Parts-Based Face Verification Using Local Frequency Bands

Christopher McCool and Sébastien Marcel

Idiap Research Institute, Centre du Parc, CH-1920 Martigny, Switzerland
christopher.mccool@idiap.ch, sebastien.marcel@idiap.ch

Abstract. In this paper we extend the Parts-Based approach of face verification by performing a frequency-based decomposition. The Parts-Based approach divides the face into a set of blocks which are then considered to be separate observations, this is a spatial decomposition of the face. This paper extends the Parts-Based approach by also dividing the face in the frequency domain and treating each frequency response from an observation separately. This can be expressed as forming a set of sub-images where each sub-image represents the response to a different frequency of, for instance, the Discrete Cosine Transform. Each of these sub-images is treated separately by a Gaussian Mixture Model (GMM) based classifier. The classifiers from each sub-image are then combined using weighted summation with the weights being derived using linear logistic regression. It is shown on the BANCA database that this method improves the performance of the system from an Average Half Total Error Rate of 24.38% to 15.17% when compared to a GMM Parts-Based approach on Protocol P.

1 Introduction

The face is an object that we as humans know can be recognised. It is used to verify the identity of people on a daily basis through its inclusion in passports, drivers licences and other identity cards. However, performing automatic face verification has proved to be a very challenging task. This is shown by the fact that face recognition has been an active area of research for over 25 years [1], in fact the earliest research into face recognition was conducted by Bledsoe [2] in 1966.

Many techniques have been proposed to perform face verification ranging from Principal Component Analysis (PCA) [3] and Linear Discriminant Analysis (LDA) [4] through to feature distribution modelling techniques such as Hidden Markov Models (HMMs) [5] and Gaussian Mixture Models (GMMs) [6]. A recent advance in face verification has been the effective use of feature distribution modelling techniques. The first effective method of performing face verification using feature distribution modelling was in 2002 by Sanderson and Paliwal [6]; despite the earlier work of Samaria et al. [5,7] and Nefian and Hayes [8] who used HMMs.

The GMM Parts-Based approach, introduced by Sanderson and Paliwal, has been employed by several researchers [9,10]. This method consists of dividing

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 259–268, 2009.

the face into blocks, or parts, and to then consider each block separately. The distribution of these parts is then modelled using Gaussian Mixture Modelling. This method is very different to other Parts-Based (or Component-Based) approaches which form an expert classifier each individual region or concatenate the information from the different Parts and then use a holistic-based classifier, for instance using SVMs [11].

In this paper an extension to the GMM Parts-Based approach (referred to as the Parts-Based approach from hereon) is proposed so that both Spatial and Frequency based decomposition can be performed. The frequency decomposition is achieved by collating the responses from each DCT coefficient from each block (observation) and forming a separate sub-image for each frequency. Each of these sub-images is then treated separately by a GMM based classifier. The classifiers from each sub-image are then combined using weighted summation with the weights being derived using linear logistic regression. Tests conducted on the BANCA database show that this extension is a significant improvement with the Average Half Total Error Rate being reduced from of 24.38% to 15.17% when compared to a baseline Parts-Based approach.

2 Related Work on GMM Parts-Based Face Verification

The Parts-Based approach divides the face into blocks, or parts, and treats each block as a separate observation of the same underlying signal (the face). In this method a feature vector is obtained from each block by applying the Discrete Cosine Transform and the distribution of these feature vectors is then modelled using GMMs. Several advances have been made upon this technique, for instance, Cardinaux et al. [9] proposed the use of background model adaptation while Lucey and Chen [10] examined a method to retain part of the structure of the face utilising the Parts-Based framework as well as proposing a relevance based adaptation.

2.1 Feature Extraction

The feature extraction algorithm is described by the following steps. The face is normalised, registered and cropped. This cropped and normalised face is divided into blocks (parts) and from each block (part) a feature vector is obtained. Each feature vector is treated as a separate observation of the same underlying signal (in this case the face) and the distribution of the feature vectors is modelled using GMMs. This process is illustrated in Figure 1.

