theta) \tag{3.13}$$

Assuming parameters independence, and thanks to Equation 3.9, we can specify a prior distribution on each local conditional probability distribution in the Bayesian Network, and thus optimize them independently. The update formula for the parameters in both ML and MAP settings for the proposed models can be found in Appendix B.

3.4 Summary

In this chapter, the necessary theoretical background for designing, computing probabilities and learning the proposed models was presented. We formally defined Bayesian Networks and their properties, especially concerning the correlations and independence assumptions between different random variables, through a simple example.

The problem of computing probabilities in such models, inference, was discussed, and the Junction Tree algorithm was presented in details. We also presented the Expectation-Maximization algorithm, which will be used to learn models parameters with hidden variables. In particular, both the Maximum Likelihood (ML) and Maximum A Posteriori (MAP) criterion were presented, since they will be used to build different models using the available amount of training data, which may be limited.

Chapter 4

Modelling The Face With Bayesian Networks

4.1 Proposed Model

The proposed model relies on two main assumptions. First, we believe that salient facial features such as the eyebrows, the eyes, the nose and the mouth provide enough discriminative information between individuals. This is evidenced by the numerous works performing face recognition based on salient facial features only. Such algorithms include for instance the Elastic Bunch Graph Matching algorithm [Wiskott 99], the Modular Eigenfaces [Pentland 94] or an approach based on Support Vectors Machine [Heisele 03]. In the case of statistical models, the work of Lucey et al. [Lucey 04] showed that modelling features extracted from salient regions of the face using different Gaussian Mixture Models (GMM) and then combining the results is actually more efficient than modelling the whole face using a single GMM.

Secondly, it is assumed that facial features are correlated and thus should not be considered independently. Actually, classical generative models make strong independence assumptions on the way that face image data are generated. Indeed, in the GMM framework as applied in [Sanderson 03], [Cardinaux 03] and [Lucey 04], overlapping blocks extracted from the face image are considered to be independent. The likelihood of the set of features vectors $\mathbf{v} = \{v_1, v_2, ..., v_T\}$ derived from the blocks is given by:

$$\mathcal{L}(\mathbf{v}|\theta) = \prod_{t=1}^{T} P(v_t|\theta) \tag{4.1}$$

As can be seen on Equation 4.1, feature vectors are considered to be independent (through the product). This is obviously not the case in a face image. Consider the two eyes for instance: the block containing the right eye is certainly related to the block containing the left one.

The one-dimensional HMM usually decomposes the face image vertically as a sequence of horizontal strips, and model the features extracted from each strips by a Gaussian Mixture [Samaria 94b] [Nefian 98] [Cardinaux 04]. This model is hence more powerful than a GMM, since different parts of the face are modelled by a different Gaussian Mixture. However, the independence assumption within the sequence of observations $\mathbf{v} = \{v_1, v_2, ..., v_T\}$ remains. Consider the state sequence $\mathbf{q} = \{q_1, q_2, ..., q_T\}$, the likelihood of the 1D-HMM is then given by:

$$\mathcal{L}(\mathbf{v}|\theta) = \prod_{t=1}^{T} P(v_t|q_t, \theta) \tag{4.2}$$

Approaches based on the embedded HMM [Nefian 00], also called pseudo-2D HMM [Eickeler 00] [Cardinaux 04] add another level of precision, since in this case, strips are not considered directly, but decomposed into blocks, and then modelled by an embedded HMM, which will give the emission probabilities of the main, vertical HMM. Nevertheless, all these HMM-based approaches, as well as models based on *dynamic* Bayesian Networks, such as coupled HMM [Nefian 02] introduce *structure* into the observations, but are not able to capture correlations *between* observations. Actually, all these statistical models (excepting GMMs) only constrain the *ordering* of the observations through state sequences (i.e. the nose has to be above the mouth for instance) but are not capable of modelling relationships of the form "this nose is correlated with this mouth" due to the output independence assumption (Equation 4.2).

The Bayesian Networks framework allows to design generative models capable of encoding correlations between different features extracted from a face image. The proposed model depicted in Figure 4.1 is hence trying to capture relationships between facial features. Actually, we would like to consider that certain features (i.e. observations) are generated by the same *causes*. Consider the feature vectors corresponding to the eyes for instance. It is very likely that these two observations were generated by the same hidden cause. At least, considering that these features are completely independent is, in our opinion, not true. Following the same idea, we consider, at a more global level, that all the observations extracted from the same face image are generated by the same identity. In our model, shaded nodes are representing visible observations, whereas white nodes are representing the hidden causes that generated these observations. The model can be explained as follow: the root node on the top represent a face (node F), which is composed of relationships between eyebrows (node B), eyes (node E), the nose (node N) and the mouth (node M). Hence, this variable is used to model the correlations between the different face regions. The variable F then generates certain types of facial features (such as a small nose, or broad lips for instance), represented by the nodes at the second level. And finally, these types of facial features generate the corresponding observations. Note that our model does introduce relationships between observations: if the node O_{le} is observed, information about the node O_{re} can be inferred through the node E for instance. Conversely, one can explicitly notice that such a model generates related observations by the same hidden causes, and hence overcoming the limitation that can be found in GMM for instance.

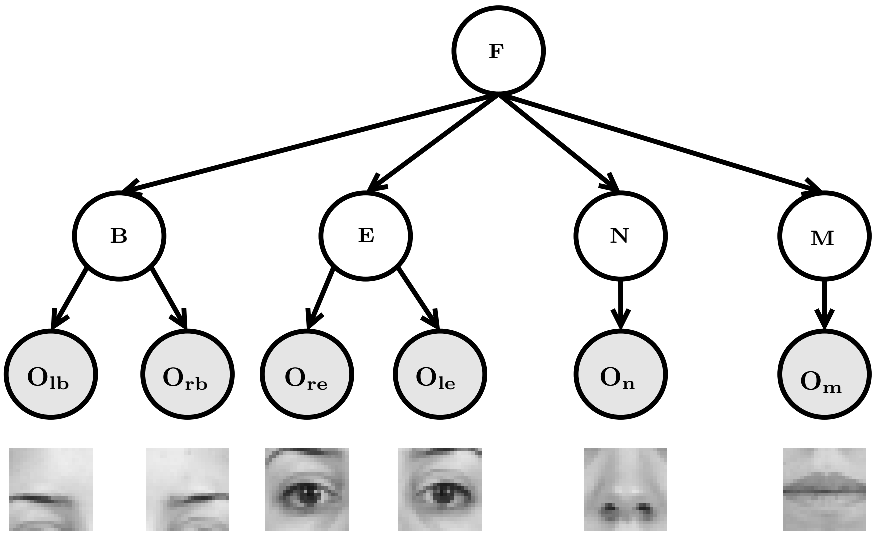


Fig. 4.1: The proposed model: observed salient facial features are generated by a tree-structured Bayesian Network. Shaded nodes represent visible observations whereas white nodes denote hidden *causes*.

In the proposed model, hidden nodes are discrete-valued and observed nodes are multivariate gaussians. Hence, the probability distributions of the nodes on the first and second level are given by (conditional) probability tables, whereas the distributions of the nodes corresponding to observations are given by conditional gaussian distributions, defined as:

$$P(O = \mathbf{o}|Pa(\mathbf{o}) = i) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp\Big(- \frac{1}{2}(\mathbf{o} - \mu_i)^T\Sigma_i^{-1}(\mathbf{o} - \mu_i)\Big) \tag{4.3}$$

where $O = \mathbf{o}$ stands for a realization of one of the observations and $Pa(\mathbf{o}) = i$ for a possible configuration of its parent. n is the dimension of the feature vector representing a particular observation. The mean μ_i and the covariance matrix Σ_i are the parameters of the conditional gaussian distribution and depend on the value of the parent node. Note also that in our model, diagonal covariance matrices are used. The parameters of the Bayesian Network to be learned are denoted by θ and consists in the entries of the (conditional) probability tables as well as the means μ_i and the covariance matrices Σ_i of the conditional gaussian distributions.

Ultimately, we are interested in finding how well a model θ fits an observed face representation $\mathbf{v}$. This is achieved by computing the probability of the observations given the model, i.e. the likelihood. Defining the set of visible observations $\mathbf{v} = (O_{lb}, O_{rb}, O_{le}, O_{re}, O_n, O_m)$, the likelihood $\mathcal{L}(\mathbf{v}|\theta)$ of a face representation is computed by first inferring the distribution of the hidden variables using the Junction Tree algorithm (Section 3.2.2), and then by summing out over the states of the hidden variables. More precisely, consider the Junction Tree corresponding to our face model and depicted in Figure 4.2. After having entered evidence $\mathbf{v}$ and run the message-passing algorithm, the joint probability distribution over all the variables $\mathbf{X}$ encoded by the tree is given by:

$$\begin{aligned} P(\mathbf{X}) &= \frac{P(F,B,\mathbf{v})P(F,E,\mathbf{v})P(F,N,\mathbf{v})P(F,M,\mathbf{v})P(B,\mathbf{v})^2P(E,\mathbf{v})^2P(N,\mathbf{v})P(M,\mathbf{v})}{P(F,\mathbf{v})^3P(B,\mathbf{v})^2P(E,\mathbf{v})^2P(N,\mathbf{v})P(M,\mathbf{v})} \\ &= \frac{P(F,B,\mathbf{v})P(F,E,\mathbf{v})P(F,N,\mathbf{v})P(F,M,\mathbf{v})}{P(F,\mathbf{v})^3} \end{aligned} \tag{4.4}$$

And the likelihood $\mathcal{L}(\mathbf{v}|\theta)$ is thus simply given by:

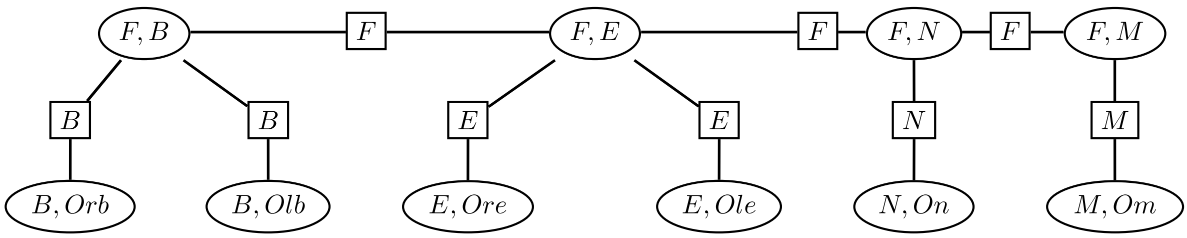


Fig. 4.2: Junction Tree corresponding to the proposed model.

$$\mathcal{L}(\mathbf{v}|\theta) = \sum_{F,B,E,N,M} \frac{P(F,B,\mathbf{v})P(F,E,\mathbf{v})P(F,N,\mathbf{v})P(F,M,\mathbf{v})}{P(F,\mathbf{v})^3} \tag{4.5}$$

4.2 Feature Extraction

In this section, we present the different steps performed to obtain the features we used together with the proposed generative model, from fiducial points localization to feature vectors construction. Feature extraction is a very important step in face processing, and recognition results may heavily depend on the representation of the observations.

There is a huge amount of different features used with different algorithms, including simple raw pixel intensities [Samaria 94a], Gabor Wavelets [Wiskott 99] [Shen 07] , SIFT features [Bicego 06], Histogram of Oriented Gradients [Prince 06] or Local Binary Patterns (LBP) [Ahonen 04] [Rodriguez 06b] to name a few. A popular feature used together with statistical models is the two-dimensional Cosine Transform (2D-DCT) [Nefian 98] [Sanderson 03] [Cardinaux 05]. Since it will be also used in this work, it is presented in more details later on.

As can be seen in Figure 4.1 our model performs classification using features extracted around salient face regions. So, these face components have first to be located in the input image. Again, there exists several approaches to facial features localization, and the interested reader may refer to [Cristinacce 04] or [Felzenszwalb 05] for examples. In this work, the well-known Active Shape Model algorithm [Cootes 95] is used to solve this task and thus described in the next section.

4.2.1 Facial Feature Localisation

Active Shape Models (ASM) were first introduced by Cootes et al. in [Cootes 95] and consists in fitting the shape of an object, using a previously learned global shape model, and represented as a set of landmark points (see Figure 4.3). A key point of this algorithm is that it is able to address the typical variability (present in the training set) encountered in the object of interest.

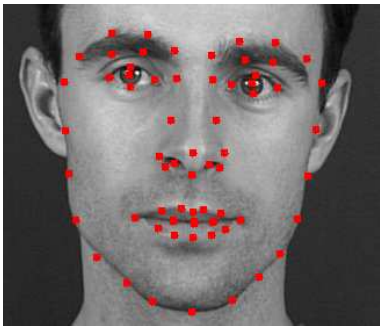


Fig. 4.3: Landmark points of the Active Shape Model.

Model Building To build a deformable model of the shape, one requires a training set where examples of the object are labelled with landmark points. The model is then built by following these steps:

1. Align the shapes in the training set through Procrustes Analysis
2. Compute the mean shape $\bar{x}$, using x_i (1D vectors containing the landmarks coordinates)
3. Find the difference of each shape in the training set with the mean shape $dx_i = \bar{x} - x_i$
4. Get the eigenvectors of the covariance matrix of the dx_i, this will give the modes of variations

Finally, any shape (i.e. set of landmarks) can be represented by the following approximation:

$$x = \bar{x} + \Phi b \tag{4.6}$$

where Φ is the matrix containing the leading eigenvectors of the distance covariance matrix, and b is a vector of weights. The parameters b are constrained such that the generated shapes remains in the allowed deformable domain.

Image Search Once the model for the shape of interest is generated, it could be used in image search to find occurrences of the modeled object. To do so, an iterative procedure is applied:

1. Start with a rough approximation on the location of the object
2. For each point, examine a region of the image around the point
3. Determine the best displacement to a better match, using a model of local observations
4. Check that newly computed points remains in the allowed deformable shape
5. Iterate from point 2 until convergence

In our work, Local Binary Patterns (LBP) are used to model the local observations, as described in [Keomany 06]. Note also that, as in the original ASM, constraints are added to the displacement of each point, such that the shape of the object does not diverge.

4.2.2 Preprocessing

Once the facial features have been localized by the Active Shape Model, a square block is extracted around selected salient regions of the face (see Figure 4.1). An important step then consists in applying an image preprocessing algorithm, which usually aims at minimizing the difference in image appearance. Actually, most of the variation within the appearance of the image corresponding to the same identity is introduced by the difference in illumination. Indeed, it was shown in [Adini 97] that such variation may be larger than variation due to the difference in identity.

There exists different preprocessing algorithm to cope with the difference in lighting conditions. The simplest one is certainly histogram equalization [Acharya 05], and is widely applied in the context of face recognition. There exist more sophisticated algorithms trying to explicitly remove the illumination component of an image. Examples include the Center-Surround Retinex and Multiscale Retinex [Jobson 97a] [Jobson 97b], the Gross & Brajovic's algorithm [Gross 03] or the method proposed by Tan et al. [Tan 07]. Figure 4.4 shows examples of a face preprocessed by these algorithms. In this work however, the simple histogram equalization is used, since it has been shown to perform well when considering local observations [Short 06].

Fig. 4.4: Examples of image preprocessing algorithms: original image, histogram equalization [Acharya 05], multiscale retinex [Jobson 97a], Gross & Brajovic [Gross 03] and Tan [Tan 07].

4.2.3 Two dimensional Discrete Cosine Transform

The two-dimensional Discrete Cosine Transform (2D-DCT) is a compression technique well-suited to image data. It has a strong energy compaction property, since image data can be approximated with relatively few coefficients. Actually, it is the de-facto standard for many compression algorithms, including MPEG and JPEG for instance. In our framework, 2D-DCT is used to perform dimensionality reduction, but has also the advantage of minimizing the amount of noise in the feature vectors [Sanderson 05]. Such features has been already used in the context of face recognition: either with a simple nearest neighbor classifier [Ekenel 05] or together with statistical models [Nefian 98] [Sanderson 03] [Cardinaux 05].

In our case, squared blocks of size NxN are extracted around salient facial features, based on the landmarks position given by the ASM. They are then decomposed in terms of orthogonal 2D DCT basis functions. The result is a squared matrix with coefficients given by:

$$C(v,u) = \alpha(v)\alpha(u)\sum_{y=0}^{N-1}\sum_{x=0}^{N-1} b(y,x)\beta(y,x,v,u) \quad \text{for} \quad v,u = 0,1,2,...,N-1 \tag{4.7}$$

where

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } v = 0 \\ \sqrt{\frac{2}{N}} & \text{for } v = 1,2,...,N-1 \end{cases} \tag{4.8}$$

and

$$\beta(y,x,v,u) = \cos\left[\frac{(2y+1)v\pi}{2N}\right]\cos\left[\frac{(2x+1)u\pi}{2N}\right] \tag{4.9}$$

The coefficients are then ordered according to a zig-zag pattern, reflecting the amount of information stored. The final observation vector for each block is then built by retaining the M most significant coefficients.

4.3 Experiments

In this section, experiments are presented to assess the validity of the assumptions made to design the proposed generative model for the face. First, we empirically show that the proposed model indeed learns to cluster different types of facial features. Secondly, our model is compared to another generative model where different salient regions of the face are considered independently. Experiments show that adding explicit correlations between information extracted from face data helps at modelling identity, since face authentication results are improved. Finally, we compare the proposed system with other statistical generative models widely used in face recognition (GMM, 1D-HMM and Pseudo-2D HMM), but also with two popular appearance-based approaches: PCA and LDA (see Section 2.1).

4.3.1 Learning Facial Features Types

The purpose of this experiment is to verify if the proposed model actually learns the different types of facial features. We believe that an identity may be explained by a combination of facial features and that a person is recognized thanks to say, his/her round eyes that comes together with a thin nose and broad lips for instance. To do so, we first would like to see if our model is capable of learning typical difference in the appearance of salient facial features among individuals.

