

Bayesian Networks to Combine Intensity and Color Information in Face Recognition

Guillaume Heusch^{1,2} and Sébastien Marcel¹

¹ Idiap Research Institute, Centre du Parc, rue Marconi 19, P.O. Box 592, 1920 Martigny, Switzerland

² Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
{guillaume.heusch,sebastien.marcel}@idiap.ch

Abstract. We present generative models dedicated to face recognition. Our models consider data extracted from color face images and use Bayesian Networks to model relationships between different observations derived from a single face. Specifically, the use of color as a complementary observation to local, grayscale-based features is investigated. This is done by means of new generative models, combining color and grayscale information in a principled way. Color is either incorporated at the global face level, at the local facial feature level, or at both levels. Experiments on the face authentication task are conducted on two benchmark databases, XM2VTS and BANCA. Obtained results show that integrating color in an intelligent manner improves the performance over a similar baseline system acting on grayscale only, but also over an Eigenfaces-based system where information from different color channels are treated independently.

1 Introduction

Face recognition is an active research area, probably because of its numerous applications, ranging from video surveillance to human-computer interaction for instance. Hence, there exists numerous systems allowing to recognize people based on their face image. The vast majority of such existing approaches typically act on grayscale images only, since color is usually considered to introduce high variability. Nevertheless, it was shown that color plays an important role in human face recognition [1] [2]. 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. [3] 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 [4]. This result was later confirmed in a study by Gutta et al. on the larger FERET database [5]. Another interesting study is due to Sadeghi et al. [6]: different channels from numerous colorspace 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 [7] consists in extracting color features to use them as input to an Elastic Graph Matching algorithm. Again, color features were shown to perform better than grayscale-based ones.

Bayesian Networks provide an elegant framework to describe relationships (and hence correlations) between different pieces of information. In this work, our aim is to derive models describing the process that generates observations of different nature extracted from face images. Such generative models will then be used for recognition purposes. Specific generative models for face recognition were recently proposed by Heusch and Marcel [8]. The authors proposed a tree-structured Bayesian Network to describe data extracted from grayscale face images. More precisely, they assumed that observations derived from salient facial features are related to each other, and hence tried to model correlations between such observations. Going one step further, we believe that such models are suitable to model correlations between local, grayscale-based features and other information, such as color. Hence, in this contribution, new models integrating color at the global face level, the local facial feature level but also at both levels are derived. Experimental evaluation on face authentication is carried out on the XM2VTS [9] and the BANCA [10] databases. Results show that integrating color intelligently into dedicated generative models may help at reducing the authentication error rate, at least when the training and testing acquisition conditions are quite similar. The best proposed model, correlating grayscale and color at both the local and the global level show a significant improvement in performance as compared to a similar baseline model acting on the luminance channel only [8], but also performs better than the Eigenfaces-based system acting on color channels independently [3]. Obtained results thus suggest that color is of valuable information when combined to grayscale in a coherent manner.

In the next section, Bayesian Networks are briefly introduced, before describing the proposed models and the features in Section 3. Section 4 details the experimental framework and section 5 describes the databases and discusses the obtained results. Finally, a conclusion is drawn in Section 6.

2 Bayesian Networks

A Bayesian Network is a probabilistic graphical model representing the joint probability distribution over a set of random variables, and having the ability to encode dependencies among these variables [11]. It is specified as a directed acyclic graph, where nodes represent random variables and directed links represent causal relationships between these variables. Defining a set of random variables $\mathbf{U} = (x_1, \dots, x_n)$, the joint probability defined by a Bayesian Network is given by the following chain rule:

$$P(\mathbf{U}) = \prod_{i=1}^n P(x_i | \text{Parents}(x_i)) \quad (1)$$

where $\text{Parents}(x_i)$ denotes the set of parents of the node representing the variable x_i . Hence, a Bayesian Network is fully defined by the structure of the

graph, and by its parameters: the conditional distribution of each variable given its parents. Computing probabilities in the network is referred to as *inference*. It is typically done to update the state knowledge of a set of hidden variables when other variables, referred to as *evidence*, have been observed. In our case, inference is carried out thanks to the Junction Tree Algorithm [12]. Another important issue is how to learn such models from data. Learning in Bayesian Networks may either refer to structure learning, parameters learning or both. Since in this work, the structure is derived according to prior domain knowledge, the focus is made on learning the parameters. As the proposed models contain hidden variables, a natural choice is the Expectation-Maximization (EM) algorithm [13].

