

Distance Measures for Gabor Jets-Based Face Authentication: A Comparative Evaluation

Daniel González-Jiménez¹, Manuele Bicego², J.W.H. Tangelder³,
B.A.M Schouten³, Onkar Ambekar³, José Luis Alba-Castro¹, Enrico Grosso²,
and Massimo Tistarelli⁴

¹ TSC Department, University of Vigo, Vigo (Spain)

{danisub,jalba}@gts.tsc.uvigo.es

² DEIR - University of Sassari, Sassari (Italy)

{bicego,grosso}@uniss.it

³ CWI, Amsterdam (The Netherlands)

{J.W.H.Tangelder,B.A.M.Schouten,Onkar.Ambekar}@cwi.nl

⁴ DAP - University of Sassari, Alghero (Italy)

tista@uniss.it

Abstract. Local Gabor features (jets) have been widely used in face recognition systems. Once the sets of jets have been extracted from the two faces to be compared, a proper measure of similarity (or distance) between corresponding features should be chosen. For instance, in the well known Elastic Bunch Graph Matching (EBGM) approach and other Gabor-based face recognition systems, the cosine distance was used as a measure. In this paper, we provide an empirical evaluation of seven distance measures for comparison, using a recently introduced face recognition system, based on Shape Driven Gabor Jets (SDGJ). Moreover we evaluate different normalization factors that are used to pre-process the jets. Experimental results on the BANCA database suggest that the concrete type of normalization applied to jets is a critical factor, and that some combinations of *normalization + distance* achieve better performance than the classical cosine measure for jet comparison.

1 Introduction

Face analysis and recognition is an important and active research area [15], whose interest has increased in recent years for both theoretical and application-driven motivations. Among the huge amount of approaches presented in this context, a wide group is based on the extraction of a particular class of features from points on the face, automatically or manually found. Different features have been proposed, like DCT [11], Local Binary Patterns [13], SIFT [12,10] and others, showing different characteristics, in terms of robustness, ease of computation, computational requirements etc. In this context, Gabor features received great attention, and several methods for face recognition using these features have been proposed (see [1,2,3,4] among others — and [5] for a recent survey). One of the most important algorithms falling in that category is the Elastic Bunch Graph Matching (EBGM) approach proposed by Wiskott et al. [1]. In

this technique, Gabor responses (jets) are extracted from a set of (the so-called) fiducial points, located at specific face regions (eyes, tip of the nose, mouth. . .). On the other hand, a recently proposed method in [4] focuses on the selection of an *own* set of points and features for a given client by exploiting facial structure (Shape-Driven Gabor Jets or SDGJ). Although both approaches used cosine distance in order to compare corresponding features, this choice is not motivated, neither with a theoretical nor with an experimental evaluation. To the best of our knowledge, the only evaluation of distances for Gabor jet comparison was performed in [14], where the authors concluded that Manhattan (or city block) distance outperformed both cosine and euclidean distances. However, it is not explicitly described, neither in [14] nor in other research papers dealing with Gabor jets-based face recognition systems, whether jets have been previously normalized or not. In this paper we propose a more extensive evaluation, comparing seven different distances for measuring similarities between Gabor jets, as well as assessing the impact of the concrete normalization method that is applied to jets before comparison. Finally, three different resolutions of input images are tested in order to provide a more complete set of results.

The paper is organized as follows. Section 2 briefly introduces the Shape-driven Gabor Jets (SDGJ) algorithm: Section 2.1 describes the set of Gabor filters used for feature extraction, and Section 2.2 presents the algorithm used to map points between two faces. The different distances involved in the comparison are introduced in Section 3. Experimental results are given in Section 4. Finally, conclusions and future research lines are drawn in Section 5.

2 Shape-Driven Gabor Jets (SDGJ)

In this approach [4], the selection of points is accomplished by exploiting shape information. Lines depicting face structure are extracted by means of a ridges and valleys detector [6], leading to a binary representation that sketches the face. In order to select a set of points from this sketch, a dense rectangular grid ($n_x \times n_y$ nodes) is applied onto the face image and each grid node is moved towards its nearest line of the sketch. Finally, a set of points $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$, with $n = n_x \times n_y$ are obtained. This set of points samples the original sketch, as it can be seen in figure 1.