Let us define a dataset containing n faces coming from different identities: $\mathcal{D} = \{x_1, x_2, ..., x_n\}$, where x_i represent data extracted from a face image: $x_i = \{O_{lb}, O_{rb}, O_{le}, O_{re}, O_n, O_m\}$ (i.e. a set of feature vectors). A model θ is learned with $\mathcal{D}$ using the EM algorithm and the Maximum Likelihood criterion, as described in Section 3.3. Recall that hidden nodes of the proposed model are meant to represent different types of facial features.

Let us define $B(x_i)$, $E(x_i)$, $N(x_i)$ and $M(x_i)$ as functions taking values in $\{0, 1, ..., J\}$ and representing the type of the corresponding facial feature for the face representation given by x_i. Actually the value of such a function is given by the most probable value of the hidden node corresponding to the facial feature. For instance, the type of eyebrows for the face represented by x_i will be given by: $B(x_i) = \underset{j}{\text{argmax}} \quad P(B = j|x_i, \theta)$.

Once the model has been learned, each example x_i is again given to the model. The probability distributions of the hidden nodes are inferred using the Junction Tree algorithm. We then record, for each x_i, the value of $B(x_i)$, $E(x_i)$, $N(x_i)$ and $M(x_i)$. We are now able to cluster the different x_i according to their facial feature type. Consider the set E^j for instance, which represent the eyes type j. Formally E^j is defined as: $E^j = \{x \in \mathcal{D} \quad \text{such that} \quad E(x) = j\}$.

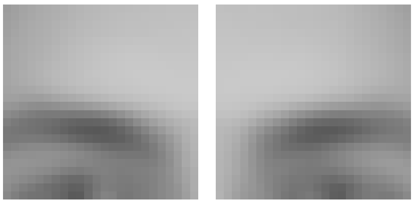
In order to get a visual representation of the different facial features types, the mean image of the extracted blocks (see Figure 4.1) is computed for each type of each facial feature. Figure 4.5 shows some examples of the different types of facial features obtained this way.

It can thus be seen that our model indeed learns different types of facial features, and that an identity could be represented as a combination of the different facial features. Figure 4.6 shows two synthesized faces using mean images as defined above. Note that a difference in identity could be clearly seen between these two different faces. Indeed, the face on the left is looking like a male whereas the face on the right resembles more to a woman.

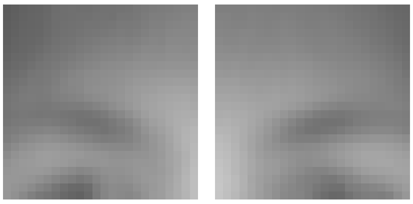
4.3.2 Face Authentication: Introducing Correlations

The main assumption that drove us towards the proposed model is that relationships may exist between salient facial features, and as such, information extracted from face images should not be treated independently. To verify this hypothesis, we compare the performance on the face authentication task of our model to another generative model based on Gaussian Mixtures Model, but where features extracted from face parts are assumed to be independent.

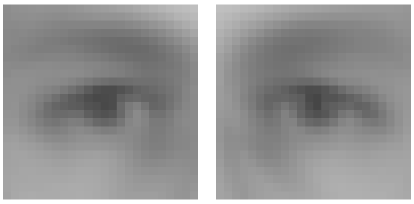
The model used for comparison purpose is a variation of the so-called *Partial Shape Collapse* GMM (PSC-GMM) [Lucey 04]. Actually, this model is quite close to ours. Indeed, our proposed model could be seen as a collection of GMMs related to each other, since observations are assumed to be Gaussians,



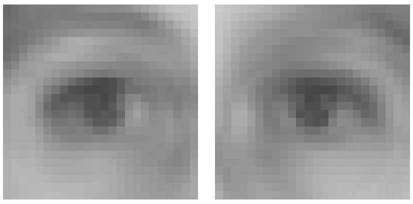
(a) eyebrows for $j = 1$



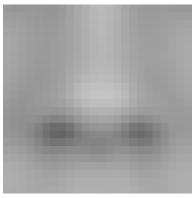
(b) eyebrows for $j = 7$



(c) eyes for $j = 5$

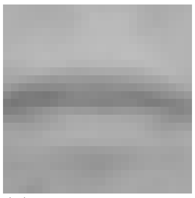


(d) eyes for $j = 6$



(e) nose for $j = 1$

(f) nose for $j = 7$



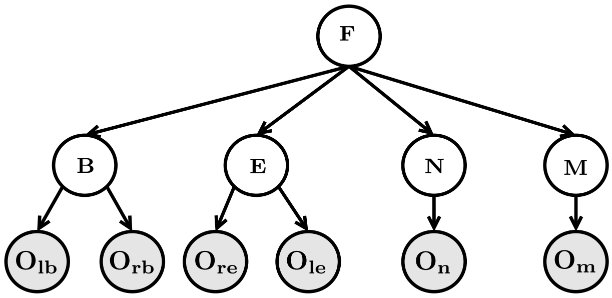
(g) mouth for $j = 4$

(h) mouth for $j = 7$

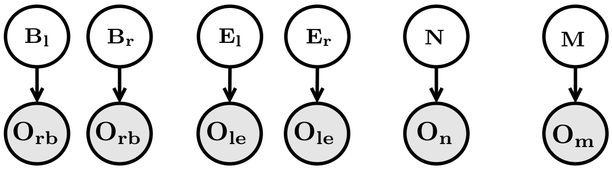
Fig. 4.5: Examples of mean images obtained for different facial features types.

Fig. 4.6: Synthetic faces generated with the mean image of different facial features types.

and the *causes* (i.e. nodes representing the types of facial features) are discrete. Figure 4.7 shows the difference between the two models. Note that in the case of the PSC-GMM, a classifier is built for each facial feature, and results are then combined using the sum-rule [Kittler 98]. In the proposed model, termed as BNFace, two extra layers of relationships are introduced between the observations. Firstly, the symmetric facial features, eyebrows and eyes, are assumed to be generated by the same underlying process. Secondly, all facial features are related to each other thanks to the node at the top.



(a) The proposed model: BNFace



(b) The PSC-GMM [Lucey 04]

Fig. 4.7: Difference between the proposed model BNFace and the PSC-GMM: a generative model where facial features are treated independently.

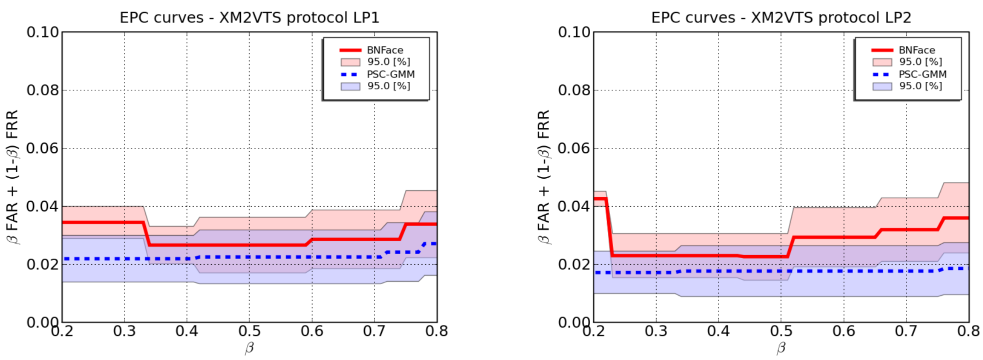
Experimental Setup Presented results were obtained using extracted blocks of 24x24 pixels. So, for each facial feature, blocks centered on the corresponding landmark point given by the ASM are cropped from the original image. Besides, and for each facial feature, additional blocks with shifts of 2, 4 and 6 pixels in each direction are also extracted. This is done in order to increase the number of training data, but also to be robust to small localization errors. Note that this technique has already been used in [Martinez 02]. Hence, for a single observation corresponding to a particular facial feature, 25 such blocks are obtained. Then, histogram equalization is applied to each block to enhance its contrast. Finally, the first 64 coefficients were retained from the 2D-DCT on the blocks.

Regarding the model parameters itself, the cardinality of the hidden nodes were set to 3 at the top level, and to 8 at the second level, and the adaptation parameter α for the MAP training was set to 0.4 (see Appendix B). The parameters for the feature extraction and the model, as well as the decision threshold τ were selected by minimizing the Equal Error Rate (EER) on the evaluation set. For the PSC-GMM, we used 512 gaussians as suggested in [Lucey 04] for each of the six GMMs corresponding to the six extracted facial features. Note also that, unlike Lucey's original work, the regions corresponding to the cheeks, the nose bridge and the forehead are not considered here. Actually, we give the PSC-GMM the exact same features as the one used in the BNFace model. This is done to clearly have an idea of the modelling power of each of the two models.

Results & Discussion Results are presented using the Half Total Error Rate (HTER) computed on clients and impostor accesses. We also present Expected Performance Curves (EPCs) to compare the two models, since they provide a more complete description of the behavior of the systems in different settings (i.e. relative importance given to the False Acceptance Rate with respect to the False Rejection Rate). Moreover, a modified version of the standard proportion test, as described in [Bengio 04] is used in order to compute 95% confidence intervals around EPCs.

System	LP1	LP2
BNFace	2.65	2.26
PSC-GMM	**2.25**	**1.76**

Tab. 4.1: HTER performance on the test set of XM2VTS: BNFace and PSC-GMM



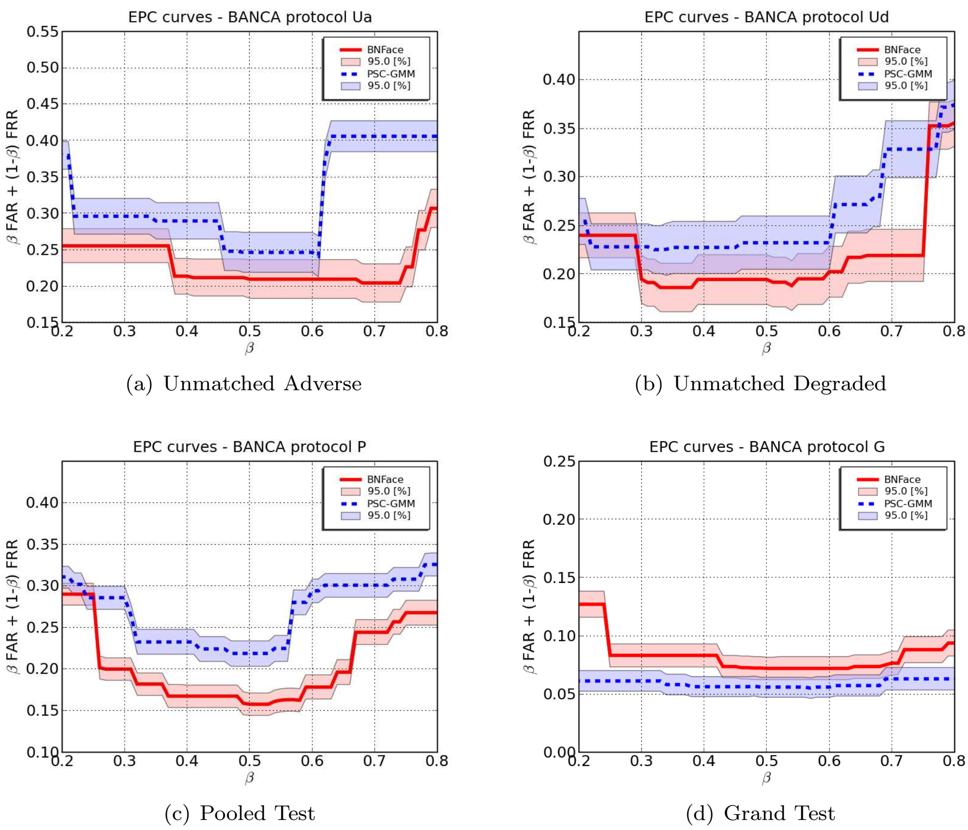
(a) LP1 (b) LP2

Fig. 4.8: EPC curves on the test set of XM2VTS : BNFace and PSC-GMM.

Results obtained on the XM2VTS database can be seen in Table 4.1 and Figure 4.8. They show that there is basically no difference in authentication performances between the two systems. Actually, it can be seen on Figure 4.8 that the confidence intervals are largely overlapping along the curves. This could be explained by the closely controlled acquisition conditions of this database. Indeed, such conditions may prevent the two systems to reveal their respective modelling power. However, one can still see that the PSC-GMM is performing a bit better than the proposed model, which suggest that adding relationships between observations is not really useful in such a case.

System	Mc	Ua	Ud	P	G
BNFace	2.21	**20.93**	**19.42**	**15.76**	7.19
PSC-GMM	**1.83**	24.62	23.21	21.87	**5.59**

Tab. 4.2: HTER performance on the test set $g2$ of BANCA: BNFace and PSC-GMM.



(a) Unmatched Adverse

(b) Unmatched Degraded

(c) Pooled Test

(d) Grand Test

Fig. 4.9: EPC curves on the test set $g2$ of BANCA: BNFace and PSC-GMM.

On the other hand, results on the more challenging BANCA database are more interesting, since in this case, there is a clear difference between the two investigated models. Note that we did not show the EPCs for the Mc protocol, since the conditions, as well as the obtained results, are very similar to the ones with the XM2VTS database, and thus similar conclusions can be drawn. Results show that when the training and testing data sets are acquired in good conditions (protocol Mc), the simpler PSC-GMM is performing better than the proposed model (by a relatively small margin though). On the other hand, one can see that in the case of mismatch in training and testing conditions (i.e. when the acquisition conditions of the gallery and the probe images are different), such as the Ua, Ud and P protocols, the proposed model is consistently performing better than the PSC-GMM. This is especially true on the P protocol, where the difference in performance is statistically significant for a wide range of operating points (see Figure 4.9(c)). This result is interesting, since the protocol P can be viewed as a summary of the overall performance of a face authentication system across different conditions. It could also be seen as the most challenging setting, since training uses images acquired under controlled conditions and testing is done across all the conditions (see Table 2.1 and Figure 2.3).

Two main observations can be stated based on these results. First, when the training and testing conditions are different, modelling the relationships between facial features clearly helps at recognizing identities. Second, it also suggests that the BNFace model has a better generalization ability to unknown conditions than the competing model.

4.3.3 Face Authentication: Comparison to Other Systems

In this section, we compare the proposed generative model, BNFace, to classical statistical models such as GMM, 1D-HMM and Pseudo 2D-HMM as applied in [Cardinaux 05]. We also provide results with two popular appearance-based algorithms: PCA [Turk 91] and LDA [Belhumeur 97].

Note also that presented results were obtained following the *strict* usage of the protocols defined with each database. Hence, for the XM2VTS database, we use 600 images corresponding to all client training data to train the world models, but also to build PCA and LDA projection matrices. For the BANCA database, the additional set containing 10 images of 30 individuals (acquired under all conditions) was used for the same purposes. In particular, we do not use any other corpus or database, nor mirroring the available images to build either world models or subspace representations, as it was sometimes done in other studies [Cardinaux 05] [Messer 04]. Doing this way enables a fair comparison among the different approaches, since exactly the same data and experimental protocols were used for each tested system. Note also that it may explain the difference with results previously published in the literature.

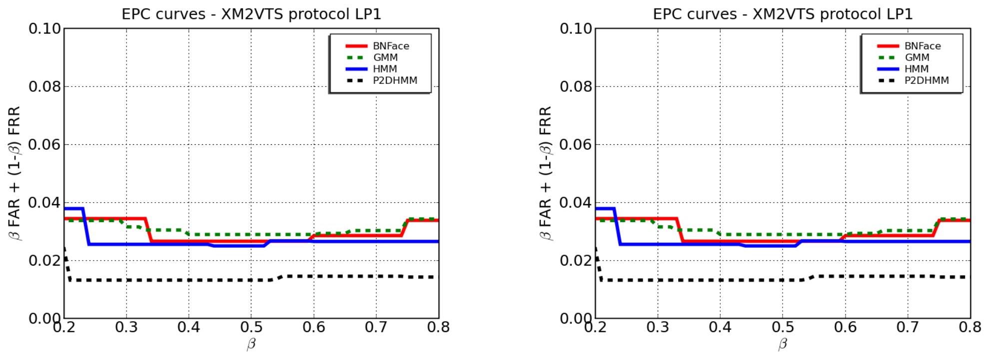
Experimental Setup Parameters for the proposed model are the same as described in the previous section. Regarding the other approaches, faces were first cropped from the original images, resized to 64x80 pixels, converted to grayscale and preprocessed with histogram equalization. Features for the GMM, HMM and Pseudo-2DHMM were extracted using the feature extraction scheme described in [Cardinaux 05]. For the PCA and LDA approaches, 90% of the variance was retained for the PCA projection, resulting in a feature vector of dimension 238 on the XM2VTS database, and 73 on the BANCA database. The nearest neighbor classifier is used with a normalized standard correlation as the metric to perform classification.

Results & Discussion Again, obtained results are presented using both HTER and EPCs. Note that since some of the results are clear enough, not all the EPCs will be presented here. Indeed, we will concentrate on the proposed model and the models obtaining the closest performances. For the sake of clarity, we also omit the 95% confidence intervals.

System	LP1	LP2
BNFace	2.65	2.26
GMM	2.89	1.69
1D-HMM	2.49	1.49
P2D-HMM	**1.31**	**0.45**
PCA	6.83	4.29
LDA	5.83	3.64

Tab. 4.3: HTER performance on the test set of XM2VTS: BNFace and Generative models.

Results obtained on the XM2VTS database show that local feature based statistical models are far better than the two holistic approaches (Table 4.3). It thus suggest that, in the exact same training conditions, generative models are particularly well-suited to tackle the face authentication task. Note that this results could be surprising compared to what can be found in the literature, but we would like to insist on the fact that the exact same data and training procedure where applied to all the systems, thus allowing a fair comparison.



(a) LP1 (b) LP2

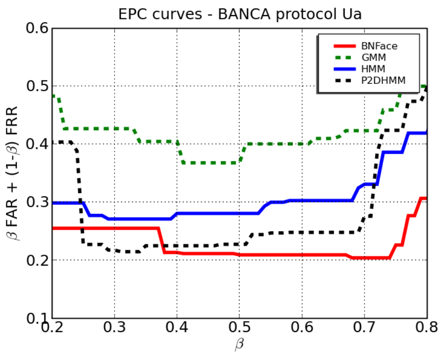
Fig. 4.10: EPC curves on the test set of the XM2VTS database: BNFace and Generative models.