3 Proposed Models

3.1 Reference Model

The model presented in [8] relies on the assumption that facial features are related to each other. To model such relationships, a tree-structured Bayesian Network is proposed. This generative model assumes that there exists pairwise relationships between observations derived from grayscale images. Namely, relationships between eyebrows and eyes, eyes and nose and nose and mouth are considered. This model performs better than a simpler generative model (based on Gaussian Mixtures Model) where independence between facial features is assumed. However, it only acts on local features derived from intensity images and does not take advantage of the Bayesian Networks framework to integrate other source of information in a smart manner.

3.2 Color at the Global Level

The first proposed model is depicted in Figure 1 and should be understood as follow: the root node is used to relate various information describing the face. A face thus consists in a relationship between different 'types' of facial features (nodes **B**: eyebrows, **E**: eyes, **N**: nose and **M**: mouth). In addition, 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. Finally, the different types of facial feature, as well as the type of color, generates the corresponding observations extracted from the face image. Note also that, unlike the reference model [8], a single hidden node is used to model the relationship between the different observations. Actually, global color information has to be related to the whole face rather than to pairwise relationships between facial features.

3.3 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

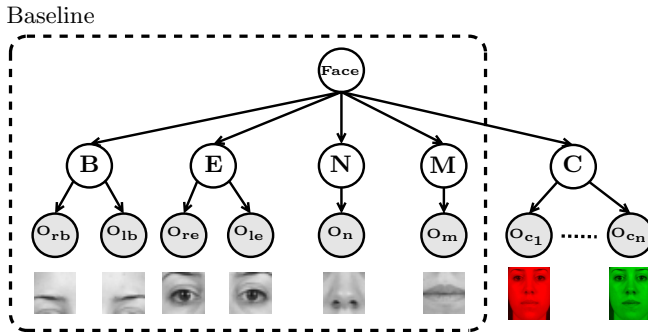


Fig. 1. Bayesian Network model for the face incorporating color information at the global face level. Gray nodes represent the observations extracted from the face image. White nodes are the hidden variables describing 'types' of observations.

feature is not only *explained* using grayscale appearance, but also with color information, as depicted in Figure 2. 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 color information directly at the facial feature level: we choose to correlate the grayscale appearance of a particular facial feature with its color information instead of correlating color of symmetric features together. The purpose of the second additional layer is to separate the appearance of the facial feature from its color.

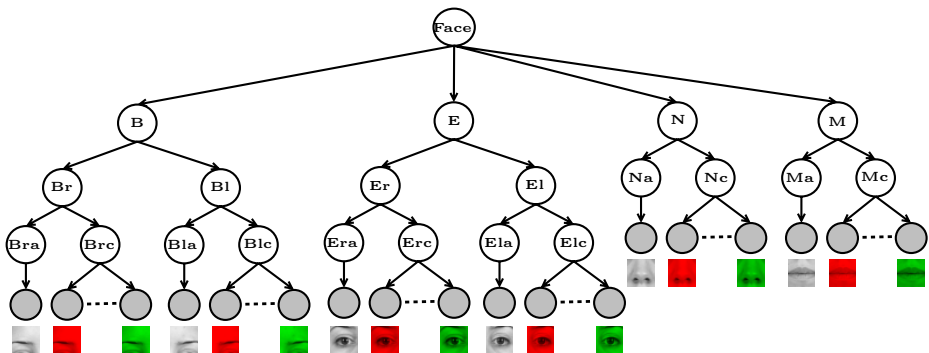


Fig. 2. Bayesian Network model for the face incorporating color information at the local facial feature level

3.4 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 2, where the branch of the global model (Figure 1) corresponding to color information is added.

3.5 Feature Extraction

As observations are derived around facial features, they are first located in the face image using an Active Shape Model (ASM) [14]. For local grayscale observations, the same feature extraction scheme as in [8] is applied here. The original color image is converted to grayscale. Multiple squared windows are then cropped around each facial feature, by adding shifts of a variable amount of pixels. Each extracted window is preprocessed using histogram equalization in order to enhance its contrast. Finally, a feature vector is obtained by applying a two-dimensional Discrete Cosine Transform (2D-DCT) on each preprocessed window.