The feature vectors from each block are obtained by applying the DCT. Even advanced feature extraction methods such as the DCTmod2 method [6] use the DCT as their basis feature vector; the DCTmod2 feature vectors incorporate spatial information within the feature vector by using the deltas from neighbouring blocks. The advantage of using only DCT feature vectors is that each DCT coefficient can be considered to be a frequency response from the image (or block). This property is exploited by the JPEG standard [12] where the coefficients are ranked in ascending order of their frequency.

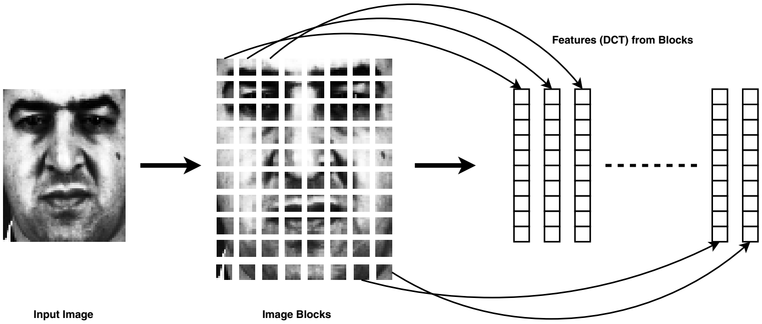


Fig. 1. A flow chart of describing the extraction of feature vectors from the face image for Parts-Based approaches

2.2 Feature Distribution Modelling

Feature distribution modelling is achieved by performing background model adaptation of GMMs [9,10]. The use of background model adaptation is not new to the field of biometric authentication in fact it is commonly used in the field of speaker verification [13]. Background model adaptation first trains a world (background) model Ω_{world} from a set of faces and then derives the client model for the i^{th} client Ω^{i}_{client} by adapting the world model to match the observations of the client.

Two common methods of performing adaptation are mean only adaptation [14] and full adaptation [15]. Mean only adaptation is often used when there are few observations available because adapting the means of each mixture component requires fewer observations to derive a useful approximation. Full adaptation is used when there are sufficient observations to adapt all the parameters of each mode. Mean only adaptation is the method chosen for this work as it requires fewer observations to perform adaptation, this is the same adaptation method employed by Cardinaux et al. [9].

2.3 Verification

A description of the Parts-Based approach is not complete without defining how an observation is verified. To verify an observation, $\boldsymbol{x}$, it is scored against both the client (Ω^{i}_{client}) and world (Ω_{model}) model, this is true even for methods that do not perform background model adaptation [6]. The two models, Ω^{i}_{client} and Ω_{world}, produce a log-likelihood score which is then combined using the log-likelihood ratio (LLR),

$$h(\boldsymbol{x}) = \ln(p(\boldsymbol{x} \mid \Omega^{i}_{client})) - \ln(p(\boldsymbol{x} \mid \Omega_{world})), \tag{1}$$

to produce a single score. This score is used to assign the observation to the world class of faces (not the client) or the client class of faces (it is the client) and consequently a threshold τ has to be applied to the score $h(\boldsymbol{x})$ to declare (verify) that $\boldsymbol{x}$ matches to the i^{th} client model Ω^{i}_{client} when $h(\boldsymbol{x}) \geq \tau$.

3 Local Frequency Band Approach

The method proposed in this paper is to divide the face into separate blocks and to then decompose these blocks in the frequency domain. This can be achieved by treating the frequency response from each block separately to form frequency sub-images. This method is applied to the DCT feature vectors obtained by applying the Parts-Based approach. Each coefficient can be considered independently because each coefficient of the DCT is orthogonal.

The technique is summarised as follows: (1) the face is cropped and normalised to a 68 × 68 image, (2) this image is divided into 8 × 8 blocks with an overlap of 4 pixels in the horizontal and vertical axes, (3) the DCT coefficients from each block are separated and used to form their own frequency sub-image, and (4) a feature vector is formed by taking a block from the frequency sub-image and vectorising the block. The way in which the frequency sub-images are formed is demonstrated in Figure 2.