Regarding the comparison between the generative models themselves, one can see that the best performance is obtained using the Pseudo 2D-HMM (P2D-HMM). This can be explained by the fact that this model is able to add structure to the observations in the form of two-dimensional spatial relationships along the whole face image. Note also that this model is the most complex with respect to the other generative models, and thus may be able to model faces in a more precise way. Conducted experiments also show that the other generative models, GMM, HMM and BNFace yield almost the same performances. Again, we believe that this database may be too easy to fully reveal the difference between the different models.

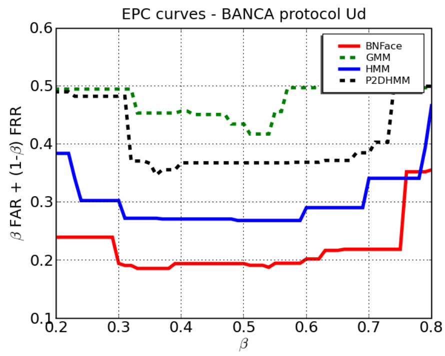
System	Mc	Ua	Ud	P	G
BNFace	2.21	**20.93**	**19.42**	**15.76**	7.19
GMM	1.67	36.79	43.49	33.64	7.87
1D-HMM	2.72	28.08	26.86	24.73	9.04
P2D-HMM	**0.99**	22.76	36.79	27.63	**5.34**
PCA	21.89	41.76	39.33	32.49	26.21
LDA	23.69	41.54	43.69	36.23	28.55

Tab. 4.4: HTER performance on the test set $g2$ of BANCA: BNFace and Generative models.

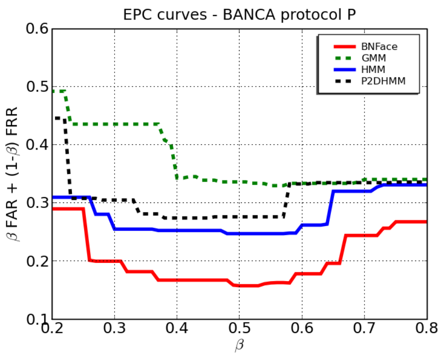
Once again, results on the BANCA database are more interesting: the different approaches exhibit different behaviors (Table 4.4 and Figure 4.11). As before, one can see that generative models usually perform better than the holistic based systems. This could be explained by the fact that there is only



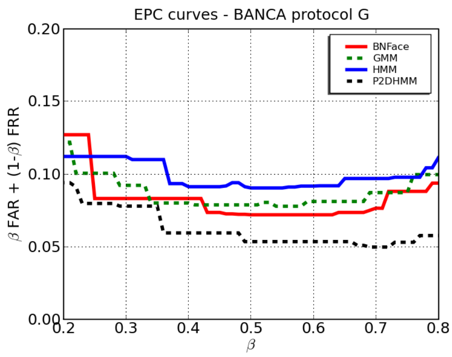
(a) Unmatched Adverse



(b) Unmatched Degraded



(c) Pooled Test



(d) Grand Test

Fig. 4.11: EPC curves on the test set $g2$ of BANCA: BNFace and Generative models.

a few data to build the projection matrices, and this is known to affect the performance. So, generative models are generally more suitable to handle scenarios where only few training data are available, which is a desirable property in real-life scenario. A comparison of the proposed model with GMM, HMM and Pseudo-2DHMM may reveal the advantages and the drawbacks at the models level, when applied to face authentication.

It must be noted that BNFace performs way better than the simple GMM-based system on uncontrolled conditions. This result is particularly interesting since it tends to support two stated hypothesis. First, only a subset of the face image, corresponding to salient facial features, is sufficient to perform authentication. Indeed, in the GMM framework, blocks of pixels are extracted from the whole face image. It thus confirms the results noted in other studies [Martinez 02] [Heisele 03]: systems based on local features are less affected by variations in acquisition conditions. Second, it also suggests that blocks extracted from the face image are correlated and hence should not be treated as if they were independent, as already noticed in the previous experiment (Section 4.3.2).

HMM-based systems are also worth to compare to, since on the contrary to GMM, they actually introduce structure on the features extracted from the face image. However, they act on the structure of the observations, since they constrain their ordering only. Even if these models are performing generally better than the simple GMM approach in unmatched conditions, obtained results are worse than with the BNFace model. It thus suggests that adding explicit correlations between the different facial features is indeed valuable, especially in case of mismatched conditions.

Nevertheless, the P2D-HMM achieves better results that the proposed system in some of the protocols. Results obtained with this approach hence suggest that the two-dimensional spatial ordering along the *entire* face image are important. Note also that this model is the most complex across all generative models. Indeed, it uses much more client-specific parameters to describe a face and is also much more computationally demanding than the proposed system. In Table 4.5, we report the computational time to perform the three tasks involved in face authentication: world-model training, client-model adaptation for one individual (computed on average over the clients), and authentication time for one individual (also computed on average). These quantities were obtained using the BANCA database on a computer with an AMD Athlon 2,6 GHz. We also report the number of client-specific parameters for the proposed system and P2D-HMM. As can be seen on the Table, BNFace's authentication time is for instance more than ten times smaller. Besides, the number of client-specific parameters is also greatly reduced.

System	world model training time	client model adaptation time	individual authentication time	number of parameters
P2D-HMM	3520 sec	$\sim$ 220.2 sec	$\sim$ 9.8 sec	73'726 [Cardinaux 05]
BNFace	86.9 sec	$\sim$ 3.1 sec	$\sim$ 0.6 sec	867

Tab. 4.5: Computational time on BANCA.

4.3.4 Face Authentication: Automatic Face Detection

In this section, results are presented using automatic detection of the face. Actually, the vast majority of face recognition algorithm's performances are provided using manually registered faces. Since face detection (or localization) is the necessary first step to any other face analysis task in a realistic scenario, we believe that robustness to imperfectly located faces is worth investigating. Besides, it has been shown [Rodriguez 06a] that performance of some systems may dramatically drop if the face is automatically (and hence not perfectly) detected.

Face Detection In order to detect the face in the input image, a variant of the face detector proposed by Fröba and Ernst [Fröba 04] is used. The detector employs local features based on the Modified Census Transform (MCT), which represent each location of the image by a binary pattern computed from a 3x3 pixel neighborhood. Each input image is scanned and all possible windows in a given scale range are analyzed. Each window is then classified as containing a face or not. The classification is carried out using a cascade classifier in a similar way than in [Viola 01]. Overlapping windows labelled as faces are then merged together so as to provide a unique bounding box containing the detected face. Eyes position are then inferred from the position and the scale of the bounding box. Note that if a face is missed by the detector, eyes position are estimated from other images of the same individual within the same recording session, but where the face was effectively detected.

Experimental Setup First, faces are automatically located using the face detector described above. The face detector has been trained using face images coming from the following databases: CMU, BioId, AR and Purdue. Hence, no prior knowledge on the face images used in the authentication experiments were introduced in the detection process. Then, the same experimental setup as before is used here. Note however that in this case, the cropping of the faces is done using automatically detected eyes position, hence introducing differences in scale and translation, as can be seen in Figure 4.12. Regarding the proposed the model, and since there is no direct face cropping, facial features are still localized by the ASM, which is now initialized with the eyes location inferred from face detection. The parameters for both feature extraction and the different models remain the same.

Results & Discussion Note that here we present results on the BANCA database only, since it is the more interesting due to its challenging conditions. Table 4.6 provide HTER performance for the different investigated protocols.

Fig. 4.12: Illustration of cropped faces using manually located eyes (first row) and automatically located eyes (second row). Note the variations in scale between column 2 and 4 for instance.

System	Mc	Ua	Ud	P	G
BNFace	6.96	21.73	**27.69**	**20.98**	8.15
GMM	4.26	34.17	37.76	31.23	9.63
1D-HMM	10.16	29.01	34.10	29.13	11.91
P2D-HMM	**1.67**	**21.41**	34.81	26.15	**5.63**
PCA	24.46	40.61	39.68	36.00	25.41
LDA	28.88	44.46	43.33	39.86	34.98

Tab. 4.6: HTER performance on the test set $g2$ of BANCA with automatic registration

It can be seen from these results that some of the systems are not much affected when the face is registered automatically. It is especially true with appearance-based algorithms, PCA and LDA. This result is surprising, but could be explained by noting that these algorithms already yield to quite high error rates when manual face registration was used. Actually, errors may be due to the intrinsic poor modelling ability of the algorithms themselves rather than to automatic registration. The same remark could also be applied to the GMM and 1D-HMM systems. Note also that the larger degradation is attained in controlled conditions (Protocol Mc), this suggest that in this case, most of the errors are effectively introduced by the non-perfect detection of the face. Actually, localization error is certainly the most prominent factor affecting performance, whereas in the mismatched cases (Protocol Ua, Ud and P), the contribution of approximation on localization on the overall face verification error may be balanced with the other sources of variations, such as image quality and illumination.

Again, the best performances are obtained using both BNFace and the P2D-HMM approach. In order

System	Mc	Ua	Ud	P	G	average
BNFace	3.14	1.03	1.42	1.33	1.13	1.61
P2D-HMM	1.68	0.94	0.94	0.94	1.05	1.11

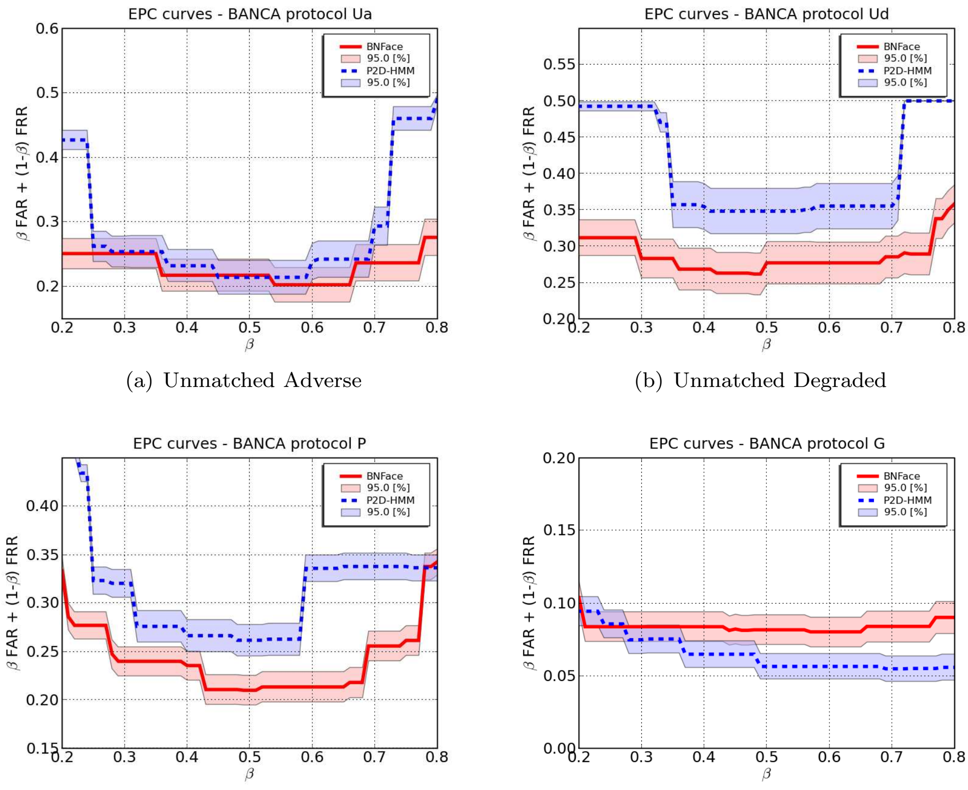
Tab. 4.7: Degradation ratio in face verification performance between automatic and manual localization of the face for the two best models.

to have a deeper understanding on how these two systems are affected by non-perfect face detection, Table 4.7 presents the degradation factor between results obtained with manual and automatic registration. The degradation factor is obtained by taking the ratio between the HTER obtained with automatic face detection and the HTER obtained with manual annotations.

As can be seen on this table, the BNFace approach is more affected by errors introduced by automatic face detection, whereas the P2D-HMM is more tolerant to such approximations (it even performs better in some cases). As a matter of fact, errors in eyes localization are certainly propagated to other facial features during the ASM search procedure. Hence, our model may be more affected than the P2D-HMM: the latter considers only the eyes position to crop the entire face image. However, absolute performance between the two systems remains comparable. A deeper comparison is provided with EPC curves and their 95% confidence intervals (Figure 4.13). Note that we did not present the results for the Mc protocol, since in this case the P2D-HMM outperforms the proposed system by a large margin (see Table 4.6). The curves show that our model is particularly well-suited to unmatched conditions: this is particularly evidenced by the results obtained on the Ud and P protocols, where BNFace achieves a significant lower error rate than P2D-HMM. In the Ua protocol, performance of both system are almost identical, whereas in the protocol G, the P2D-HMM performs a little bit better than BNFace.

4.4 Summary

In this chapter, we introduced a new statistical generative model dedicated to the face object. Our model considers observations extracted around salient facial features, and attempts to relate them to each other. In comparison to classical generative models, the proposed model is able to add explicit correlations to the observed variables, and should thus be theoretically better suited to model an identity. Actually, the model allows to encode properties on the way observations extracted from a face image could be



(a) Unmatched Adverse

(b) Unmatched Degraded

(c) Pooled Test

(d) Grand Test

Fig. 4.13: EPC curves on the test set $g2$ of BANCA with automatic registration.

generated. In particular, similar features are considered to be generated by the same underlying cause, a property that classical generative models could not take into account.

Conducted experiments showed that the proposed model is able to cluster faces according to facial feature types, such as a big nose or thin lips for instance, thus supporting our idea that an identity could be retrieved based on a particular combination of different facial features. In an authentication scenario, this new model was shown to perform better than a corresponding model where facial features are treated independently, again supporting the main hypothesis that features extracted from a face image are related, and should not be considered to be independent. Our model was also compared to currently used statistical models and yield better or equivalent results. This is particularly true in case of mismatched conditions between gallery and probe images, which is a desirable property in a real-world scenario. Moreover, and at the same level of performance, our model is much less complex and thus more computationally efficient than existing approaches.

Chapter 5

Using Auxiliary Information

5.1 Motivation

In this chapter, we show how to use the developed Bayesian Networks framework for face authentication using complementary information. Actually, the proposed model in Chapter 4 only acts on facial features extracted form grayscale images. For instance, all kind of geometrical or configural information is lost in our model. Different studies on face recognition (both in automatic, computer-based systems and in cognitive research) advocate for the use of other source of information to properly recognize faces. As seen before, the Bayesian Networks framework allows to easily design models based on prior domain knowledge. In this chapter, several extended generative models are hence proposed to incorporate different observations as complementary clues to the information extracted around salient facial features. Adding such information is motivated by recent studies in human face processing, such as [Sinha 06], which tries to understand what information people actually use to recognize other individuals.

For instance, it has been empirically shown that a low resolution of a familiar face could be easily recognized by humans. Hence, recognition does not rely on facial features only. Another existing debate is whether the human visual system relies on a piecewise processing or if it is considering the face as a whole [Schwaninger 02]. Actually, conducted experiments showed that recognition relies on both featural and configural information. Regarding facial features configuration, early automatic face recognition systems were relying on such geometrical information [Kanade 73] [Nixon 85]. Cognitive studies also show that this kind of information is used by human beings. Actually, it is shown that distance ratio between face parts may be more important that distance themselves. Other evidences used in the human visual system are including face shape and color for instance [Yip 02]. In this chapter, we would like to see if adding such information as complementary clues to the appearance of facial features helps at discriminating identities.

In the Bayesian Networks framework, it is straightforward to embed other source of information into the generative model. This is usually done by adding nodes into the network, and by arranging them thanks to the prior knowledge we may have on the correlations between the different represented random variables. Here, the focus will be made on adding holistic representation of the face, but also geometrical and color information.

5.2 Low-Resolution & Holistic Representation

It has been empirically shown [Sinha 02] that humans are partially relying on an holistic, low resolution representation of the face to perform face recognition. Actually, this capability is emphasized in the case of familiar faces, as can be seen in Figure 5.1. Moreover, conducted experiments in [Yip 02] show that subjects can achieve a 50% recognition rate at a resolution of only 7x10 pixels, and that the minimum resolution to reach good recognition is no more than 19x27 pixels. Such a result thus demonstrates that fine featural details are not necessary for human beings. In the case of computer-based systems, some algorithms are especially dedicated to face recognition at a low resolution (usually made to work at a distance, or from video sequences). Examples can be found in [Lee 06] and [Shakhnarovich 02], where faces of size 16x16 and 22x22 pixels are considered respectively. These studies suggest that holistic, low-resolution representation is valuable in the context of automatic recognition. Hence in this section we present a mean to incorporate such information into the first proposed model.

Fig. 5.1: Examples of familiar faces at low resolution [Sinha 06].

Adding an holistic representation is done by simply adding nodes in the network. In this particular case, the previous model is augmented with two nodes: one for the observation (node $\mathbf{O_h}$), and another hidden one corresponding to the cause of this observation (node **H**, which represents the "holistic type"). Figure 5.2 depicts the model incorporating the low resolution holistic representation.