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.

4 Face Authentication and Performance Measures

Face authentication consists in confirming (or denying) a client's claim supported by its face image. In such a framework, either the claimant provides its real identity, either it is trying to fool the system (it is then referred to as an *impostor*). The system has thus to make a decision on whether the claimant is a true client or an impostor. Since modelling all possible impostors is not feasible, a so-called *world model* is trained thanks to the EM algorithm with the Maximum Likelihood (ML) criterion [13] using data coming from different identities. In face authentication, there are usually few training examples available for each client, and hence Maximum Likelihood estimates of the parameters for the client

specific models may be inaccurate. To tackle this problem, a form of Maximum A Posteriori (MAP) adaptation [15] is used to adapt client models from a nearby distribution, given by the world model. This approach was already successfully applied to this task [16] [8].

When using generative models, authentication decision is typically performed by taking the likelihood ratio between the model corresponding to the claimed identity and the world model, which is used to represent arbitrary impostors. Given a client's claim supported by its face representation \mathbf{X} (i.e. the set of observations derived from the face image, as depicted on Figures 1 and 2), the decision is made according to:

$$\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\theta_C) - \log p(\mathbf{X}|\theta_{world}) \quad (2)$$

where $p(\mathbf{X}|\theta_C)$ is the probability that the client's model θ_C has generated the data \mathbf{X} and $p(\mathbf{X}|\theta_{world})$ is the probability that the data were generated by an impostor. Based on a threshold τ , the claim is accepted if $\Lambda(\mathbf{X}) \geq \tau$.

In a face authentication framework, two kinds of error can occur: either the true claimant is rejected (false rejection), or an impostor is accepted (false acceptance). Hence, authentication results are typically presented using the Half Total Error Rate, which combines the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) and is defined as:

$$HTER(\tau) = (FAR(\tau) + FRR(\tau))/2 \quad [\%] \quad (3)$$

5 Experiments and Results

In this work, we are interested in seeing if color is a valuable source of information for the face authentication task and, more importantly, we also would like to know if combining color and grayscale intelligently is better than treating such information independently. Color is here encoded in the HSV colorspace, since it was previously shown to be suitable for various computer vision tasks. Results are reported using the three proposed models (referred to as *Global-HSV*, *Local-HSV* and *Both-HSV*), but also with three baseline algorithms. The first one is similar to [8] and consists in the first proposed model, but where the color part has been discarded (see Figure 1) and is referred to as *Local-gray*. The second and third one are our own implementations of the traditional Eigenfaces algorithm [4] (*PCA-gray*) and of the color-based Eigenfaces (*PCA-HSV*) proposed in [3].

Results presented throughout this work are obtained using the following settings for feature extraction: windows of size 24x24 pixels were extracted from the original images based on the results given by the ASM, with shifts of 2, 4 and 6 pixels in each directions. Regarding global color information, faces windows of size 64x80 pixels were first cropped and then subsampled to 24x24 pixels. We kept the first 64 DCT coefficients to build the final feature vectors. For the Eigenfaces-based system, 95% of the variance was kept, and the standard correlation was used as the metric. All these hyper-parameters, as well as the threshold τ , were selected by minimizing the Equal Error Rate (EER: when the FAR equals the FRR) on a separate validation set.

5.1 Experiments on the XM2VTS Database

The XM2VTS database [9] contains 295 identities, among which 200 are used as true clients and the remainder are used to simulate impostors. Recordings were acquired during four sessions under controlled conditions and covering a time period of five months. Along with the database, two experimental protocols, specifying which images have to be used for training, validation and testing have been defined. In Table 1, HTER performance with its 95% confidence interval is reported for the various systems using both XM2VTS protocols.