2.1 Extracting Textural Information

A set of 40 Gabor filters $\{\psi_m\}_{m=1,2,\dots,40}$, with the same configuration as in [1], is used to extract textural information. These filters are convolution kernels in the shape of plane waves restricted by a Gaussian envelope, as it is shown next:

$$\psi_m(\mathbf{x}) = \frac{\|\mathbf{k}_m\|^2}{\sigma^2} \exp\left(\frac{-\|\mathbf{k}_m\|^2 \|\mathbf{x}\|^2}{2\sigma^2}\right) \left[\exp(i \cdot \mathbf{k}_m \mathbf{x}) - \exp\left(\frac{-\sigma^2}{2}\right) \right] \quad (1)$$



Fig. 1. Left: Original rectangular dense grid. Center: Valleys and ridges sketch. Right: Grid adjusted to the sketch.

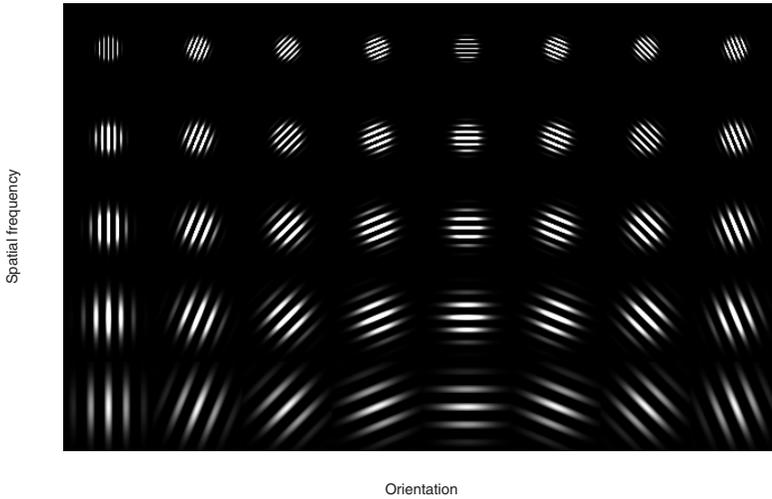


Fig. 2. Real part of the set of 40 (8 orientations \times 5 scales) Gabor filters used in this paper

where \mathbf{k}_m contains information about scale and orientation, and the same standard deviation $\sigma = 2\pi$ is used in both directions for the Gaussian envelope. Figure 2 shows the real part of the 40 Gabor filters used in this paper.

The region surrounding a pixel in the image is encoded by the convolution of the image patch with these filters, and the set of responses is called a jet, \mathcal{J} . So, a jet is a vector with 40 complex coefficients, and it provides information about an specific region of the image. At each shape-driven point $\mathbf{p}_i = [x_i, y_i]^T$, we get the following feature vector:

$$\{\mathcal{J}_{\mathbf{p}_i}\}_m = \sum_x \sum_y I(x, y) \psi_m(x_i - x, y_i - y) \tag{2}$$

where $\{\mathcal{J}_{\mathbf{p}_i}\}_m$ stands for the m -th coefficient of the feature vector extracted from \mathbf{p}_i . So, for a given face with a set of points $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$, we get n Gabor jets $\mathcal{R} = \{\mathcal{J}_{\mathbf{p}_1}, \mathcal{J}_{\mathbf{p}_2}, \dots, \mathcal{J}_{\mathbf{p}_n}\}$.

2.2 Mapping Corresponding Features

Suppose that shape information has been extracted from two images, \mathcal{F}_1 and \mathcal{F}_2 . Let \mathcal{S}_1 and \mathcal{S}_2 be the sketches for these incoming images, and let $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ be the set of points for \mathcal{S}_1 , and $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ the set of points for \mathcal{S}_2 . In the SDGJ approach, there does not exist any a priori correspondence between points nor features (i.e. there is no label indicating which pair of points are matched). So, in order to compare jets from both faces, the authors of [4] make use of a point matching algorithm based on shape contexts [7], obtaining a function ξ that maps each point from \mathcal{P} to a point within \mathcal{Q} :

$$\xi(i) : \mathbf{p}_i \implies \mathbf{q}_{\xi(i)} \tag{3}$$

with an associated cost denoted by $C_{\mathbf{p}_i \mathbf{q}_{\xi(i)}}$ [4]. Finally, the feature vector from \mathcal{F}_1 , $\mathcal{J}_{\mathbf{p}_i}$, will be compared to $\mathcal{J}_{\mathbf{q}_{\xi(i)}}$, extracted from \mathcal{F}_2 .