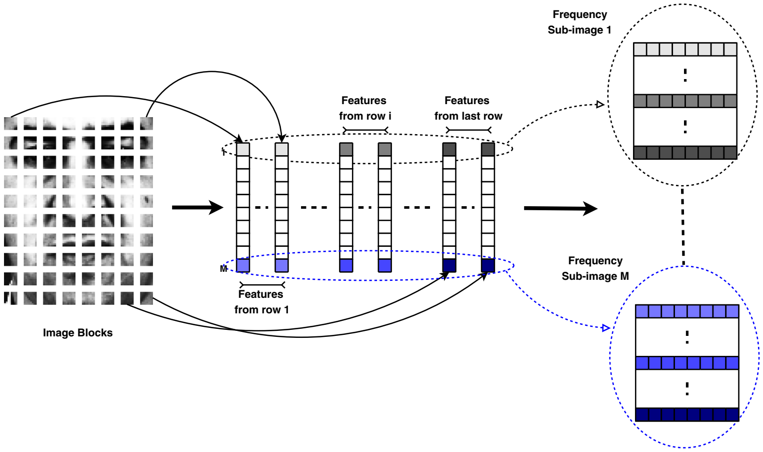
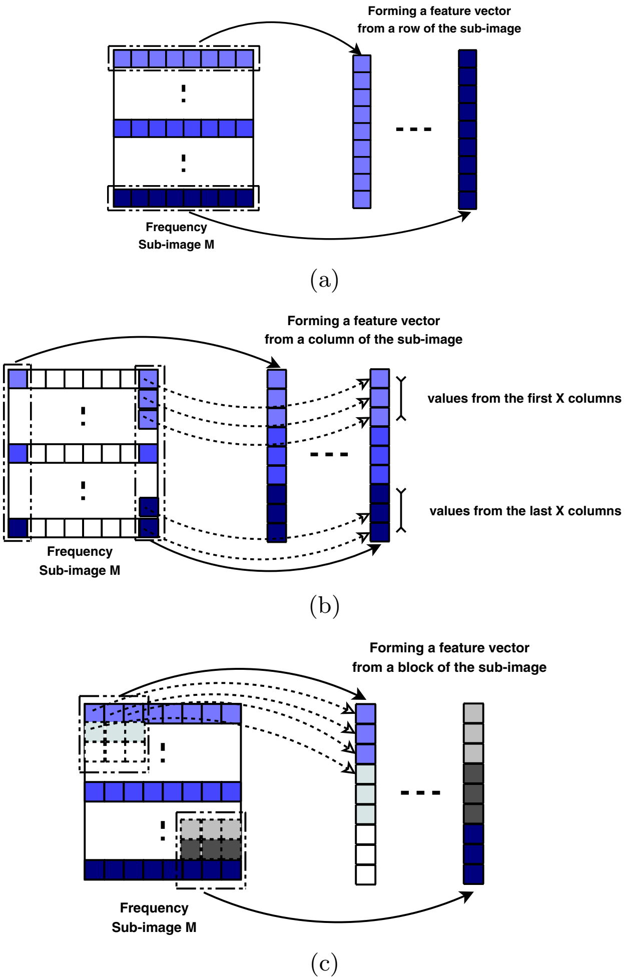


Fig. 2. The figure above describes how the face can be decomposed into separate frequency sub-bands (sub-images)

3.1 Motivation

To illustrate the differences between the frequency decomposition approach and the full Parts-Based approach the following statements are made. For the Parts-Based approach it is often stated that the face is broken into blocks and the distribution of each block is then modelled [6,10], however, another stricter statement would be that the frequency information from each block is simultaneously modelled since each dimension of the feature vector represents a different sampling frequency of the DCT. By contrast the frequency decomposition approach separates the frequency information from each local block and forms a feature vector from the resulting frequency sub-images. Many feature vectors are formed from a frequency sub-image and then modelled using background model adaptation, thus, the image is decomposed in both the spatial domain and the frequency domain.