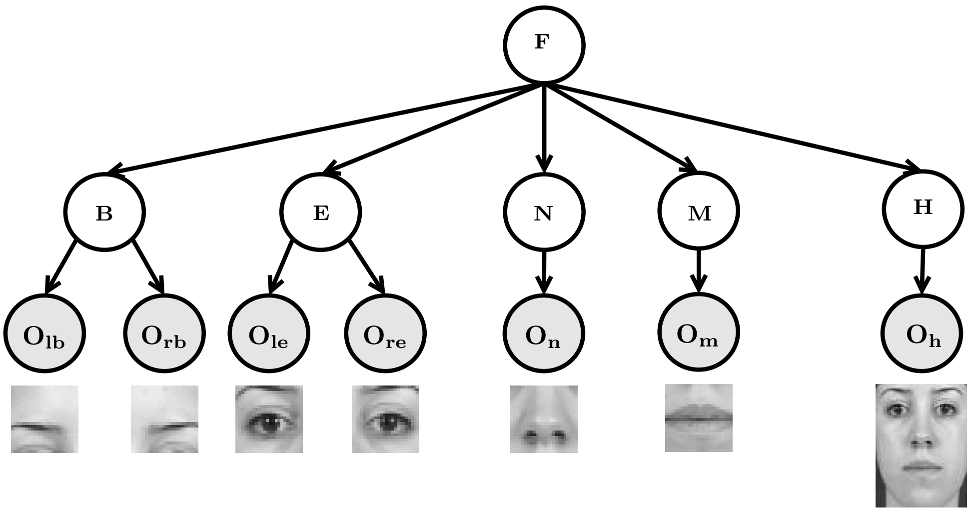


Fig. 5.2: Generative model where holistic information has been added.

5.2.1 Representing Holistic Information

To derive a feature vector containing the holistic representation, we first extract a subwindow containing the face. It is then subsampled to discard the featural details. Finally, the feature vector containing global appearance information is obtained by projecting the subsampled face window onto the Principal Component subspace. Note also that since external features have been shown to play a role in both human [Jarudi 03] and automatic [Lapedriza 06] face recognition, two different ways of extracting holistic information were investigated: we either extract only the inner face part containing mainly the facial features, or we extract the face in a way to include external features such as the ears and the chin for instance. Examples of these two face extractions can be seen in Figure 5.3.

(a) internal face (b) external face

Fig. 5.3: Example of the two different face extraction approaches.

5.2.2 Results & Discussion

The experimental framework is basically the same as in the previous cases, the only difference being on the feature vector representing low resolution information. In this case, the projection on the PCA subspace is made by retaining 90% of the variance (the resulting dimension is given in the tables). The HTER performance is presented on the different protocols of the BANCA database, and for different resolutions of the cropped face window. Table 5.1 show results for the holistic representation using the internal face and in Table 5.2 for the one with the external face.

System	Mc	Ua	Ud	P	G	dim
BNFace	**2.21**	20.93	**19.42**	**15.76**	**7.19**	NA
BNFace-LR-16x16	3.88	22.85	25.13	21.47	7.39	33
BNFace-LR-32x32	2.31	**20.77**	21.99	19.74	7.25	38
BNFace-LR-64x64	2.28	21.70	22.02	20.82	7.90	39

Tab. 5.1: HTER Performance on the test set $g2$ of BANCA with internal holistic information.

System	Mc	Ua	Ud	P	G	dim
BNFace	2.21	**20.93**	**19.42**	**15.76**	7.19	NA
BNFace-LR-16x16	**2.18**	22.92	24.90	20.98	**6.65**	40
BNFace-LR-32x32	4.33	21.25	25.19	20.57	8.02	49
BNFace-LR-64x64	3.46	22.88	26.89	21.34	7.25	50

Tab. 5.2: HTER Performance on the test set $g2$ of BANCA with external holistic information.

Results suggests that the holistic low resolution representation of the face as a complementary clue is not particularly useful for the face authentication task in our settings. In fact, the best performance with such information is really close to the baseline system. Moreover, this happens when there is a match between the training and testing conditions (protocol Mc and G). An interesting results though is that the best performance with holistic representation is attained when the lowest image resolution is used, supporting the results found in cognitive studies [Yip 02]. Regarding the difference between the usage of internal or external face representation, it is hard to clearly state which is the most suitable, since performances are more or less the same. It thus suggests that the two representations may carry useful information.

5.3 Geometrical Information

In the first proposed model, all geometric information is discarded, since the BNFace generative model only acts on the appearance of salient facial features. Unlike HMM-based approaches for instance, no constraints on the ordering of the observations are present. In human face recognition however, several studies suggest that facial features are processed in their context. Moreover, it was also shown that the global face shape plays an important role [Yip 02], and especially when the color information is not present, as it is the case in our first model. Besides, early automatic face recognition systems [Kanade 73] [Nixon 85] relied on measurements between facial features, hence such information also provide valuable information on the identity.

In our case, facial features have to be located by the ASM anyway. It could thus be interesting to exploit this result to add geometrical information into the generative process to model a face. Such observations could be incorporated either at the global face level or at the local facial features level. Global information includes face shape, face size or distance between facial features for instance, whereas local features measurements could be the eyes size or the mouth width. In this section, we thus propose new models incorporating structural information by using geometrical measurements as complementary clues to facial features appearance.

5.3.1 Encoding Structural Information

Both global and local information is considered here. These measurements are derived from the landmarks obtained thanks to the ASM. In order to be consistent across the different face images, landmarks are first rotated such that the eyes lies at the same height, then translated to a new origin given by the eyes

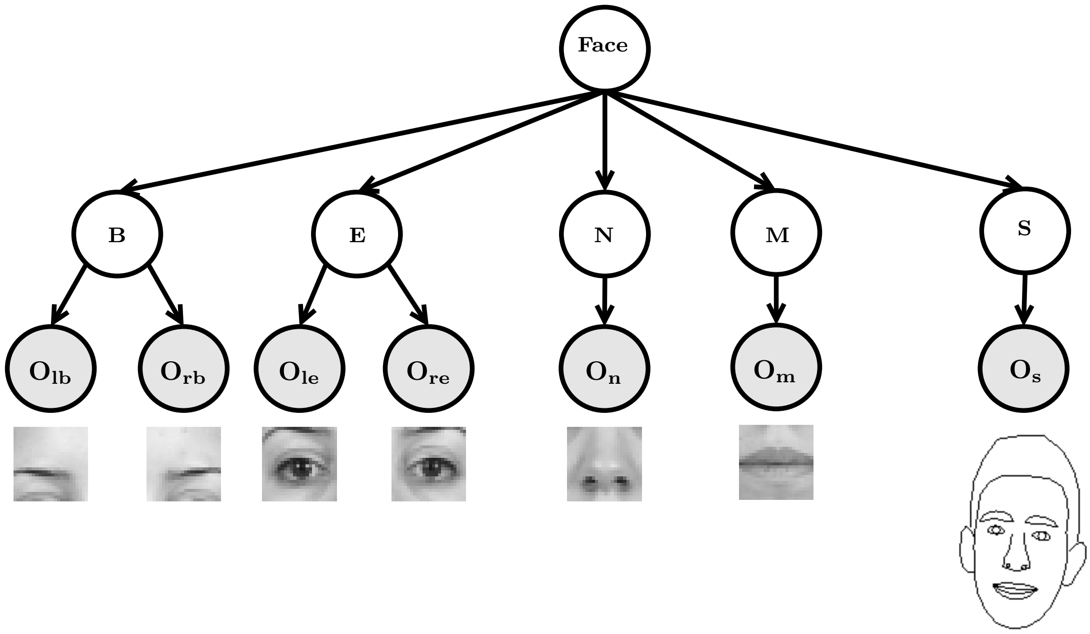


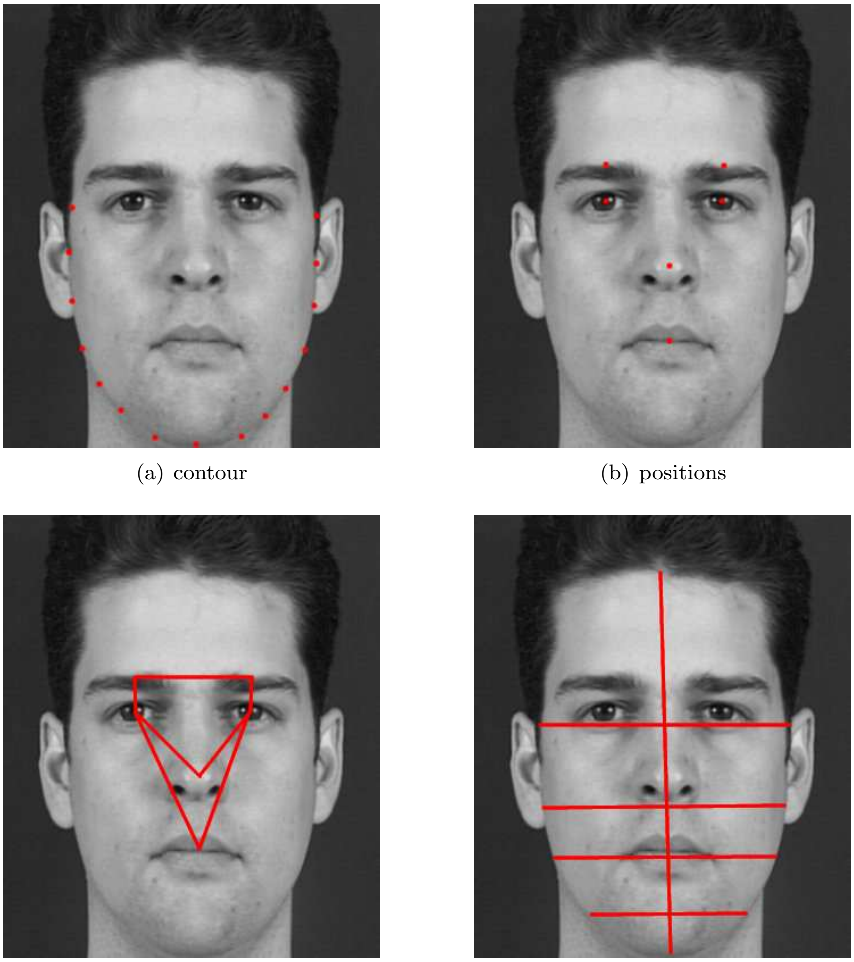
Fig. 5.4: The model to incorporate global geometric information.

center, and finally scaled so that the distance between the eyes remains constant.

Global Measurements In this case, we are interested to encode geometrical information at the global face level. An extra observation is thus incorporated in the network with two additional nodes: one for the observation (node $\mathbf{O_s}$), and another hidden one corresponding to the cause of this observation (node **S**). The model with global geometric information is depicted in Figure 5.4. Five different measures are considered to be used as global geometrical and shape/spatial context information:

- **All landmarks** Here we consider the (x, y) position for each landmark given by the ASM (see Figure 4.3). Note that in this case, PCA has been applied on the landmarks in order to decorrelate the data and reducing the dimensionality. The final feature vector is hence of dimension 25.

- **Contour landmarks** In this case, only the landmarks that correspond to the global contour of the face are considered and concatenated. It is meant to provide a representation of the face shape and consists in 15 points (Figure 5.5(a)).



(a) contour (b) positions

(c) distances (d) face size

Fig. 5.5: Different geometrical information at the global level.

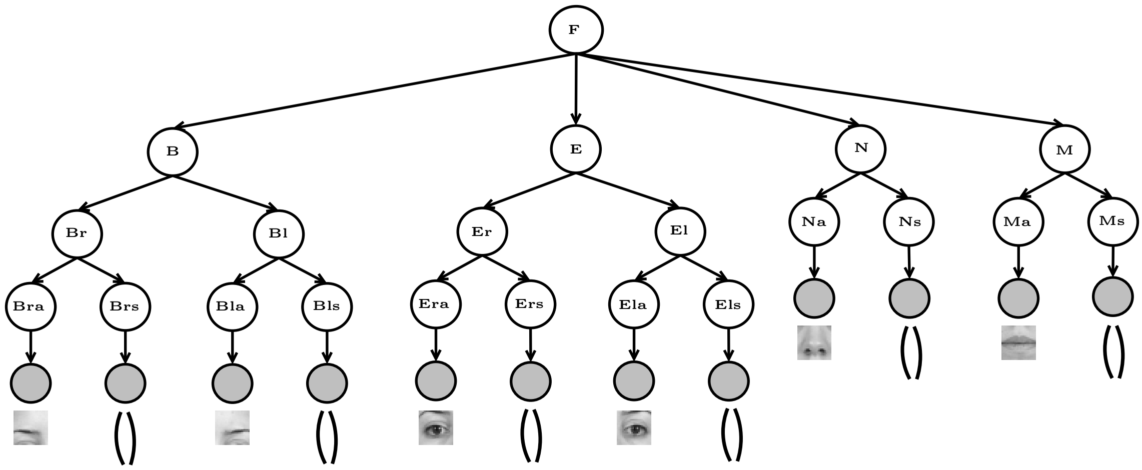


Fig. 5.6: The model to incorporate local geometric information.

- **Facial features positions** The landmarks corresponding to the extracted facial features are here considered. Hence, the position of the eyebrows, eyes, nose and mouth are concatenated to form the feature vector for geometrical observations (Figure 5.5(b)).

- **Distances between facial features** Configural information has been shown to be appropriate to various face analysis task, such as gender and race classification [Sun 06]. More precisely, it consists in the following measurements:

  - distances between eyebrows and eyes
  - distances between eyes and nose tip
  - distances between eyes and mouth corners
  - distances between eyebrows

  The resulting feature vector is obtained by concatenating these 7 measurements (Figure 5.5(c)).

- **Face size** The measurements describing the face size are obtained by recording the height of the face (from the tip of the chin to the beginning of the hairs) and also the width at different height levels: above and below the ears, at the mouth and below the mouth. The final feature vector then contains these 5 distances (Figure 5.5(c)).

Local Measurements We also considered geometric information at the local feature level. Our goal here is to explain a type of facial feature not only by its appearance but also by its position, its size or its shape. To do so, several nodes are added to embed this information, and now several observations are associated to each salient facial features. The corresponding model can be seen in Figure 5.6. Note that two additional layers of hidden nodes were introduced in this model. The first one aims at separating the left and right part of 'symmetric' observations (such as eyebrows and eyes, the separation is made through nodes **Br**, **Bl**, **Er** and **El**). This is done to incorporate geometrical information directly at the facial feature level: we choose to correlate the grayscale appearance of a particular facial feature with its geometrical information instead of correlating geometrical information of symmetric features together. The purpose of the second additional layer is to separate the appearance of the facial feature from its geometrical information. Note that now, the facial feature type does not depend on its appearance only, but on both its appearance and a geometrical information, such as its size for instance. In the local case, we considered 3 different measurements:

- **Local features position** The (x, y) position of each facial feature (see Figure 5.5(b)). The difference here is that the position information is not concatenated, but associated to the corresponding facial feature, as depicted in Figure 5.6.
- **Local features shape** Here, the landmark describing the contour of the facial feature are considered for each salient region (Figure 5.7(a)).
- **Local features size** The vertical and horizontal size of the feature, as depicted on Figure 5.7(b). Note that we do not consider the vertical size of the mouth, since it heavily depends on the expression.

5.3.2 Results & Discussion

In Table 5.3 and Table 5.4, HTER results for the different models and the different geometrical measurements are presented. As a first observation, one can see that embedding such information does not improve the performance much: in some cases, it even degrades it. However, it seems that some of the measurements carry valuable information. Actually, the best results on average are obtained when using sizes. At the global level, the model incorporating face size reaches almost identical performance

(a) local shape (b) local size

Fig. 5.7: Different geometrical information at the local feature level.

System	Mc	Ua	Ud	P	G
BNFace	**2.21**	20.93	19.42	15.76	7.19
BNFace-Global-All	9.49	27.34	28.62	23.21	**6.60**
BNFace-Global-Contour	4.04	22.79	**18.85**	17.93	9.93
BNFace-Global-Positions	2.82	22.21	19.36	16.06	7.20
BNFace-Global-Distance	3.11	**20.35**	19.04	15.93	6.76
BNFace-Global-Size	2.37	20.42	19.68	**15.45**	7.01

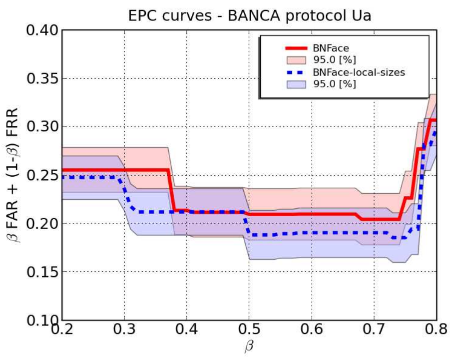
Tab. 5.3: HTER Performance on the test set $g2$ of BANCA for global geometrical information.

as the baseline model. However, it is quite difficult to state which global information is the most valuable, since results vary across the different protocols. At the local level, the model taking feature sizes into account performs better than the model using appearances of facial features only, but not significantly (Figure 5.8). It thus suggest that adding such information may be useful for recognizing identities.

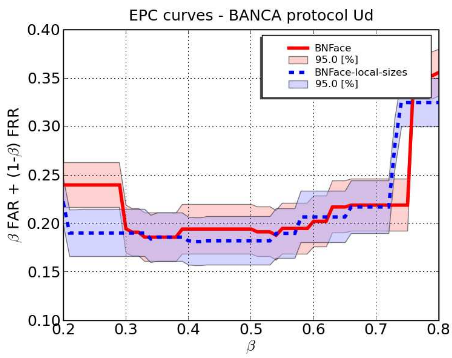
It must be noted that information based on geometrical measurements may be unstable, since it heavily depends on the results obtained by the ASM, which in turn depends on the pose and the expression. Moreover, information such as distances between facial features or facial features position may be severely affected when the pose of the face in the input image is not frontal.

System	Mc	Ua	Ud	P	G
BNFace	2.21	20.93	19.42	**15.76**	7.19
BNFace-Local-Shape	6.99	25.16	20.29	18.91	9.21
BNFace-Local-Position	4.46	22.50	20.16	17.61	6.96
BNFace-Local-Size	**1.92**	**18.69**	**19.01**	15.79	**5.84**

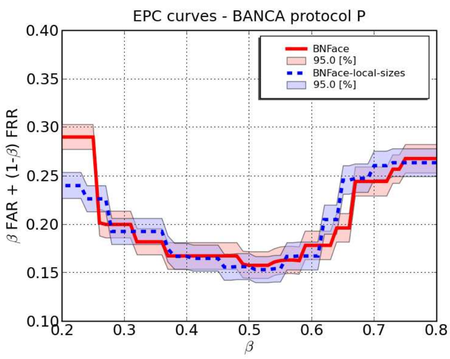
Tab. 5.4: HTER Performance on the test set $g2$ of BANCA for local geometrical information.