3 Distance Between Faces

Let $\mathcal{R}_1 = \{\mathcal{J}_{\mathbf{p}_1}, \mathcal{J}_{\mathbf{p}_2}, \dots, \mathcal{J}_{\mathbf{p}_n}\}$ be the set of jets calculated for \mathcal{F}_1 and $\mathcal{R}_2 = \{\mathcal{J}_{\mathbf{q}_1}, \mathcal{J}_{\mathbf{q}_2}, \dots, \mathcal{J}_{\mathbf{q}_n}\}$ the set of jets extracted from \mathcal{F}_2 . Before computing distances, each jet \mathcal{J} is processed as follows:

1. Each complex coefficient is replaced by its modulus, obtaining \mathcal{J}' .
2. The obtained vector can be either normalized (to have unit L_1 or L_2 norm for instance) or not. Although some of the distances that will be introduced next, such as cosine distance, are invariant to these normalizations, some of them are not, and it seems that the concrete type of normalization applied to jets could be a critical point. In this paper, the three possibilities described above (no normalization, L_1 normalization and L_2 normalization) will be evaluated. Hence, given a vector, \mathcal{J}' , comprising the moduli of jet coefficients, we divide it by a normalization factor α given by:
 - No normalization: $\alpha = 1$.
 - L_1 normalization: $\alpha = \sum_i |\mathcal{J}'_i|$.
 - L_2 normalization: $\alpha = \sqrt{\sum_i (\mathcal{J}'_i)^2}$.

We will denote the resulting vector by \mathbf{J} ($\mathbf{J} = \mathcal{J}'/\alpha$) and, for the sake of simplicity, we will maintain the name of jet. The distance function between the two faces, $\mathcal{DF}(\mathcal{F}_1, \mathcal{F}_2)$ is given by:

$$\mathcal{DF}(\mathcal{F}_1, \mathcal{F}_2) = \Upsilon_{i=1}^n \{ \mathcal{D}(\mathbf{J}_{\mathbf{p}_i}, \mathbf{J}_{\mathbf{q}_{\xi(i)}}) \} \tag{4}$$

where $\mathcal{D}(\mathbf{J}_{\mathbf{p}_i}, \mathbf{J}_{\mathbf{q}_{\xi(i)}})$ represents the distance used to compare corresponding jets, and $\Upsilon_{i=1}^n \{ \dots \}$ stands for a generic combination rule of the n local distances $\mathcal{D}(\mathbf{J}_{\mathbf{p}_1}, \mathbf{J}_{\mathbf{q}_{\xi(1)}}), \dots, \mathcal{D}(\mathbf{J}_{\mathbf{p}_n}, \mathbf{J}_{\mathbf{q}_{\xi(n)}})$. In the EBGM approach [1] and other

Gabor-based face recognition systems, it is proposed to use a normalized dot product to compare jets. In this work, we assess the performance of the system varying $\mathcal{D}(\dots)$, i.e. we compare the following distances:

1. Cosine distance (negated normalized dot product as used in [1]).

$$\mathcal{D}(X, Y) = -\cos(X, Y) = -\frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} \tag{5}$$

2. Manhattan distance (L_1 metrics or city block distance)

$$\mathcal{D}(X, Y) = L_1(X, Y) = \sum_{i=1}^n |x_i - y_i| \tag{6}$$

3. Squared Euclidean Distance (sum of squared errors-SSE)

$$\mathcal{D}(X, Y) = \text{SSE}(X, Y) = \sum_{i=1}^n (x_i - y_i)^2 \tag{7}$$

4. Chi square distance

$$\mathcal{D}(X, Y) = \chi^2(X, Y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{x_i + y_i} \tag{8}$$

5. Modified Manhattan distance

$$\mathcal{D}(X, Y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i| \sum_{i=1}^n |y_i|} \tag{9}$$

6. Correlation-based distance

$$\mathcal{D}(X, Y) = -\frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2\right)}} \tag{10}$$

7. Canberra distance

$$\mathcal{D}(X, Y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|} \tag{11}$$

In the definition of all presented distances (equation (5) to equation (11)), n stands for the length of the vector, i.e. $n = 40$. It is easy to see that both cosine and correlation-based distances are invariant to α , i.e. the type of normalization (L_1 , L_2 or no normalization at all) applied to jets does not change the result. It is also straightforward to realize that the Modified Manhattan distance is equivalent to the Manhattan distance when jets are normalized to have unit L_1 norm.