Table 1. HTER Performance on the test set of XM2VTS

System		HTER [%]	
		LP1	LP2
Baselines	Local-gray	2.74 (0.96)	2.43 (0.84)
	PCA-gray	5.32 (1.07)	4.28 (1.07)
	PCA-HSV	3.09 (0.87)	1.94 (0.49)
Proposed	Global-HSV	2.33 (0.90)	1.25 (0.69)
	Local-HSV	1.98 (0.87)	0.98 (0.60)
	Both-HSV	1.22 (0.59)	0.59 (0.42)

One can first remark that color is indeed a valuable information. Actually, results obtained with the color-based Eigenfaces algorithm significantly outperforms the classical Eigenfaces, as expected. However, it reaches almost the same performance as the reference generative model acting on grayscale only. This suggest that treating color channels as independent sources of information is not really successful. Note that, on the other hand, the proposed generative models integrating color in a principled way yields better performance than the similar model based on grayscale features only (*Local-gray*), again emphasizing the usefulness of color information in face processing.

Correlating color and grayscale information (*Global-HSV*) seems to be better than treating such information independently (*PCA-HSV*). This is again evidenced by the results obtained with the proposed generative models: they all perform better than the color-based Eigenfaces system. Regarding the different proposed generative models, we can see that modelling color at the facial feature level consistently yields better results than using information derived from the whole face image. This result is not surprising since the local model is inherently more accurate than the global one. Note finally that the best performance is achieved with the model taking both global and local color information into account (*Both-HSV*), suggesting that both clues are valuable and complementary to describe an identity.

5.2 Experiments on the BANCA Database

To assess the validity of our approach, experiments were also carried out with the baseline systems and the proposed generative models on the more challenging

BANCA database [10]. This database contains 52 clients (English corpus), equally divided into two groups $g1$ and $g2$ used for validation and test respectively. An additional set with 10 images of 30 other subjects is also provided as the world model. In this database, image acquisition was performed according to three different scenarios: Controlled (high-quality camera, uniform background, controlled lighting), Degraded (webcam, non-uniform background) and Adverse (high-quality camera, arbitrary conditions). Examples of acquisition conditions can be seen on Figure 3. There exists several experimental protocols defining which scenarios and which images have to be used for enrollment and testing. In this study, the protocols Mc, Ua, Ud, P and G have been considered.

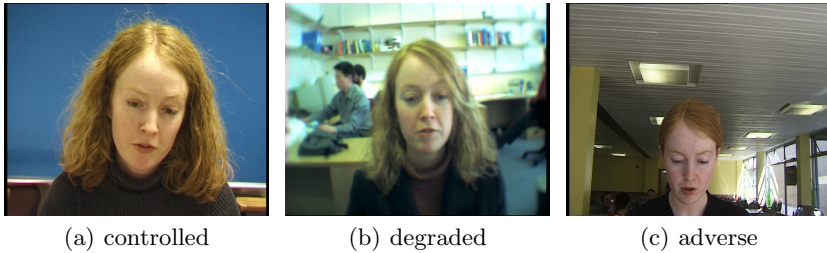


Fig. 3. Example of the different scenarios in the BANCA database

In Table 2, HTER performance with its 95% confidence interval for the different protocols is reported on the test set $g2$. 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. *Global* and *Both*). Indeed, the baseline generative model based on grayscale features (*Local-gray*) performs better when there is a strong mismatch between training and testing conditions. 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

Table 2. HTER on the test set $g2$ of the BANCA database

System		HTER [%]				
		Mc	Ua	Ud	P	G
Baselines	Local-gray	2.24 (0.93)	20.51 (2.64)	19.90 (2.59)	16.52 (1.38)	6.85 (0.95)
	PCA-gray	20.38 (2.62)	41.60 (3.17)	37.02 (3.17)	34.51 (1.80)	21.84 (1.56)
	PCA-HSV	14.71 (2.30)	34.07 (3.11)	32.34 (3.08)	29.78 (1.72)	18.48 (1.47)
Proposed	Global-HSV	5.19 (1.41)	28.11 (2.94)	32.08 (3.04)	24.22 (1.63)	11.50 (1.22)
	Local-HSV	1.89 (0.87)	17.24 (2.43)	20.77 (2.65)	18.80 (1.41)	5.58 (0.86)
	Both-HSV	3.21 (1.09)	21.79 (2.72)	22.02 (2.66)	19.94 (1.52)	6.31 (0.93)

improves the performance, and clearly outperforms the Eigenfaces-based systems. 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 3).