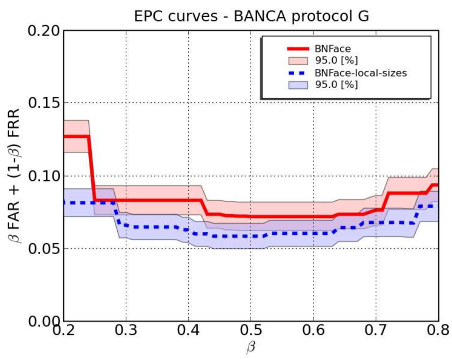
(a) Unmatched Adverse



(b) Unmatched Degraded



(c) Pooled Test



(d) Grand Test

Fig. 5.8: EPC curves for the test set $g2$ of BANCA: BNFace and BNFace-Local-Sizes.

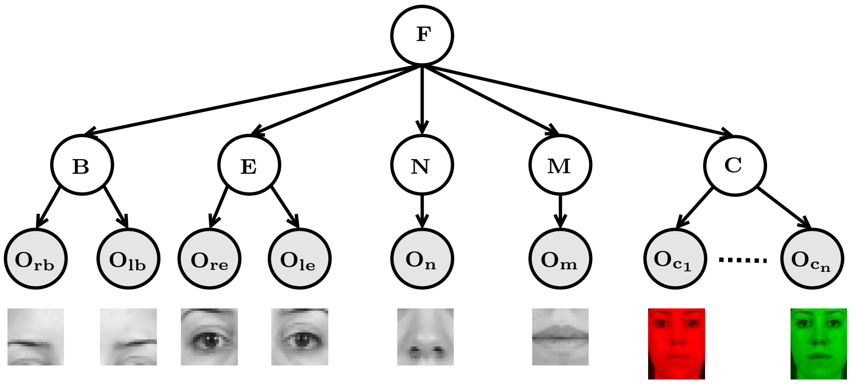


Fig. 5.9: The model to incorporate global color information.

5.4 Color Information

Color was also shown to play an important role in human face recognition [Russell 06] [Sinha 06]. It is thus likely that it may also carry useful information for computer-based systems. Surprisingly, only a few studies are using color in automatic face recognition systems. Torres et al. [Torres 99] developed a color Eigenfaces system, where Principal Component Analysis (PCA) is independently applied on each color channels and results are then combined for final classification. They showed that an improvement is obtained over traditional Eigenfaces acting on grayscale images [Turk 91]. This result was later confirmed in a study by Gutta et al. on the larger FERET database [Gutta 01]. Another interesting study is due to Sadeghi et al. [Sadeghi 07]: different channels from numerous colorspaces are first classified independently thanks to Linear Discriminant Analysis (LDA). An optimal subset of such classifiers is then found, and selected classification scores are combined using Support Vector Machines. Another approach proposed in [Jones 06] consists in extracting color features, and to use them as input to an Elastic Graph Matching algorithm. Color features were generally shown to perform better than grayscale-based ones.

Color at the Global Level The first proposed model is again a simple extension of the generative model proposed in Chapter 4 (Figure 5.9). In addition to the appearance of facial features, a 'type' of color is also modelled through node **C**. This hidden node *causes* observations derived in each color channel: it is hence assumed that information coming from different color channels are explicitly correlated. This color type then generates the corresponding observations extracted from the face image.

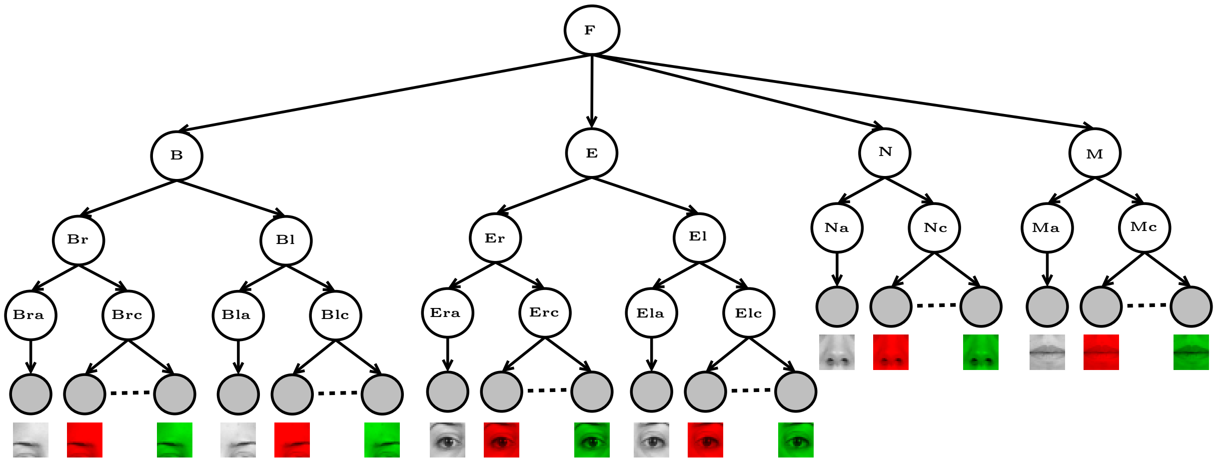


Fig. 5.10: The model to incorporate local color information.

Color at the Local Level To model the face more accurately, we also propose to incorporate color information at the local facial feature level. For this purpose, every type of facial feature is not only *explained* using grayscale appearance, but also with color information, as depicted in Figure 5.10. Again, two additional layers of hidden nodes were introduced in this model, as it was already the case with geometrical information. This is done for the same purposes: introducing color directly at the facial features level.

Color at Both Levels Color information at the global level is used to represent the global skin color of the face. On the other hand, color information at the local level aims at modelling the color of the associated facial feature. Hence, these two different observations may provide complementary information. As a consequence, we also propose a model combining both global and local information. This model consists in the local model depicted in Figure 5.10, where the branch of the global model (Figure 5.9) corresponding to global color information is added.

5.4.1 Encoding Color Features

Global Color Observations In this framework, our aim is to extract skin color information. Hence, a bounding box containing only the inner part of the face is cropped from the color image, based on eyes position. After being preprocessed by histogram equalization, it is subsampled to yield a low-resolution representation of the face, which discards details and thus mainly contains skin-colored pixels. Finally, feature vectors representing color are obtained by decomposing each color channel in terms of 2D-DCT.

Local Color Observations Regarding the color observations at the local feature level, the same windows as for the grayscale observations are cropped (using shifts as well), but from the original color image. Then, each extracted window is preprocessed by histogram equalization. Finally, feature vectors are obtained by decomposing each color channel in each window in terms of 2D-DCT.

Color information is encoded in the Hue-Saturation-Value (HSV) colorspace. Defining max (respectively min) as the maximum (resp. minimum) among the R, G and B values at a given pixel position, corresponding HSV values are obtained through a non-linear transformation of the RGB values:

$$\begin{aligned}
H &= \begin{cases} \text{undefined} & \text{if} \quad max = min \\ 60 \cdot \frac{G-B}{max-min} & \text{if} \quad max = R \text{ and } G \geq B \\ 60 \cdot \frac{G-B}{max-min} + 360 & \text{if} \quad max = R \text{ and } G \leq B \\ 60 \cdot \frac{G-B}{max-min} + 120 & \text{if} \quad max = G \\ 60 \cdot \frac{G-B}{max-min} + 240 & \text{if} \quad max = B \end{cases} \\
S &= \begin{cases} 0 & \text{if}\, max = 0 \\ 1 - \frac{min}{max} & \text{otherwise} \end{cases} \\
V &= max
\end{aligned} \tag{5.1}$$

Note that the HSV colorspace was already considered for various face-related task, such as face detection [Ikeda 03] and face recognition [Torres 99]. Moreover, this colorspace was shown to reach better performance than the RGB colorspace.

5.4.2 Results & Discussion

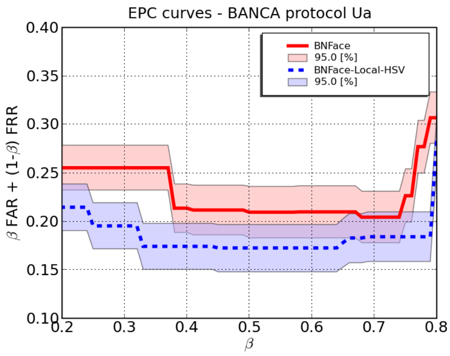
Results are reported using the three proposed models (referred to as BNFace-Global-Color, BNFace-Local-Color and BNFace-Both-Color), but also with two baseline algorithms. The first one is the first

System	Mc	Ua	Ud	P	G
BNFace	2.21	20.93	**19.42**	**15.76**	7.19
PCA-HSV	14.71	34.07	32.34	29.78	18.48
BNFace-Global-HSV	5.19	28.11	32.08	24.22	11.50
BNFace-Local-HSV	**1.89**	**17.24**	20.77	18.80	**5.58**
BNFace-Both-HSV	3.21	21.79	22.02	19.94	6.31

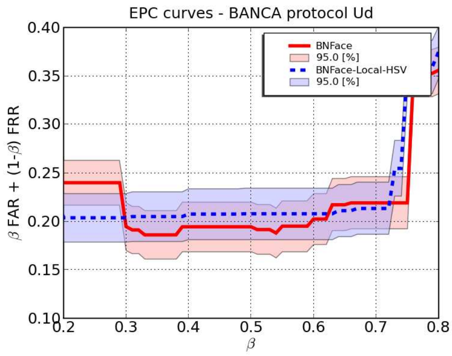
Tab. 5.5: HTER Performance on the test set $g2$ of BANCA for color information.

proposed model in Chapter 4. The second one is the color-based Eigenfaces (PCA-HSV) proposed in [Torres 99], and where color information from different channels is processed independently. Note that for the systems based on Bayesian Networks, the same experimental setup and the same learning parameters are again used here.

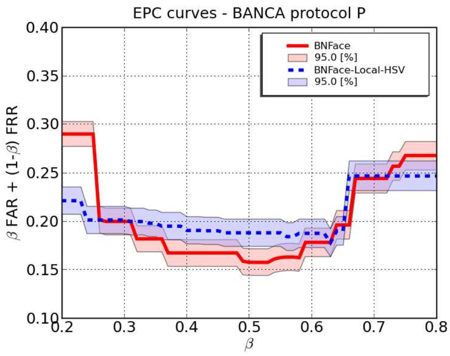
In Table 5.5, HTER performance for the different protocols is reported on the test set of the BANCA database. Obtained results show that when the acquisition conditions are not well-controlled, global skin color is not a reliable clue anymore. This is evidenced by the performance obtained with the Eigenfaces-based system, but also with generative models taking this information into account (i.e. BNFace-Global-Color and BNFace-Both-Color). Nevertheless, our results suggest that information derived from different color channels should not be treated independently, but correlated to each other, as evidenced by the performance of BNFace-Global-Color versus PCA-HSV. The baseline generative model based on grayscale features still performs better when there is a strong mismatch between training and testing conditions (protocols Ud and P). However, modelling the color at the local facial feature level achieves good results: when the training and testing conditions are the same (protocols Mc and G), local color information improves the performance, and outperforms both BNFace and the Eigenfaces-based systems. In Figure 5.11, we also present EPCs for the baseline system and the system using local color information. Although results are not clearly better than the baseline (as evidenced by the overlap between the confidence intervals), obtained performances with local color suggest that this information is indeed valuable. One can also see that local information is less affected by variations in acquisition conditions, as compared to global color information. Another interesting result is obtained with protocol Ua: in this case, even if the training/testing conditions are different, the model integrating color performs better than the baseline system. This can be explained by the fact that the same acquisition device was used in both scenarios, and thus color seems to remain consistent across controlled and adverse conditions (see Figure 2.3).



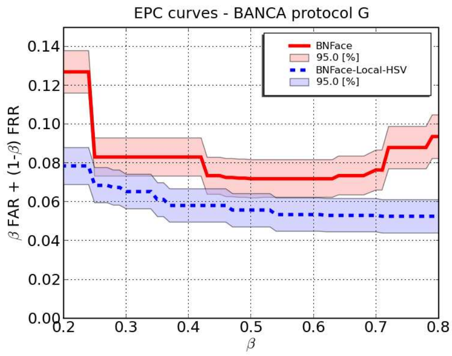
(a) Unmatched Adverse



(b) Unmatched Degraded



(c) Pooled Test



(d) Grand Test

Fig. 5.11: EPC curves for the test set $g2$ of BANCA: BNFace and BNFace-Local-HSV.

5.5 Summary

In this chapter we presented extensions of the first proposed model defined in Chapter 4. These new generative models were designed using results obtained from cognitive studies in human face recognition. They incorporate auxiliary information that are complementary to grayscale image patches, and that may help to describe the identity of an individual. Additional clues introduced into the model are an holistic representation of the face, geometrical information and color information.

Experiments conducted show that under some circumstances, adding auxiliary information may indeed help at recognizing identities. However and in our settings, holistic information showed no significant improvements over the baseline system. On the other hand, the use of geometrical information was shown to be helpful, provided that the face is more or less frontal. Embedding size information at the global face level and at the local features level yields better authentication performance. Color information was also shown to be helpful, but only at the local level and provided that the training and testing conditions match.

Chapter 6

CONCLUSION

6.1 Summary

This thesis dealt with the broad topic of Face Recognition, and more particularly aimed to tackle the face authentication problem. First, current existing approaches were briefly reviewed, and an emphasis was made on generative statistical models, since they have been proven to be successful to solve this task. However, currently used statistical models are not dedicated to handle information extracted from face data. Furthermore, such models have strong independence assumptions that are not in accordance with the way face data are generated. Indeed, generative models used so far are unable to capture correlations between information extracted from face images and does not model the fact that certain observations may be generated by the same cause.

We thus proposed new generative models, based on the Bayesian Networks framework, and especially dedicated to handle the face. The first proposed model relies on two main assumptions: first, salient facial features are supposed to be sufficient to discriminate individuals and second, such features should not be considered to be independent. Indeed, we stated that an identity is best described by a combination of different facial features, and as such, observations corresponding to salient facial features should be related. Moreover, we made the hypothesis that particular observations should be generated by the same underlying process, and going one step further, all observations extracted from a face image are generated by the same cause, which is the identity. Hence, we conducted experiments to verify that the proposed model indeed learns facial feature types, and their relationships, and we also compared our model to existing approaches in the context of face authentication. These experiments revealed the following results:

- The model is able to learn types of facial features, and the existing relationships between them. Actually, the model can be used to retrieve meaningful information about the face, since it is able to cluster people according to their facial feature types, such as a big nose for instance.

- The model is capable to perform recognition based on these pieces of information. This is evidenced by face authentication experiments where our model is shown to perform better than a generative model where facial features are treated independently. This is particularly true when the training and testing sets are acquired in different conditions.

- Our approach is competitive with respect to the state-of-the-art generative models previously used in face authentication, and far better than popular and widely-used appearance-based systems.

- In general, our model has a better generalization ability (as evidenced by mismatched conditions). This result could be explained by the fact that acting on certain features only, and not considering the whole face, is in fact more robust to little variations in acquisition conditions. Moreover, modelling relationships between observations seems to add valuable information in the authentication process, and especially when the the training and testing conditions do not match.

- The proposed models are by far less complex, in terms of the number of parameters, than the pseudo 2D-HMM. Therefore, our models is much more computationnaly efficient than the pseudo 2D-HMM, which is the current generative models obtaining the best performances.

We then extend the first approach with auxiliary information derived from the face image. This was motivated by recent results in cognitive research on human face recognition. As suggested by recent findings in this field, we augmented the first proposed model with various other information. This was done by incorporating additional clues at either the global face level or at the local facial feature level. In particular, we investigated the use of (1) low-resolution representation of the face, (2) geometrical measurements and (3) color information. Results obtained with these models showed that, under some circumstances, considering additional clues is useful at recognizing individuals. Experiments conducted show that under some circumstances, adding auxiliary information may indeed help at recognizing identities. However results obtained with these approaches should be carefully interpreted. Indeed, such pieces of information could be unstable, mainly depending on the image acquisition. Actually, the imaging device and the illumination conditions play a crucial role when using holistic and color information, whereas the face pose strongly affects geometrical measurements. Nevertheless, under relatively closely controlled conditions, adding additional sources of information is actually helpful.

Most importantly and what could be considered as the key result of this thesis work, is the introduction of a novel framework for the design of generative models dedicated to the face. It is actually now straightforward to design new models that are able to handle and relate different kind of observations extracted from face images.

6.2 Suggestions for Future Research

Since this work was the first, to the best of our knowledge, to propose the usage of Bayesian Networks as generative model for recognizing faces, future research directions are manifold. In the following, we thus list some suggestions for further investigations.

Observations & Network Structure In this work, we decided to use blocks around salient facial features as the primary source of observations to describe a face. Other auxiliary information were also embedded into network architecture and inspired by cognitive research. However, it is not known, as of today, which information is the most useful for face recognition (for both humans and automatic systems), and how to relate different pieces of information to each other.

Indeed, proposed models were designed according to our prior intuition. Actually, we do not know if relationships between the different processed sources of information truly exists, and how they should be encoded. There exists a large amount of research in structure learning (see for instance [De Campos 09] and references therein), and also some work towards an automatic discovery of causal relationships between random variables [Freedman 99] as well as hidden variables discovery [Elidan 00] [Elidan 04]. This field is however still an active research area, and current algorithms are not able to handle high-dimensional real-world data such as images.

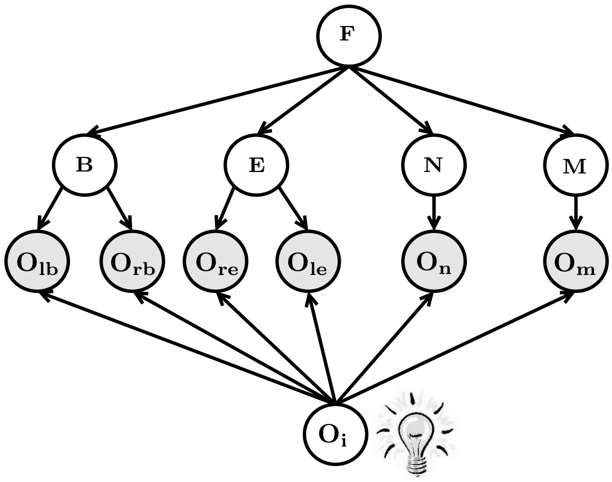


Fig. 6.1: Example of a model taking the illumination conditions into account. In this example, the node representing the illumination acts on the observations: their appearance is conditioned by the lighting conditions.