4 Face Authentication on the BANCA Database

4.1 Database and Experimental Setup

We have used the english part of the BANCA database on protocol Matched Controlled (MC) to test the different measures. For this database, face and voice were recorded using both high and low quality microphones and cameras. The subjects were captured in three different scenarios: controlled, degraded and adverse over 12 different sessions spanning three months. This subset of the BANCA database (english part) consists of 52 people, half men and half women.

In order to propose an experimental protocol, it is necessary to define a *development* set, on which the system can be adjusted by setting thresholds, etc. and an *evaluation* set, where the system performance can be assessed. For this reason, two disjoint subsets were created, namely $G1$ and $G2$, each one with 26 people (13 men and 13 women). So, when $G1$ is used as development set, $G2$ is used for evaluation and vice versa.

In the experiments carried out, three specific operating conditions corresponding to three different values of the Cost Ratio, $R = FAR/FRR$, namely $R = 0.1$, $R = 1$, $R = 10$ have been considered. Assuming equal a priori probabilities of genuine clients and impostor, these situations correspond to three quite distinct cases:

- $R = 0.1$, FAR is an order of magnitude less harmful than FRR ,
- $R = 1$, FAR and FRR are equally harmful,
- $R = 10$, FAR is an order of magnitude more harmful than FRR .

The so-called Weighted Error Rate (WER) given by:

$$WER(R) = \frac{FRR + R \cdot FAR}{1 + R} \quad (12)$$

was calculated for the test data of groups $G1$ and $G2$ at the three different values of R . The average WER was reported as final performance measure.

4.2 Results

In order to provide a more complete set of results, we repeated the experiments using images at 3 different resolutions (55×51 , 150×115 and 220×200 pixel images). We used the median rule to fuse the n local distances, i.e. $\mathcal{Y}_{i=1}^n \{ \dots \} \equiv$ median. Moreover, in protocol MC there are 5 training images to build the client model. Whenever a test image claims a given identity, this

test face is compared to each of the 5 training images. Hence, we get 5 scores which are combined (using again the median rule) to obtain the final score ready for authentication. Tables 1, 2 and 3 show the obtained results changing the normalization factor (α) applied to jets (no normalization, L_1 normalization and L_2 normalization respectively). If no normalization is applied to jets, the best performing distance is cosine. The remaining choices achieve significantly

Table 1. Average WER (%) using several distance measures $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ to compare jets and different resolution of input images (jets are *not* normalized)

| $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ | Input Image Resolution | | |
|--|------------------------|------------------|------------------|
| | 55×51 | 150×115 | 220×200 |
| Cosine | 5.13 | 4.64 | 3.40 |
| Manhattan | 8.54 | 10.56 | 13.16 |
| SSE | 8.45 | 9.45 | 12.60 |
| χ^2 | 6.53 | 7.53 | 10.28 |
| Modified Manhattan | 8.18 | 9.71 | 12.72 |
| Correlation | 6.70 | 7.01 | 5.45 |
| Canberra | 6.10 | 5.05 | 5.11 |

Table 2. Average WER (%) using several distance measures $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ to compare jets and different resolution of input images (jets are normalized to have unit L_1 norm)

| $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ | Input Image Resolution | | |
|--|------------------------|------------------|------------------|
| | 55×51 | 150×115 | 220×200 |
| Cosine | 5.13 | 4.64 | 3.40 |
| Manhattan | 5.28 | 3.73 | 2.90 |
| SSE | 4.49 | 3.53 | 3.17 |
| χ^2 | 5.33 | 4.03 | 3.01 |
| Modified Manhattan | 5.28 | 3.73 | 2.90 |
| Correlation | 6.70 | 7.01 | 5.45 |
| Canberra | 5.26 | 3.84 | 3.63 |

Table 3. Average WER (%) using several distance measures $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ to compare jets and different resolution of input images (jets are normalized to have unit L_2 norm)

| $\mathcal{D}(\mathbf{J}_{p_i}, \mathbf{J}_{q_{\xi(i)}})$ | Input Image Resolution | | |
|--|------------------------|------------------|------------------|
| | 55×51 | 150×115 | 220×200 |
| Cosine | 5.13 | 4.64 | 3.40 |
| Manhattan | 5.93 | 6.20 | 4.68 |
| SSE | 5.14 | 4.61 | 3.42 |
| χ^2 | 5.60 | 5.29 | 3.80 |
| Modified Manhattan | 4.75 | 3.10 | 3.34 |
| Correlation | 6.70 | 7.01 | 5.45 |
| Canberra | 5.07 | 3.69 | 3.88 |