Forming a feature vector
from a row of the sub-image
Frequency
Sub-image M
Forming a feature vector
from a column of the sub-image
values from the first X columns
values from the last X columns
Frequency
Sub-image M
Forming a feature vector
from a block of the sub-image
Frequency
Sub-image M

(a)

(b)

(c)

Fig. 3. Forming the feature vectors (a) along the row of the frequency sub-image, (b) along the columns of the frequency sub-image and (c) using blocks of the frequency sub-image

A side effect of working on the frequency sub-images is that the feature vectors formed from these sub-images will retain extra spatial information. This is because the Parts-Based approach obtains an observation from a block, however, the frequency decomposition approach gets the response from each block and then forms a feature vector using responses from several blocks. This means that the feature vectors formed from the frequency sub-images will actually span several blocks when compared to the Parts-Based approach, for instance the feature vector could be formed from a frequency sub-image by spanning an entire row or column of the image.

3.2 Feature Extraction

Three methods of forming a feature vector from the frequency sub-images are examined, these are to form a feature vector: (1) across the row of the frequency sub-image (row-based approach), (2) across the column of the frequency sub-image (column-based approach), or (3) from a 4×4 block of the frequency sub-image which is then vectorised (block-based approach). The choice of a 68×68 image results in frequency sub-images of size 16×16 which allows for the fair comparison of the three different feature extraction methods as each method will result in feature vectors of dimension $D = 16$ with $o = 16$ observations from each frequency sub-image. A visualisation of these three methods is provided in Figure 3.

3.3 Classifier

Having obtained these feature vectors a classifier is formed using the same background model adaptation approach that was used for the Parts-Based approach [9]. Each local frequency sub-band (k) produces a separate classifier (C_k) and these classifiers are then combined using weighted linear score fusion, $C_{w\_sum} = \sum_{k=1}^{K} \beta_k C_k$. This fusion technique is used as it was shown by Kittler et al. [16] that the sum rule (which is what weighted linear classifier score fusion abstracts to be) is robust to estimation errors. The weights, β_k, for the classifiers are derived using an implementation of linear logistic regression [17].

4 Results and Discussion

The database used for these experiments is the BANCA English database [18]. The images are cropped and scaled to a size of 68×68 pixels with a distance between the eyes of 33 pixels. Illumination normalisation is applied to each image as a two stage process, the image is histogram equalised and then encoded using a Local Binary Pattern (LBP) [19]. This is the same normalisation strategy employed by Heusch et al. [20] where the parameters used for the LBP are $R = 2$ and $P = 8$.

Results for the BANCA database are presented as the average Half Total Error Rate (HTER) using g_1 and g_2 in a cross-validated manner. Parameters such as the optimal number of mixture components M are derived using the development set of Protocol P and used as a constant throughout the remaining tests. The decision threshold and other parameters such as classifier weights, β_k, are derived on the independent development (tuning) set for each protocol.

Two baseline systems are considered for this work, one system uses DCT feature vectors and the other uses DCTmod2 feature vectors. DCTmod2 feature vectors are examined as it was previously found to be more robust than DCT feature vectors [6]. The size of the feature vectors, $D = 15$ for DCT feature vectors and $D = 18$ for DCTmod2 feature vectors, was chosen based on work conducted in [6]. Both baseline systems were developed using 68×68 face images with a varying number mixtures $M = [100, 200..., 500]$.

Results on the Development set of the P protocol found that, for both manual and automatic eye positions, the system using DCT feature vectors with $M = 100$ mixture components provides the best performance. This system was then used to produce full results for all of the BANCA protocols and are presented in Table 1. It can be seen that there is a consistent degradation in performance of, on average, 2.37% in absolute performance difference when automatic eye locations are used.

The effect of using manual and automatic eye positions is examined since any deployed face verification system will need to cope with errors introduced from an automatic face detection system. The manually annotated eye positions were provided with the BANCA database and the automatically annotated eye positions were obtained using a face detector based on a cascade of LBP-like features [21] [1]. There were 93 images (out of 6,540 images) where the automatic face detector could not find the face, these images were excluded from training, development and evaluation of the automatic systems.

The initial experiments indicated that all of the local frequency sub-band approaches, across all BANCA protocols, provided significantly improved performance when compared to the baseline system. It can be seen from the results in Table 2 that the optimal local frequency sub-band approach is the column-based approach followed by the block-based approach and finally the row-based approach. The column-based approach provides an absolute improvement of 9.21% for Protocol P with the HTER reducing from 24.38% to 15.17%; the frequency sub-band systems are optimised in a similar manner to the baseline systems, however, because there are fewer observations ($o = 16$ observations for each frequency sub-image whereas $o = 256$ for the Parts-Based approach) the number of mixtures were constrainted to $M = [5, 10, ..., 25]$.