Continuous Hidden Variables The different proposed models in this thesis were all designed using discrete hidden variables. This is somewhat a limitation, since it is difficult to a priori decide on the number of different color types for instance. Hence, a natural extent would be to consider such hidden causes as continuous variables. However, doing so brings new challenges, and especially regarding numerical computations with high-dimensional data. Actually, experiments using such models (termed as Linear Gaussian Models) have been performed, but the numerous approximations made due to numerical issues are not well understood yet, and yield to poor face authentication results.

External Variations Another possible research direction is to embed external source of variations directly in the different models. For instance, pose and illuminations are known to influence the appearance of a particular feature. Hence, it is possible to add variables corresponding to such sources of variations. Figure 6.1 shows a possible example of a model taking illumination conditions into account. Actually, there already exists some studies taking quality measures into account, but they usually act on classification scores [Poh 09] and not on data directly. Hence, an interesting research direction would be the embedding of such measures directly in the learning and recognition process.

Multi-modal Biometric Verification Since in the Bayesian Networks framework, it is pretty easy to incorporate different sources of information in a principled way, such generative models should also be well-suited to handle multi-modal biometric measurements. Actually, an identity could be modelled using the correlations between different modalities, such as 3D face data, or iris images for instance. An example of a system that continuously asses the identity of a user using multiple biometrics can be found in [Muncaster 06].

Appendix A

MESSAGE PASSING

Section 3.2.2 defines the absorption procedure to pass a message between two nodes in a Junction Tree. Absorption induces several mathematical operations applied on potentials: marginalization, multiplication and division. Of course, the potentials and the associated operations are not the same depending on the distribution of the variables in the clusters. In this work, we thus present the absorption procedure for two types of distributions of the clusters potentials: discrete and linear gaussian. For this purpose, we will use the simple example depicted in Figure A.1.

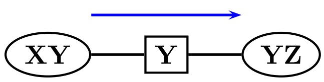


Fig. A.1: Absorption procedure between two clusters.

A.1 Discrete Case

In this case, potentials ϕ are simply represented as tables. The dimension of a potential is then given by the product of the cardinalities of the variables present in the cluster. Entries of the table are given by the result of a real-valued function on the configuration of the variables.

Consider for instance a cluster a with variable X and Y with cardinality I and J respectively. The dimension of the cluster potential $\phi(a)$ is thus given by $K = I \cdot J$, and the potential entry k is given by:

$$\phi(a)_k = P(X = i, Y = j) \tag{A.1}$$

Now consider the separator b containing the variable Y only, and a cluster c with variables Y and Z (as depicted in Figure A.1). The absorption procedure is in this case:

$$\phi^*(b) = \sum_X \phi(a) \tag{A.2}$$

$$\phi^*(c) = \phi(c) \cdot \left(\frac{\phi^*(b)}{\phi(b)}\right) \tag{A.3}$$

Equation A.2 consists simply in performing the marginalization over the nodes that are not present in b. Note also that the multiplication and the division in the Equation A.3 is done term-by-term.

A.2 Linear Gaussian Case

To derive the absorption procedure in this case we first have to define the conditional linear gaussian distribution, and the corresponding potentials obtained with this distribution. Consider a variable X with parent Y, the conditional linear gaussian distribution of X is given by:

$$P(X = x|Y = y) = \mathcal{N}(\mu_x + W \cdot y, \Sigma_x) \tag{A.4}$$

where $\mathcal{N}$ is the normal (gaussian) distribution. In its moment form, the multivariate normal distribution is given by:

$$\mathcal{N}(x, \mu, \Sigma) = \frac{1}{(2\pi)^{-\frac{n}{2}}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)) \tag{A.5}$$

where n is the dimension of the variable x, and the parameters of the distribution are (μ, Σ). This expression could be transformed to obtain the following canonical form:

$$\mathcal{N}(g, h, K) = \exp(g + h^t x - \frac{1}{2}x^t K x) \tag{A.6}$$

where the parameters (g, h, K) are given by:

$$g = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Sigma)) - \frac{1}{2}\mu^t K \mu. \tag{A.7}$$

$$h = \Sigma^{-1}\mu, \tag{A.8}$$

$$K = \Sigma^{-1}, \tag{A.9}$$

Conditional Gaussian Potential Recall that a potential is a function on the variables in the cluster and that one has to multiply factors from the original distribution to form the potential (Section 3.2.2). Hence we have to define the multiplication between conditional gaussian distribution. Note that before multiplying two distributions, we have to be sure that the dimension of each vector and matrix are the same. If not, the process of *extension* is applied by inserting 0s wherever its necessary. Now, the multiplication and the division operations are performed in the following way [Murphy 99a]:

1. Convert the distributions to canonical form

2. Extend them to the same domain

3. Multiplication and division are then defined by:

$$(g_1, h_1, K_1) \times (g_2, h_2, K_2) = (g_1 + g_2, h_1 + h_2, K_1 + K_2) \tag{A.10}$$

$$(g_1, h_1, K_1) \div (g_2, h_2, K_2) = (g_1 - g_2, h_1 - h_2, K_1 - K_2) \tag{A.11}$$

In order to be able to perform the absorption procedure and hence to update the cluster potentials, we also need to define the marginalization operation. Consider the example depicted in Figure A.1, to perform absorption, we first have to compute:

$$\phi(Y) = \int_X \phi(XY) \tag{A.12}$$

To do so, we first have to partition the h and K parameters into the components that will be kept and the one being marginalized over:

$$h = \begin{pmatrix} h_X \\ h_Y \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{pmatrix}$$

The new canonical parameters for the conditional gaussian potential $\phi(Y)$ are then given by:

$$g_Y = g + \frac{1}{2}\Big(d_1 \log(2\pi) - \log(|K_{XX}|) + h_X^t K_{XX}^{-1} h_X\Big) \tag{A.13}$$

$$h_Y = h_Y - K_{YX} K_{XX}^{-1} h_X \tag{A.14}$$

$$K_Y = K_{YY} - K_{YX} K_{XX}^{-1} K_{XY}. \tag{A.15}$$

Appendix B

PARAMETERS UPDATE RULES

This appendix derives the parameters update rules in the EM algorithm used to train the different models. In the face authentication framework, we first learn the parameters for the world model (see Section 2.2) using the EM algorithm with the Maximum Likelihood (ML) criterion, and then each client is adapted using Maximum A Posterior (MAP) training in the setting of model adaptation [1] [Gauvain 94]. The parameters are learned thanks to a training set $\mathcal{D} = \{\mathbf{v}_1, ..., \mathbf{v}_P\}$. Note that each example $\mathbf{v}_p$ of the training set actually represents a set of observations O_k (see Figure 4.1).

B.1 Maximum Likelihood Criterion - Discrete Case

In this section, we derive the update formula for the (conditional) probability distributions of the node in the Bayesian Networks representing the proposed models. Thanks to the approximation of Equation 3.9, each local distribution can be maximized independently. There are however three different cases in the proposed models, corresponding to three different probability distributions:

The Root In our case, the root node is a discrete random variable with cardinality J. Hence the update rule for the EM algorithm in Maximum Likelihood settings is given by [Bauer 97]:

$$P^{new}(X = i) = \frac{\sum_{p=1}^{P} P^{old}(X = i|\mathbf{v}_p)}{\sum_{p=1}^{P} \sum_{j=1}^{J} P^{old}(X = j|\mathbf{v}_p)} \tag{B.1}$$

[1] Note that in the linear conditional gaussian case with continuous hidden variables, we did not derive the formulas for the MAP criterion.

Hidden Variables Again, hidden variables in the proposed models are discrete, the only difference being the presence of a parent Y. Hence, the update rule is quite similar to the previous case and given by [Bauer 97]:

$$P^{new}(X = i|Y = k) = \frac{\sum_{p=1}^{P} P^{old}(X = i, Y = k|\mathbf{v}_p)}{\sum_{p=1}^{P} \sum_{j=1}^{J} P^{old}(X = j, Y = k|\mathbf{v}_P)} \tag{B.2}$$

Observed Variables In this case, the probability of the observation is given by a conditional gaussian (Equation 4.3), and is thus similar than the probability of a particular gaussian in a GMM, considering the parent node to be the mixture weight. Hence, the maximum likelihood parameter estimation for the mean and the variance are given by:

$$\mu_i^{new} = \frac{\sum_{p=1}^{P} P^{old}(X = i|O_x = o_x^p) \cdot o_x^p}{\sum_{p=1}^{P} P^{old}(X = i|O_x = o_x^p)} \tag{B.3}$$

$$\Sigma_i^{new} = \frac{\sum_{p=1}^{P} P^{old}(X = i|O_x = o_x^p) \cdot (o_x^p - \mu_i^{new})(o_x^p - \mu_i^{new})^T}{\sum_{p=1}^{P} P^{old}(X = i|O_x = o_x^p)} \tag{B.4}$$

B.2 Model Adaptation - Discrete Case

In the case of model adaptation [Gauvain 94], a prior is set on the parameters. In our case, this prior information will come from the world model trained using the ML criterion (since there is usually enough data to reach good ML estimates). Client models are then learned by an adaptation of the world model and using the few available client-specific training data. In practice, it was shown [Reynolds 00] [Cardinaux 05] that adapting only the means of the gaussian observations was sufficient to properly model client data. Hence, this strategy is also used here. Defining μ_i^{world} as the means of the world model, the MAP update rule for the means of the conditional gaussian variables for the client models is given by [Gauvain 94]:

$$\mu_i^{new} = \frac{\tau_i \mu_i^{world} + \sum_{p=1}^{P} P^{old}(X = i | O_x = o_x^p) \cdot o_x^p}{\tau_i + \sum_{p=1}^{P} P^{old}(X = i | O_x = o_x^p)} \tag{B.5}$$

where τ_i is a parameter of the prior Normal-Wishart density. Following the constraints made on the prior distribution ([Gauvain 94], Eqs. (47) and (48)) and the observations made in [Reynolds 00] that parameter-dependant adaptation coefficients only show a minor improvement, we can reduce Equation B.5 to:

$$\mu_i^{new} = \alpha \mu_i^{world} + (1 - \alpha) \frac{\sum_{p=1}^{P} P^{old}(X = i | O_x = o_x^p) \cdot o_x^p}{\sum_{p=1}^{P} P^{old}(X = i | O_x = o_x^p)} \tag{B.6}$$

Hence, we only use one extra hyperparameter α when training client models. α is referred to as the adaptation coefficient and reflects the importance made on the prior. Actually, α is used as a weighting factor between the prior distribution and the new obtained statistics, which may be unreliable due to the lack of training data.

B.3 Maximum Likelihood Criterion - Continuous Case

In linear gaussian models, all the variables in the network follow the same distribution, hence there are no distinct formulas for the parameters update rule, on the contrary to the previous model with discrete hidden variables and gaussian observations. Recall also that, thanks to Equation 3.9, the maximization can be done independently on the parameters for each variables.

To derive the update formulas for the variables in linear gaussian models, we will consider a node X having a single parent Y of dimension d_Y. The conditional gaussian distribution is given by:

$$p(X = x | Y = y) = (2\pi)^{-\frac{d_Y}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\big(x - (\mu + Wy)\big)^t \Sigma^{-1} \big(x - (\mu + Wy)\big)\Big) \tag{B.7}$$

Since there are hidden variables in our model, we will use the expected value E of the unobserved variables to get the log-likelihood of each distribution given the observations:

$$\mathcal{L}(\theta, \mathbf{v}) = -\frac{1}{2}\sum_{p=1}^{P} E\Big[\log(\det(\Sigma)) + (x_p - Wy_p - \mu)^t \Sigma^{-1}(x_p - Wy_p - \mu) \mid \mathbf{v}_p\Big], \tag{B.8}$$

The parameters of this distribution are the mean μ, the covariance matrix Σ and the regression matrix W. To simplify further notations, we define the following statistics:

$$S_{XX^t} = \sum_{p=1}^{P} E(XX^t|\mathbf{v}_p) \tag{B.9}$$

$$S_{X^tX} = \sum_{p=1}^{P} E(X^tX|\mathbf{v}_p) \tag{B.10}$$

$$S_X = \sum_{p=1}^{P} E(X|\mathbf{v}_p) \tag{B.11}$$

$$S_{YY^t} = \sum_{p=1}^{P} E(YY^t|\mathbf{v}_p) \tag{B.12}$$

$$S_Y = \sum_{p=1}^{P} E(Y|\mathbf{v}_p) \tag{B.13}$$

$$S_{YX^t} = \sum_{p=1}^{P} E(YX^t|\mathbf{v}_p) \tag{B.14}$$

$$S_{XY^t} = \sum_{p=1}^{P} E(XY^t|\mathbf{v}_p) \tag{B.15}$$

In order to find the update formulas, we simply derive the likelihood with respect to each of the parameters W, μ and Σ, and then set the derivative to zero.

Regression Matrix W

Using the following identity:

$$\frac{\partial\big((Xa+b)^t C(Xa+b)\big)}{\partial X} = (C + C^t)(Xa+b)a^t \tag{B.16}$$

and setting $X = -W, a = y_p, b = x_p - \mu, C = \Sigma^{-1}$, we have:

$$\begin{aligned}
\frac{\partial}{\partial W} L(\theta) &= \frac{1}{2} \sum_{p=1}^{P} 2 \cdot \Sigma^{-1} \cdot E\big[(x_p - W y_p - \mu) \cdot y_p^t\big] \\
&= \Sigma^{-1} \Big(\sum_{p=1}^{P} E(XY^t|\mathbf{v}_p) - W \sum_{p=1}^{P} E(YY^t|\mathbf{v}_p) - \mu \sum_{p=1}^{P} E(Y^t|\mathbf{v}_p) \Big)
\end{aligned}$$

Setting this expression equal to zero and using notations introduced earlier, we finally reach:

$$W = \big(S_{XY^t} - \mu \cdot S_Y^t\big) \cdot S_{XX^t}^{-1}. \tag{B.17}$$

Mean μ

Using the same identity (Equation B.16) and setting $X = \mu, a = -1, b = x_p - W y_p$ et $C = \Sigma^{-1}$, we have:

$$\begin{aligned}
\frac{\partial}{\partial \mu} L(\theta) &= \frac{1}{2} \sum_{p=1}^{P} 2 \cdot \Sigma^{-1} \cdot E\big[(x_p - W y_p - \mu)\big] \\
&= \Sigma^{-1} \Big(\sum_{p=1}^{P} E(X|\mathbf{v}_p) - W \sum_{p=1}^{P} E(Y|\mathbf{v}_i) - \mu \Big)
\end{aligned}$$

which yields to:

$$\mu_x = S_X - W \cdot S_Y. \tag{B.18}$$

Note that if the regression matrix is null, or does not exist, we have the classical formulation for the mean estimator.

Covariance Matrix Σ

To derive the update formula for the covariance matrix, we first need the following identities:

$$\begin{aligned}\frac{\partial \ln |X|}{\partial X} &= (X^t)^{-1} && \text{(B.19)}\\ \frac{\partial (a^t X b)}{\partial X} &= ab^t && \text{(B.20)}\end{aligned}$$

We then have:

$$\begin{aligned}\frac{\partial}{\partial \Sigma^{-1}} L(\theta) &= \frac{1}{2}\sum_{p=1}^{P}\Sigma - \frac{1}{2}\sum_{p=1}^{P} E\big[(x_p - Wy_p - \mu)(x_p - Wy_i - \mu)^t|\mathbf{v}_p\big]\\ &= \frac{P}{2}\Sigma - \frac{1}{2}\sum_{p=1}^{P}\big[x_p x_p^t - x_p y_p^t W^t - x_p\mu^t + Wy_p y_p^t W^t + Wy_p x_p^t + Wy_p\mu^t - \mu x_p^t + \mu y_p^t W + \mu\mu^t|\mathbf{v}_p\big]\end{aligned}$$

which yields to:

$$\hat{\Sigma} = \frac{1}{P}\Big(S_{XX^t} - S_{XY^t}\cdot W^t - S_X\cdot\mu^t + WS_{YY^t}W^t + W\cdot S_{YX^t} + \ WS_Y\mu^t - \mu S_X^t + \mu S_Y^t W^t\Big) + \mu\cdot\mu. \quad \text{(B.21)}$$

Appendix C

A Simple Yet Efficient Preprocessing Algorithm

C.1 Introduction

One of the major problem in face authentication systems is to deal with variations in illumination. In a realistic scenario, it is very likely that the lighting conditions of the probe image does not correspond to those of the gallery image, hence there is a need to handle such variations. Existing algorithms to cope with changes in lighting conditions can be divided into two main categories: model-based and preprocessing. In this appendix, we will focus on the preprocessing approach. For model-based approaches, interested readers may refer to [Georghiades 01] or [Basri 03].

The aim of such preprocessing algorithms is to minimize the effect of illumination variations prior to classification. Preprocessing has two main advantages: it is completely stand-alone and thus can be used with any classifier. Moreover, no training images are required to model illumination variations, since such algorithms turn the original image into another representation, which is supposed to be less sensitive to illumination variations. A study of preprocessing algorithms applied to face authentication [Short 04] showed the superiority of the model proposed by Gross & Brajovic [Gross 03]. However, this approach involves the resolution of a partial differential equation, which is either computationally expensive (brute force resolution) or approximate (iterative methods). Moreover, the algorithm depends on a normalization parameter λ, and hence face authentication cannot be performed in a fully automated way, since this parameter has to be manually set *a priori*. Another recent and simple algorithm was proposed in [Tan 07]: in this case, several operations are applied to the face image in order to yield an illumination-free representation. These operations are gamma correction, filtering with a difference of gaussians (DoG) filter and contrast equalization. Again, several parameters for the different steps have to be properly set.