6 Conclusion

In this contribution, new generative models based on Bayesian Networks were proposed to tackle the face authentication task. The purpose of these models was to integrate color in a principled way into a local-feature based model acting on grayscale observations only. To do so, new models were derived: they combine color and grayscale information either at the global face level, at the local facial feature level or at both levels. Face authentication experiments were conducted on two different benchmark databases. Obtained results showed that improvement can be gained when color is combined to grayscale as additional information. Namely, we showed that the proposed models are suitable for the face authentication task, at least when the acquisition conditions between enrollment and testing are quite similar. In particular, the model taking color into account at both the global and the local level significantly outperforms a similar baseline system acting on grayscale-based features, as well as a color-based Eigenfaces algorithm [3]. Correlating different sources of information thus seems to be more effective than treating them independently. However, when there is a strong mismatch between training and testing conditions, color information may become confusing. An obvious possible future direction is hence the investigation of other colorspaces, and particularly the one taking the illuminant into account (i.e. CIE-XYZ and its derivatives). Besides, it would also be interesting to combine more than one colorspace representation, since this approach was shown to yield good results [6].

Acknowledgements

This work has been funded by the GMFace project of the Swiss National Science Foundation (SNSF) and by the MOBIO project 7th Framework Research Programme of the European Union (EU). Softwares were implemented using the TorchVision library¹ and experiments were carried out using the PyVerif framework².

References

1. Russell, R., Sinha, P., Biedermann, I., Naderhouser, M.: Is Pigmentation Important For Face Recognition? Evidence From Contrast Negation. *Perception* 35, 749–759 (2006)

¹ <http://torch3vision.idiap.ch>

² <http://pyverif.idiap.ch>

2. Sinha, P., Balas, B., Ostrovsky, Y., Russel, R.: Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE, Special Issue on Biometrics: Algorithms and Applications* 94(11), 1948–1962 (2006)
3. Torres, L., Reutter, J.Y., Lorente, L.: The Importance of the Color Information in Face Recognition. In: *IEEE Intl. Conf. on Image Processing (ICIP)*, vol. 3, pp. 627–631 (1999)
4. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. In: *IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–591 (1991)
5. Gutta, S., Huang, J., Chengjun, L., Wechsler, H.: Comparative Performance Evaluation of Gray-Scale and Color Information for Face Recognition Tasks. In: Bigun, J., Smeraldi, F. (eds.) *AVBPA 2001. LNCS*, vol. 2091, pp. 38–43. Springer, Heidelberg (2001)
6. Sadeghi, M., Khoshrou, S., Kittler, J.: SVM-Based Selection of Colour Space Experts for Face Authentication. In: Lee, S.-W., Li, S.Z. (eds.) *ICB 2007. LNCS*, vol. 4642, pp. 907–916. Springer, Heidelberg (2007)
7. Jones, C.I., Abott, A.L.: Color Face Recognition by Hypercomplex Gabor Analysis. In: *IEEE Intl. Conf. on Automatic Face and Gesture Recognition (AFGR)*, pp. 126–131 (2006)
8. Heusch, G., Marcel, S.: Face Authentication with Salient Facial Features and Static Bayesian Network. In: Lee, S.-W., Li, S.Z. (eds.) *ICB 2007. LNCS*, vol. 4642, pp. 878–887. Springer, Heidelberg (2007)
9. Messer, K., Matas, J., Kittler, J., Lüttin, J., Maitre, G.: XM2VTSDB: The Extended M2VTS Database. In: *Intl. Conf. Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 72–77 (1999)
10. Bailly-Baillière, E., et al.: The Banca Database and Evaluation Protocol. In: Kittler, J., Nixon, M.S. (eds.) *AVBPA 2003. LNCS*, vol. 2688, pp. 625–638. Springer, Heidelberg (2003)
11. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco (1988)
12. Cowell, G., Dawid, P., Lauritzen, L., Spiegelhalter, J.: *Probabilistic Networks and Expert Systems*. Springer, Heidelberg (1999)
13. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood From Incomplete Data via the EM Algorithm. *The Journal of Royal Statistical Society* 39, 1–37 (1977)
14. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active Shape Models: Their Training and Applications. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
15. Gauvain, J.L., Lee, C.H.: Maximum A Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. *IEEE Trans. on Speech and Audio Processing* 2(2), 291–298 (1994)
16. Cardinaux, F., Sanderson, C., Bengio, S.: User Authentication via Adapted Statistical Models of Face Images. *IEEE Trans. on Signal Processing* 54(1), 361–373 (2005)