Table 1. This table presents the average HTER for the baseline Parts-Based verification system across all of BANCA protocols

	P	G	Ud	Ua	Mc	Md	Ma
manual annotations	24.38%	12.38%	24.60%	29.46%	8.21%	16.35%	18.91%
automatic annotations	26.34%	15.33%	27.86%	30.51%	11.88%	18.63%	20.33%

The performance of the frequency sub-band approach when using automatic eye locations is superior to that of the baseline system. This is true across all test conditions except for the block-based approach on Protocol Mc, see Table 3 for full results. Examining the results in more detail it becomes obvious that the column-based approach degrades in a similar manner as the baseline system with an average absolute performance difference of 2.51%. This means that for both manual and automatic eye locations the column-based system significantly outperforms the baseline system. However, a different result appears for the row-based and block-based approaches whose performance degrades significantly,

[1] This detector has been implemented with torch3vision.idiap.ch

Table 2. This table presents the average HTER for the local frequency sub-band approaches on manually annotated eye locations for all of the BANCA protocols. Each system has $M_{row} = 15$, $M_{blk} = 20$ and $M_{col} = 5$ mixture components respectively. Highlighted are the best performing systems.

	P	G	Ud	Ua	Mc	Md	Ma
row-based	19.418%	7.24%	18.19%	23.06%	7.18%	9.98%	12.42%
block-based	16.74%	**5.72%**	14.46%	25.03%	**5.51%**	**7.20%**	**10.29%**
column-based	**15.17%**	5.80%	**13.40%**	**19.20%**	6.73%	7.88%	11.04%

Table 3. This table presents the average HTER for the local frequency sub-band approaches on automatically annotated eye locations for all of the BANCA protocols. Each system has $M_{row} = 15$, $M_{blk} = 20$ and $M_{col} = 5$ mixture components respectively. Highlighted are the best performing systems.

	P	G	Ud	Ua	Mc	Md	Ma
row-based	24.57%	9.97%	25.02%	27.81%	10.37%	16.66%	**15.00%**
block-based	21.86%	9.93%	23.04%	25.64%	12.34%	14.21%	16.79%
column-based	**16.64%**	**7.4%**	**17.50%**	**21.06%**	**8.07%**	**10.26%**	15.89%

Table 4. This table presents state-of-the-art results on the BANCA database using the average HTER (%) on the Mc protocol for Manual eye annotations

	SB_{Col}	BN	PSC-GMM	GMM
HTER on g1 (%)	10.13	9.01	11.31	N/A
HTER on g2 (%)	3.33	5.41	11.34	8.9

their HTER increases by an absolute average of 4.56% and 5.55% respectively when using automatic eye locations.

The result for the automatic eye locations demonstrates that the choice of feature vector formation has a significant impact on the local frequency sub-band approach. The column-based approach has empirically been shown to be more robust to localisation errors than either the row-based or block-based approaches. The column-based also performs better than either the row-based or block-based approaches for most of the test conditions. This fact could be explained by suggesting that the features of the face are more stable when scanned in a vertical manner, particularly when there is misalignment of the face image. However, this argument has not been explored fully and so it forms the basis of future work for this technique.

Finally, this system compares well to state-of-the-art techniques for the BANCA database. Previous state-of-the-art face verification systems, tested on the Mc protocol of BANCA, are taken from the work of Heusch and Marcel [22] who provide results from a Bayesian Network classifier (BN), a Partial Shape Collapse GMM classifier (PSC-GMM) and a state-of-the-art GMM Parts-Based classifier[2] are provided. The results are reproduced in Table 4 where it can be

[2] Results for the state-of-the-art GMM Parts-Based classifier are only available for g2.

seen that the column-based sub-band (SB_{Col}) approach performs very competitively when compared to the more complex BN approach.