Here, we propose to use Local Binary Patterns (LBP) [Ojala 02] as a preprocessing step prior to the face authentication task. The LBP is a nonparametric and computationnaly simple descriptor of local texture patterns. Moreover, it is invariant to monotonic grayscale transformation, hence the LBP representation may be less sensitive to changes in illumination.

C.2 Proposed Approach

The local binary pattern (LBP) is a non-parametric operator which describes the local spatial structure of an image. Ojala et al. [Ojala 02] first introduced this operator and showed its high discriminative power for texture classification. At a given pixel position (x_c, y_c), LBP is defined as an ordered set of binary comparisons of pixel intensities between the center pixel and its eight surrounding pixels (Figure C.1). The decimal form of the resulting 8-bit word (LBP code) can be expressed as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{C.1}$$

where i_c corresponds to the grey value of the center pixel (x_c, y_c), i_n to the grey values of the 8 surrounding pixels, and function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \,. \end{cases} \tag{C.2}$$

By definition, the LBP operator is unaffected by any monotonic gray-scale transformation which preserves the pixel intensity order in a local neighborhood. Note that each bit of the LBP code has the same

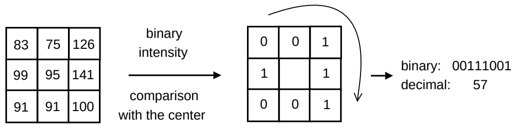


Fig. C.1: The LBP operator.

significance level and that two successive bit values may have a totally different meaning. Actually, The LBP code may be interpreted as a kernel structure index.

Later, Ojala et al. [Ojala 96] extended their original LBP operator to a circular neighborhood of different radius size. Their $LBP_{P,R}$ notation refers to P equally spaced pixels on a circle of radius R. In this work, we use the $LBP_{8,2}$ operator which is illustrated in Figure C.2.

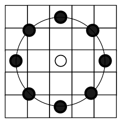


Fig. C.2: The extended LBP operator with (8,2) neighborhood.

Due to its texture discriminative property and its very low computational cost, LBP is becoming very popular in pattern recognition. Recently, LBP has been applied for instance to motion detection [Heikkilä 04], visual inspection [Turtinen 05], image retrieval [Takala 05], face detection [Jin 04] and face recognition [Zhang 05, Zhang 04, Ahonen 04]. In these previous works a face image was usually divided in small regions. For each region, a cumulative histogram of LBP codes, computed at each pixel location within the region, was used as a feature vector. However, in the present work, we do not use LBP to perform feature extraction, but to preprocess the input image: the face is represented with its texture patterns given by the LBP operator at each pixel location (Figure C.3).

C.3 Experiments and Results

In our case, the LBP operator will be used as a preprocessing step. The experimental framework is exactly the same as the one described in chapter 4. However, here we compare the different preprocessing algorithms. Indeed, each extracted block is preprocessed using an algorithm to cope with variations in

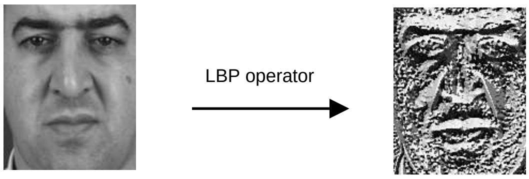


Fig. C.3: Original image (left) processed by the LBP operator (right).

illumination conditions. In particular, we investigate the use of:

- Gross & Brajovic Algorithm [Gross 03]
- Tan Algorithm [Tan 07]
- Our approach based on the LBP [Heusch 06]

Experiments are conducted on the BANCA database. Results for the Gross & Brajovic algorithm have been obtained using a value of $\lambda = 0.5$, whereas parameters for the Tan approach are $\gamma = 0.2$ for the gamma correction, $\sigma_1 = 1$ and $\sigma_2 = 2$ for the two gaussians of the DoG filter, $\alpha = 0.1$ and $\tau = 10$ in the contrast equalization operation (as suggested in [Tan 07]).

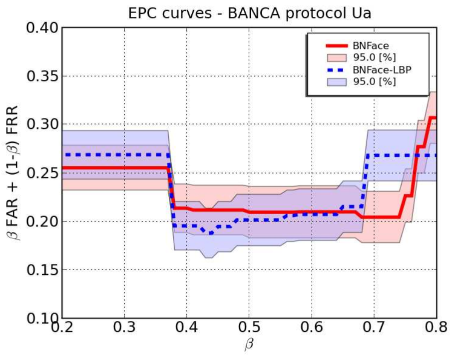
Table C.1 presents the HTER performance obtained on the test set $g2$ for different protocols of the BANCA database. It can be seen that the preprocessing applied to the blocks extracted from the face image has a huge impact on the face authentication performance. Results obtained with the two state-of-that-art results [Gross 03] [Tan 07] are somewhat astonishing, since they are degrading the result with respect to the simpler histogram equalization. However, these results could be explained because these normalization algorithms were applied on local image patches, and they were meant to process entire face images. Indeed, a similar behavior was observed in [Short 06].

Note that the proposed approach clearly outperforms the baseline histogram equalization in 3 out of 5 protocols. Actually, interesting results are obtained in the case of degraded conditions (protocol Ud): in this case, the LBP preprocessing significantly reduces the error rate as compared to histogram equalization. It thus suggest that this representation is particularly suitable when the acquisition conditions between the gallery and the probe sets are different. This is again evidenced by the results obtained on the protocol P. However, when the acquisition conditions are clean and the same for the training and testing sets (protocol Mc), results are comparable to that of histogram equalization. Note also that if these conditions match, but are not clean, the proposed preprocessing step is also valuable (protocol G) and is significantly better than all other preprocessing algorithms.

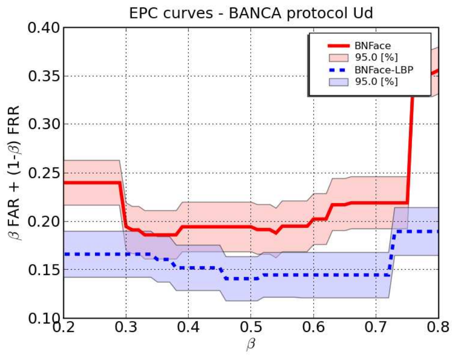
The proposed approach is thus a simple, computationnaly efficient and non-parametric preprocessing step that is able to cope with variation in acquisition conditions. It was shown to help at reducing the face authentication error rate by a significant amount, and performs better than more sophisticated illumination normalization algorithm. Note also that this preprocessing algorithm is usually computationnaly more efficient that other approaches, and also yield good performances with other face recognition systems, as shown in [Heusch 06].

System	Mc	Ua	Ud	P	G
BNFace	**2.21**	20.93	19.42	15.76	7.19
BNFace-Gross	7.47	28.88	39.55	26.03	12.48
BNFace-Tan	12.37	28.43	31.19	25.90	12.64
BNFace-LBP	**2.21**	**20.13**	**14.07**	**13.14**	**4.10**

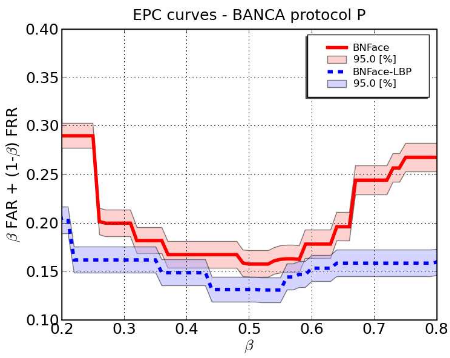
Tab. C.1: HTER performance on the test set $g2$ of BANCA with different preprocessing.



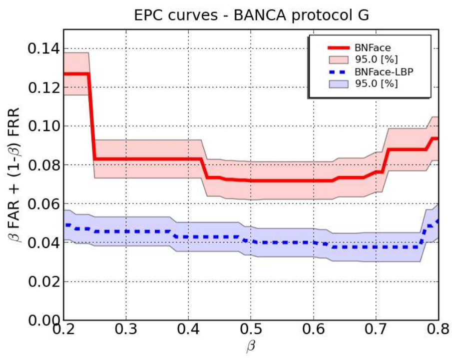
(a) Unmatched Adverse



(b) Unmatched Degraded



(c) Pooled Test



(d) Grand Test

Fig. C.4: EPC curves for the test set $g2$ of BANCA: BNFace and BNFace-LBP.

BIBLIOGRAPHY

[Acharya 05] T. Acharya & A.K. Ray. Image Processing: Principles and Applications. John Wiley and Sons, 2005.

[Adini 97] Y. Adini, Y. Moses & S. Ullman. *Face Recognition: The Problem of Compensating for Changes in Illumination Direction.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 721–732, 1997.

[Ahonen 04] T. Ahonen, A. Hadid & M. Pietikäinen. *Face Recognition With Local Binary Patterns.* In European Conference on Computer Vision (ECCV), pages 469–481. Springer, 2004.

[Ahonen 09] T. Ahonen & M. Pietikäinen. *Pixelwise Local Binary Pattern Models of Faces Using Kernel Density Estimation.* In Intl Conf. on Biometrics (ICB), pages 52–61, 2009.

[Bailly-Baillière 03] E. Bailly-Baillière*et al. The Banca Database and Evaluation Protocol.* In Intl Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA), pages 625–638, 2003.

[Bartlett 02] M. Bartlett, J. Movellan & T. Sejnowski. *Face Recognition by Independent Component Analysis.* IEEE Trans. on Neural Networks, vol. 13, no. 6, pages 1450–1464, 2002.

[Basri 03] R. Basri & D.W. Jacobs. *Lambertian Reflectance and Linear Subspaces.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pages 218–233, 2003.

[Bauer 97] E. Bauer, D. Koller & Y. Singer. *Update rules for parameter estimation in Bayesian networks.* In Uncertainty in Artificial Intelligence (UAI), pages 3–13, 1997.

[Belhumeur 97] P. Belhumeur, J. Hespanha & D. Kriegman. *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 711–720, 1997.

[Bengio 04] S. Bengio & J. Mariéthoz. *A Statistical Significance Test for Person Authentication.* In Proceedings of Odyssey: The Speaker and Language Recognition Workshop, pages 237–244, 2004.

[Bengio 05] S. Bengio, J. Mariéthoz & M. Keller. *The Expected Performance Curve.* In Intl Conf. On Machine Learning (ICML), 2005.

[Bicego 06] M. Bicego, A. Lagorio, E. Grosso & M. Tistarelli. *On The Use Of SIFT Features For Face Authentication.* In IEEE Intl. Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW), pages 35– 41, 2006.

[Cardinaux 03] F. Cardinaux, C. Sanderson & S. Marcel. *Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS.* In Intl Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA), pages 911–920, 2003.

[Cardinaux 04] F. Cardinaux, C. Sanderson & S. Bengio. *Face Verification Using Adapted Generative Models.* In IEEE Intl Conf on Automatic Face and Gesture Recognition (AFGR), pages 825–830, 2004.

[Cardinaux 05] F. Cardinaux, C. Sanderson & S. Bengio. *User Authentication via Adapted Statistical Models of Face Images.* IEEE Trans. on Signal Processing, vol. 54, no. 1, pages 361–373, 2005.

[Cohen 02] I. Cohen, N. Sebe, A. Garg, M.S. Lew & T.S. Huang. *Facial Expression Recognition From Video Sequences.* In IEEE Intl Conf. on Multimedia and Expo (ICME), volume 2, pages 121 – 124, 2002.

[Cootes 95] T.F. Cootes, C.J. Taylor, D. Cooper & J. Graham. *Active Shape Models: Their Training and Applications.* Computer Vision and Image Understanding, vol. 61, no. 1, pages 38–59, 1995.

[Cowell 99] G. Cowell, P. Dawid, L. Lauritzen & J. Spiegelhalter. Probabilistic Networks and Expert Systems. Springer, 1999.

[Cristinacce 04] D. Cristinacce, T. Cootes & I. Scott. *A Multi-Stage Approach to Facial Feature Detection.* In British Machine Vision Conference (BMVC), volume 1, pages 277–286, 2004.

[De Campos 09] C.P. De Campos, Z. Zheng & Q. Ji. *Structure Learning of Bayesian Networks Using Constraints.* In Intl Conf. On Machine Learning (ICML), pages 113–120, 2009.

[Dechter 96] R. Dechter. *Bucket Elimination: a Unifying Framework for Probabilistic Inference.* In Uncertainty in Artificial Intelligence (UAI), pages 211–219, 1996.

[Dempster 77] A. Dempster, N. Laird & D. Rubin. *Maximum Likelihood From Incomplete Data via the EM Algorithm.* The Journal of Royal Statistical Society, vol. 39, pages 1–37, 1977.

[Edwards 98] G. J. Edwards, T. F. Cootes & C. J. Taylor. *Face Recognition Using Active Appearance Models.* In European Conference On Computer Vision (ECCV), pages 581–595, 1998.

[Eickeler 00] S. Eickeler, S. Müller & G. Rigoll. *Recognition of JPEG Compressed Face Images Based on Statistical Methods.* Image and Vision Computing, vol. 18, no. 4, pages 279–287, 2000.

[Ekenel 05] H.K. Ekenel & B. Sankur. *Multiresolution Face Recognition.* Image and Vision Computing, vol. 23, no. 5, pages 469–477, 2005.

[Elidan 00] G. Elidan, N. Lotner, N. Friedman & D. Koller. *Discovering Hidden Variables: A Structure-Based Approach.* In Neural Information Processing Systems (NIPS), pages 479–485, 2000.

[Elidan 04] G. Elidan. *Learning Hidden Variables In Probabilistic Graphical Models.* PhD thesis, Hebrew University, 2004.

[Felzenszwalb 05] P. F. Felzenszwalb & D. P. Huttenlocher. *Pictorial Structures For Object Recognition.* International Journal of Computer Vision, vol. 61, no. 1, pages 55–79, 2005.

[Freedman 99] D. Freedman & P. Humphreys. *Are There Algorithms That Discover Causal Structure ?* Synthese, vol. 121, no. 1-2, pages 29–54, 1999.

[Fröba 04] B. Fröba & A. Ernst. *Face Detection With The Modified Census Transform.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 91–96, 2004.

[Gates 05] K.E. Gates. *Fast and Accurate Face Recognition Using Support Vector Machines.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), volume 3, page 163, 2005.

[Gauvain 94] J-L. Gauvain & C-H. Lee. *Maximum A Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains.* IEEE Trans. on Speech and Audio Processing, vol. 2, no. 2, pages 291–298, 1994.

[Georghiades 01] A. Georghiades, P. Belhumeur & D. Kriegman. *From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pages 643–660, 2001.

[Gross 03] R. Gross & V. Brajovic. *An Image Preprocessing Algorithm for Illumination Invariant Face Recognition.* In Intl Conf. Audio- and Video-based Biometric Person Authentication (AVBPA), pages 10–18, 2003.

[Gutta 01] S. Gutta, J. Huang, L. Chengjun & H. Wechsler. *Comparative Performance Evaluation of Gray-Scale and Color Information for Face Recognition Tasks.* In Intl Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA), pages 38–43, 2001.

[He 05] X. He, S. Yan, Y. Hu, N. Partha & H-J. Zhang. *Face Recognition using Laplacianfaces.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pages 328–340, 2005.

[Heckerman 99] D. Heckerman. Learning in Graphical Models, chapter A Tutorial on Learning With Bayesian Networks, pages 301–354. MIT Press, 1999.

[Heikkilä 04] M. Heikkilä, M. Pietikäinen & J. Heikkilä. *A Texture-based Method for Detecting Moving Objects.* In British Machine Vision Conference (BMVC), 2004.

[Heisele 03] B. Heisele, P. Ho, J. Wu & T. Poggio. *Face Recognition: Component-based Versus Global Approaches.* Computer Vision and Image Understanding, vol. 91, no. 1, pages 6–21, 2003.

[Heusch 06] G. Heusch, Y. Rodriguez & S. Marcel. *Local Binary Patterns as an Image Preprocessing for Face Authentication.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 9–14, 2006.

[Huang 96] C. Huang & A. Darwiche. *Inference in Belief Networks: A Procedural Guide.* International Journal of Approximate Reasoning, vol. 15, no. 3, pages 225–263, 1996.

[Huang 04] R. Huang, V. Pavlovic & D. Metaxas. *A Hybrid Face Recognition Method using Markov Random Fields.* In Intl Conf. on Pattern Recognition (ICPR), pages 157–160, 2004.

[Huang 05] X. Huang, S. Z. Li & Y. Wang. *Jensen-Shannon Boosting Learning For Object Recognition.* In Intl Conf. on Computer Vision and Pattern Recognition (CVPR), pages 144–149, 2005.

[Ikeda 03] O. Ikeda. *Segmentation of Faces in Video Footage using HSV Color for Face Detection and Image Retrieval.* In Intl Conf. on Image Processing (ICIP), pages 913–916, 2003.

[Jarudi 03] I. Jarudi & P. Sinha. *Internal and External Features and Face Recognition.* Technical report, M.I.T., 2003.

[Jensen 94] F. V. Jensen & F. Jensen. *Optimal Junction Trees.* In Uncertainty in Artificial Intelligence (UAI), pages 360–366, 1994.

[Jin 04] H. Jin, Q. Liu, H. Lu & X. Tong. *Face Detection Using Improved LBP under Bayesian Framework.* In Intl Conf. on Image and Graphics (ICIG), pages 306–309, 2004.

[Jobson 97a] D. J. Jobson, Z. Rahman & G. A. Woodell. *Properties and Performance of a Center/Surround Retinex.* IEEE Trans. On Image Processing, vol. 6, no. 3, pages 451–462, 1997.

[Jobson 97b] D.J. Jobson, Z. Rahman & G.A. Woodel. *A Multi-scale Retinex for Bridging the Gap between Color Images and the Human Observation of Scenes.* IEEE Transactions on Image Processing, vol. 6, no. 7, pages 965–976, 1997.

[Jones 06] C III. Jones & A. L. Abott. *Color Face Recognition by Hypercomplex Gabor Analysis.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 126–131, 2006.

[Jonsson 00] K. Jonsson, J. Matas, Y. P. Li & J. Kittler. *Learning Support Vectors for Face Verification and Recognition Biometrics and Benchmarking.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 208–213, 2000.

[Jordan 99a] M. I. Jordan, editor. Learning In Graphical Models. The MIT Press, 1999.

[Jordan 99b] M.I. Jordan, Z. Ghahramani, T.S. Jaakkolla & L.K. Saul. *An Introduction to Variational Methods for Graphical Models.* Machine Learning, vol. 37, pages 183–233, 1999.

[Kanade 73] T. Kanade. *Picture Processing by Computer Complex and Recognition of Human Faces.* Technical report, Dept. Information Science, Kyoto University, 1973.

[Keomany 06] J. Keomany & S. Marcel. *Active Shape Models Using Local Binary Patterns.* RR 06-07, IDIAP Research Institute, 2006.

[Kim 02] K-I. Kim, K. Jung & H-J. Kim. *Face Recognition Using Kernel Principal Component Analysis.* IEEE Signal Processing Letters, vol. 9, no. 2, pages 40–42, 2002.

[Kittler 98] J. Kittler, M Hatef, R.P.W. Duin & J. Matas. *On Combining Classifiers.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pages 226–239, 1998.

[Kittler 00] J. Kittler, Y. Li & J. Matas. *On Matching Scores for LDA-based Face Verification.* In British Machine Vision Conference (BMVC), pages 42–51, 2000.

[Kohir 98] V. Kohir & U. Desai. *Face Recognition Using A DCT-HMM Approach.* In IEEE Workshop on Applications of Computer Vision (WACV), pages 226–231, 1998.

[Lapedriza 06] A. Lapedriza, D. Masip & J. Vitria. *Face Verification using External Features.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 132–137, 2006.

[Lauritzen 88] S. L. Lauritzen & D. J. Speigelhalter. *Local Computations With Probabilities On Graphical Structures And Their Application To Expert Systems.* The Journal of Royal Statistical Society, vol. 50, pages 157–224, 1988.

[Lawrence 06] S. Lawrence, C. Giles Lee & A. C. Tsoi. *Convolutional Neural Networks for Face Recognition.* In Intl Conf. on Computer Vision and Pattern Recognition (CVPR), page 217, 2006.

[Lee 06] S.W. Lee, J. Park & S.W Lee. *Low Resolution Face Recognition Based on Support Vector Data Description.* Pattern Recognition, vol. 39, no. 9, pages 1809–1812, 2006.

[Liao 07] S. Liao, X. Zhu, Z. Lei, L. Zhang & S.Z. Li. *Learning Multi-scale Block Local Binary Patterns for Face Recognition.* In Intl Conf. on Biometrics (ICB), pages 828–837, 2007.

[Lucey 04] S. Lucey & 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.

[Lucey 06] S. Lucey & T. Chen. *Learning Patch Dependencies For Improved Pose Mismatched Face Verification.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), pages 909–915, 2006.

[Marcel 02] S. Marcel & S. Bengio. *Improving Face Verification Using Skin Color Information.* In Intl Conf. on Pattern Recognition (ICPR), volume 2, pages 378–381, 2002.

[Marcel 04] S. Marcel. *A Symmetric Transformation for LDA-based Face Verification.* In IEEE Intl. Conf on Automatic Face and Gesture Recognition (AFGR), pages 207–212, 2004.

[Martin 97] A. Martin, G. Doddington, T. Kamm, M. Ordowski & M. Przybocki. *The DET Curve in Assessment of Detection Task Performance.* In Eurospeech, pages 1895–1898, 1997.

[Martinez 98] A.M. Martinez & R. Benavente. *The AR Face Database.* Technical report, CVC Technical Report number 24, 1998.

[Martinez 99] A. Martinez. *Face Image Retrieval Using HMMs.* In IEEE Workshop on Content-Based Access of Image and Video Libraries, pages 35–39, 1999.

[Martinez 02] A. Martinez. *Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, pages 748–763, 2002.

[Messer 99] K. Messer, J. Matas, J. Kittler, J. Lüttin & G. Maitre. *XM2VTSDB: The Extended M2VTS Database.* In Intl Conf. Audio- and Video-based Biometric Person Authentication (AVBPA), pages 72–77, 1999.

[Messer 04] K. Messer*et al.* *Face Authentication Test on the BANCA Database.* In Intl Conf. on Pattern Recognition (ICPR), volume 4, pages 523–532, 2004.

[Muncaster 06] J. Muncaster & M. Turk. *Continuous Multimodal Authentication Using Dynamic Bayesian Networks.* In Workshop on Multimodal User Authentication, 2006.

[Murphy 99a] K. Murphy. *A Variational Approximation For Bayesian Networks With Discrete And Continuous Latent Variables.* In Uncertainty in Artificial Intelligence (UAI), pages 457–466, 1999.

[Murphy 99b] K. Murphy, Y. Weiss & M. Jordan. *Loopy Belief Propagation For Approximate Inference: An Empirical Study.* In Uncertainty in Artificial Intelligence (UAI), pages 467–475, 1999.

[Nefian 98] A. Nefian & 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.

[Nefian 00] A. Nefian & M. Hayes. *Maximum Likelihood Training of the Embedded HMM for Face Detection and Recognition.* In IEEE Intl Conf. on Image Processing (ICIP), volume 1, pages 33–36, 2000.

[Nefian 02] A. Nefian. *Embedded Bayesian Networks for Face Recognition.* In IEEE Intl Conf. on Multimedia and Expo (ICME), volume 2, pages 133–136, 2002.

[Nixon 85] M. Nixon. *Eye Spacing Measurement for Facial Recognition.* SPIE Digital Image Processing, vol. 575, pages 279–285, 1985.

[Ojala 96] T. Ojala, M. Pietikäinen & D. Harwood. *A Comparative Study of Texture Measures with Classification Based on Featured Distributions.* Pattern Recognition, vol. 29, no. 1, pages 51–59, 1996.

[Ojala 02] T. Ojala, M. Pietikäinen & T. Mäenpää. *Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pages 971–987, 2002.

[Othman 03] H. Othman & T. Aboulnasr. *A Separable Low Complexity 2D HMM with Application to Face Recognition.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pages 1229–1238, 2003.

[Pearl 88] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.

[Pentland 94] A. Pentland, B. Moghaddam & T. Starner. *View-based and Modular Eigenspaces for Face Recognition.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), pages 84–91, 1994.

[Phillips 99] P. J. Phillips. *Support Vector Machines Applied to Face Recognition.* In Neural Information Processing Systems (NIPS), pages 803–809, 1999.

[Phillips 00] P.J. Phillips, H. Moon, S.A. Rizvi & P.J. Rauss. *The FERET Evaluation Methodology for Face-Recognition Algorithms.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pages 1090–1104, 2000.

[Phillips 05] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min & W. Worek. *Overview of the Face Recognition Grand Challenge.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 947–954, 2005.

[Poh 09] N. Poh, S.Y. Wong, J. Kittler & F. Roli. *Challenges and Research Directions for Adaptive Biometric Recognition Systems.* In Intl Conf. on Biometrics (ICB), pages 753–764, 2009.

[Prince 06] S. J. Prince & J. Elder. *Tied Factor Analysis for Face Recognition Across Large Pose Changes.* In British Machine Vision Conference (BMVC), volume 889-898, 2006.

[Rajapakse 04] M. Rajapakse, J. Tan & J. Rajapakse. *Color Channel Encoding With NMF for Face Recognition.* In IEEE Intl Conf. on Image Processing (ICIP), volume 3, pages 2007–2010, 2004.

[Reynolds 00] D.A. Reynolds, T.F. Quateri & R.B. Dunn. *Speaker Verification Using Adapted Gaussian Mixture Models.* Digital Signal Processing, vol. 10, pages 19–41, 2000.

[Rodriguez 06a] Y. Rodriguez, F. Cardinaux, S. Bengio & J. Mariéthoz. *Measuring The Performance of Face Localization Systems.* Image and Vision Computing, vol. 24, no. 8, pages 882–893, 2006.

[Rodriguez 06b] Y. Rodriguez & S. Marcel. *Face Authentication Using Adapted Local Binary Pattern Histograms.* In European Conference on Computer Vision (ECCV), pages 321–332, 2006.

[Russell 06] R. Russell, P. Sinha, I. Biedermann & M. Nederhouser. *Is Pigmentation Important For Face Recognition ? Evidence From Contrast Negation.* Perception, vol. 35, pages 749–759, 2006.

[Sadeghi 07] M.T. Sadeghi, S. Khoshrou & J. Kittler. *SVM-Based Selection of Colour Space Experts for Face Authentication.* In Intl Conf. on Biometrics (ICB), pages 907–916, 2007.

[Samaria 94a] F. Samaria & A. Harter. *Parametrisation of a Stochastic Model for Human Face Identification.* In IEEE Workshop on Applications of Computer Vision (WACV), 1994.

[Samaria 94b] F. Samaria & S. Young. *HMM-based Architecture for Face Identification.* Image and Vision Computing, vol. 12, no. 8, pages 537–543, October 1994.

[Sanderson 03] C. Sanderson & K. Paliwal. *Fast Features for Face Authentication Under Illumination Direction Changes.* Pattern Recognition Letters, vol. 24, no. 14, pages 2409–2419, 2003.

[Sanderson 05] C. Sanderson, M. Saban & G. Yongsheng. *On Local Features For GMM based Face Verification.* In Intl Conf. on Information Technology and Applications, pages 650–655, 2005.

[Schwaninger 02] A. Schwaninger, S.M. Collishaw & J. Lobmaier. *Role And Interaction of Featural and Configural Processing in Face Recognition.* Journal of Vision, vol. 2, no. 7, page 602, 2002.

[Shakhnarovich 02] G. Shakhnarovich, J. W. Fisher & T. Darrell. *Face Recognition From Long-term Observations.* In European Conference on Computer Vision (ECCV), pages 851–868, 2002.

[Shen 07] L. Shen, L. Bai & M. Fairhust. *Gabor Wavelets and General Discriminant Analysis for Face Identification and Verification.* Image and Vision Computing, vol. 25, no. 5, pages 553–563, 2007.

[Short 04] J. Short, J. Kittler & K. Messer. *A Comparison of Photometric Normalisation Algorithms for Face Verification.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 254–259, 2004.

[Short 06] J. Short, J. Kittler & K. Messer. *Photometric Normalisation for Component-based Face Verification.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 119–125, 2006.

[Sim 01] T. Sim, S. Baker & M. Bsat. *The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces.* Technical report, Carnegie Mellon University, 2001.

[Sinha 02] P. Sinha. *Identifying Perceptually Significant Features For Recognizing Faces.* Proceedings of SPIE - Human Vision and Electronic Imaging VII, vol. 4662, pages 12–21, 2002.

[Sinha 06] P. Sinha, B. Balas, Y. Ostrovsky & R. Russel. *Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About.* Proceedings of the IEEE, Special Issue on Biometrics: Algorithms and Applications, vol. 94, no. 11, pages 1948–1962, 2006.

[Stephenson 00] T. Stephenson. *An Introduction to Bayesian Networks Theory and Usage.* Technical Report 00-03, Idiap Research Institute, 2000.

[Sun 06] Y. Sun, N. Sebe, T. Gevers & M. Mercera. *The Role of Featural and Configural Information in Face Classification A Simulation of the Expertise Hypothesis.* In Intl Conf. on Pattern Recognition (ICPR), volume 1, pages 1166–1170, 2006.

[Takala 05] V. Takala, T. Ahonen & M. Pietikäinen. *Block-based Methods for Image Retrieval using Local Binary Patterns.* In 14th Scandinavian Conference on Image Analysis (SCIA), pages 882–891, 2005.

[Tan 07] X. Tan & B. Triggs. *Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions.* In IEEE Intl Workshop on Analysis and Modeling of Faces and Gestures (AMFG), pages 168–182, 2007.

[Torres 99] L. Torres, J. Y. Reutter & L. Lorente. *The Importance of the Color Information in Face Recognition.* In IEEE Intl Conf. on Image Processing (ICIP), volume 3, pages 627–631, 1999.

[Turk 91] M. Turk & A. Pentland. *Face Recognition Using Eigenfaces.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), pages 586–591, 1991.

[Turtinen 05] M. Turtinen, M. Pietikäinen & O. Silven. *Visual Characterization of Paper using Isomap and Local Binary Patterns.* In Conference on Machine Vision Applications (MVA), pages 210–213, 2005.

[Viola 01] P. Viola & M. Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features.* In IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR), pages 511–518, 2001.

[Wiskott 99] L. Wiskott, J-M. Fellous, N. Krüger & C. Von Der Malsburg. *Face Recognition By Elastic Bunch Graph Matching.* In L. C. Jain, U. Halici, I. Hayashi & S. B. Lee, editors, Intelligent Biometric Techniques in Fingerprint and Face Recognition, chapter 11, pages 355–396. CRC Press, 1999.

[Yang 00] M-H. Yang, N. Ahuja & D. Kriegman. *Face Recognition Using Kernel Eigenfaces.* In IEEE Intl Conf. on Image Processing (ICIP), volume 1, pages 37–40, 2000.

[Yang 02] M-H. Yang. *Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods.* In IEEE Intl Conf. on Automatic Face and Gesture Recognition (AFGR), pages 205–211, 2002.

[Yip 02] A. Yip & P. Sinha. *Role Of Color in Face Recognition.* Journal Of Vision, vol. 2, no. 7, page 596, 2002.

[Yow 97] K. Yow & R. Cipolla. *Feature-based Human Face Detection.* Image and Vision Computing, vol. 15, no. 9, pages 713–735, 1997.

[Zhang D. 05] Zhang D., Songcan C. & Zhou. Z-H. *Two-Dimensional Non-negative Matrix Factorization for Face Representation and Recognition.* In IEEE Intl Workshop on Analysis and Modeling of Faces and Gestures (AMFG), pages 350–363, 2005.

[Zhang 04] G. Zhang, X. Huang, S.Z. Li, Y. Wang & X. Wu. *Boosting Local Binary Pattern (LBP)-based Face Recognition.* In Chinese Conference on Biometric Recognition (Sinobiometrics), pages 179–186, 2004.

[Zhang 05] W. Zhang, S. Shan, W. Gao, X. Chen & H. Zhang. *Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a Novel Non-statistical Model for Face Representation and Recognition.* In IEEE Intl. Conf. on Computer Vision (ICCV), 2005.

[Zhao 03] W. Zhao, R. Chellappa, P.J. Phillips & A. Rosenfeld. *Face Recognition: A Literature Survey.* ACM Computing Surveys, vol. 35, no. 4, pages 399–458, 2003.

Appendix D

Curriculum Vitae

Guillaume Heusch

Rue Edouard Payot 2

1005 Lausanne

Switzerland

guillaume.heusch@gmail.com

Swiss & French citizenship

31 years old

Single, no children

Education

2009 PhD in Computer Science, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Thesis: *Bayesian Networks as Generative Models for Face Recognition.*

2005 MSc in Communication Systems, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Thesis: *Lighting Normalization for Face Verification.*

Professional Experience

2005 – 2009 **Research Assistant, Idiap Research Institute, Martigny, Switzerland**
Carry out research in the field of face recognition. The aim of my work was to design, implement and test new statistical models inspired by human face recognition abilities. Proposed systems reach comparable performance to state-of-the-art systems while being more computationnaly efficient.

2008 **Intern, Siemens Corporate Research, Princeton NJ, USA**

Research work on automatic detection of particular anatomical structures in ultrasound images for medical diagnosis purposes. Strengthen collaboration with external clinician partners by presenting and explaining the project. This work resulted in an invention disclosure and a grant proposal.

2004 **Intern, Idiap Research Institute, Martigny, Switzerland**

Implemented several lighting normalization algorithms, study their effect on a face verification system, and finally reduced the error rate by 34%.

2002 **Intern, Visiowave (now G.E Security), Ecublens, Switzerland**

Realization of a program able to automatically detect stopped vehicles in highways surveillance video sequences, and then sending an alarm to the operator.

Skills

Competencies: Machine Learning, Pattern Recognition, Computer Vision, Image Processing.

Computer Programming: C, C++, Perl, Matlab, Java, PHP, SQL.

Languages: French (mother tongue), English (good knowledge), German (basics).

Extracurricular Activities

Rugby: former captain of the university team, former president, board member.

Sports: hiking, running, swimming and skiing.

Travels: USA, Carribean, Thailand, Laos, Madagascar, South Korea, European cities.

Publications

Journal Papers

* G. Heusch and S. Marcel, *A Novel Statistical Generative Model Dedicated to Face Recognition*, in press, Image & Vision Computing, Volume 28, Issue 1, pp.101-110, 2010.

Conference Proceedings

* G. Heusch and S. Marcel, *Bayesian Networks to Combine Intensity and Color Information in Face Recognition*, Intl Conf. on Biometrics (ICB), pp. 414-423, 2009.

* G. Heusch and S. Marcel, *Face Authentication with Salient Local Features and Static Bayesian Network*, Intl Conf. on Biometrics (ICB), pp. 878-887, 2007.

* N. Poh, G. Heusch and J. Kittler, *On Combination of Face Authentication Experts by a Mixture of Quality Dependent Fusion Classifiers*, 7th Workshop on Multiple Classifier Systems (MCS), pp. 344-356, 2007.

* G. Heusch, Y. Rodriguez and S. Marcel, *Local Binary Patterns as an Image Preprocessing for Face Authentication*, Intl Conf. on Face and Gesture Recognition (FGR), pp. 9-14, 2006.

* K. Messer et al. *Performance Characterisation of Face Recognition Algorithms and Their Sensitivity to Severe Illumination Changes*, Intl Conf. on Biometrics (ICB), pp. 1-11, 2006.

Technical Reports

* S. Marcel, Y. Rodriguez and G. Heusch, *On the Recent Use of Local Binary Patterns for Face Authentication*, Idiap RR 06-34, 2006.

* G. Heusch, F. Cardinaux and S. Marcel, *Efficient Diffusion-based Illumination Normalization for Face Verification*, Idiap RR 05-46, 2005.