5 Conclusions and Future Work

In this paper a novel extension to the Parts-Based approach has been introduced that results in an absolute improvement of the HTER of 9.21%. This novel extension decomposes the face into local frequency sub-bands. Feature vectors are extracted from these frequency sub-band images in one of three ways, in a: row-based manner, column-based manner or a block-based manner. It has been shown empirically that extracting the feature vectors in a column-based manner results in a robust and accurate method for performing face verification. However, it remains an open question why the column-based approach to feature formation is more robust and accurate than the row-based or block-based approaches.

Future work will examine potential reasons for the superior performance of the column-based approach. Work will also be conducted to determine if local frequency sub-band's can be merged, performing a feature level fusion, to simplify the task of classifier fusion. Finally, any future work will also need to address the possibility of applying a HMM approach to this local frequency sub-band technique.

Acknowledgements

This work has been performed by the MOBIO project 7th Framework Research Programme of the European Union (EU), grant agreement number: 214324. The authors would like to thank the EU for the financial support and the partners within the consortium for a fruitful collaboration. For more information about the MOBIO consortium please visit `http://www.mobioproject.org`

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys 35(4), 399–459 (2003)
2. Bledsoe, W.W.: The model method in facial recognition. Technical report for Panoramic Research Inc. (1966)
3. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
5. Samaria, F., Fallside, F.: Face identification and feature extraction using hidden markov models. Image Processing: Theory and Applications, 295–298 (1993)
6. Sanderson, C., Paliwal, K.K.: Fast feature extraction method for robust face verification. Electronic Letters 38(25), 1648–1650 (2002)
7. Samaria, F., Young, S.: Hmm-based architecture for face identification. Image and Vision Computing 12(8), 537–543 (1994)

8. Nefian, A., Hayes III, M.H.: Hidden markov models for face recognition. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 2721–2724 (1998)
9. Cardinaux, F., Sanderson, C., Marcel, S.: Comparison of mlp and gmm classifiers for face verification on xm2vts. In: International Conference on Audio- and Video-based Biometric Person Authentication, pp. 1058–1059 (2003)
10. Lucey, S., Chen, T.: A gmm parts based face representation for improved verification through relevance adaptation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 855–861 (2004)
11. Heisele, B., Ho, P., Wu, J., Poggio, T.: Face recognition: component-based versus global approaches. In: Computer Vision and Image Understanding, vol. 91, pp. 6–21 (2003)
12. Pennebaker, W.B., Mitchell, J.L.: JPEG still image data compression standard. Van Nostrand Reinhold, New York (1993)
13. Doddington, G., Przybocki, M., Martin, A., Reynolds, D.: The NIST speaker recognition evaluation — overview, methodology, systems, results, perspective. Speech Communication 31(2-3), 225–254 (2000)
14. Reynolds, D.: Comparison of background normalization methods for text-independent speaker verification. In: Proc. European Conference on Speech Communication and Technology (Eurospeech), vol. 2, pp. 963–966 (1997)
15. Lee, C., Gauvain, J.: Bayesian adaptive learning and MAP estimation of HMM, pp. 83–107. Kluwer Academic Publishers, Boston (1996)
16. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 226–239 (1998)
17. Brummer, N.: Tools for fusion and calibration of automatic speaker detection systems (2005), `http://www.dsp.sun.ac.za/~nbrummer/focal/index.htm`
18. Bailly-Bailliere, E., Bengio, S., Bimbo, F., Hamouz, M., Kittler, J., Mariethoz, J., Matas, J., Messer, K., Popovici, V., Poree, F., Ruiz, B., Thiran, J.P.: The banca database and evaluation protocol. LNCS, pp. 625–638. Springer, Heidelberg (2003)
19. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)
20. Heusch, G., Rodriguez, Y., Marcel, S.: Local binary patterns as an image preprocessing for face authentication. In: International Conference on Automatic Face and Gesture Recognition, pp. 9–14 (2006)
21. Fröba, B., Ernst, A.: Face detection with the modified census transform. In: IEEE Conference on Automatic Face and Gesture Recognition, pp. 91–96 (2004)
22. Heusch, G., Marcel, S.: Face authentication with Salient Local Features and Static Bayesian network. In: IEEE / IAPR Intl. Conf. On Biometrics (ICB) (2007) IDIAP-RR 07-04