worse results for all resolutions (except Canberra distance, with similar performance). In the EBGGM approach [1], the authors did not apply any normalization to jets (at least, they did not state it explicitly) and these results may support their choice of the cosine distance for jet comparison. However, the use of L_1 and L_2 normalization factors (tables 2 and 3) leads to completely different conclusions. Cosine is outperformed by other distances such as SSE or Modified Manhattan distance (MMD). If we compare the results obtained, for instance, with the MMD varying the normalization factor, we see that impressive improvements are obtained with the use of L_1 and L_2 normalization factors (WER decreases from 12.72% to 3.34–2.90% using 220×200 pixel images). Hence, we conclude that the concrete type of normalization that is applied to jets is, in fact, a critical point. In [14], the authors observed that the Manhattan distance outperformed cosine after identification experiments. According to our results, Manhattan outperforms cosine when L_1 normalization is used. Although the authors of [14] do not describe whether they have normalized their jets or not, we have obtained results supporting their finding. As stated in Section 3, both cosine and correlation-based distances are invariant to the tested normalization factors, and this is reflected in the obtained results. It is also interesting to note that, in general terms, error rates decrease (or stays approximately equal) as long as the resolution of input images grows. A clear exception occurs when testing Manhattan, SSE, χ^2 and Modified Manhattan distances without normalization. Further research is needed in order to better understand this behavior.

Other results on BANCA. The CSU Face Identification Evaluation System (<http://www.cs.colostate.edu/evalfacerec/index.html>), including several face recognition techniques such as PCA and PCA+LDA, was tested on the BANCA database using protocol MC to provide baseline results for the ICBA 2004 face authentication contest [9]. These baseline performances can be found in <http://www.ee.surrey.ac.uk/banca/icba2004/csuresults.html>. The lowest average *WER*'s using PCA and PCA+LDA were 9.71% and 6.25% respectively. On the other hand, the best performance in this competition (with manual face localization) was achieved by the Université Catholique de Louvain with an average *WER* of 1.95%.

5 Conclusions and Further Research

In this work, we have proposed an empirical evaluation of different combinations of normalization and distance measures for Gabor jet comparison. The SDGJ algorithm was tested on the BANCA database with 3 input image resolutions, 3 distinct normalization factors and 7 distance measures to compare jets. It has been shown that:

- The performance of a given distance strongly depends on the concrete pre-processing applied to jets.
- When no normalization ($\alpha = 1$) is used, the cosine distance outperforms the remaining ones. Although the authors of [1] did not explicitly state whether they have normalized their jets or not, this result would support their choice.

- The use of L_1 and L_2 normalization factors lead to completely different results, turning out that other distances, such as SSE or MMD, achieve better performance than the cosine measure.

Although we have shown that there exist better choices than cosine distance for Gabor jet comparison, no theoretical reasons supporting this fact have been provided. Recently [16], it has been demonstrated that Gabor coefficients can be accurately modeled using Generalized Gaussian Distributions (GGD's) and this finding opens new possibilities in terms of selecting optimal ways to compare jets from a theoretical point of view.

No discriminant analysis of jet coefficients has been applied, i.e. we have considered that all coefficients have the same discriminative power. Thus, the next step will be to select the most important Gabor jet coefficients according to their classification ability, a selection that should depend on the concrete facial region we are analyzing. Moreover, the function $\mathcal{I}_{i=1}^n \{ \dots \}$ used to combine the n local distances should weigh the different jet contributions according to the discriminative power of the facial regions from where they are extracted.

Acknowledgments

Part of this work was performed during the visit of Daniel González-Jiménez and Manuele Bicego to the CWI, in the framework of the Biosecure Network of Excellence mobility program.

References

1. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on PAMI* 19(7), 775–779 (1997)
2. Duc, B., Fischer, S., Bigun, J.: Face Authentication with Gabor Information on Deformable Graphs. *IEEE Transactions on Image Processing* 8(4), 504–516 (1999)
3. Smeraldi, F., Bigun, J.: Retinal Vision applied to Facial Features Detection and Face Authentication. *Pattern Recognition Letters* 23(4), 463–475 (2002)
4. González-Jiménez, D., Alba-Castro, J.L.: Shape Contexts and Gabor Features for Face Description and Authentication. In: *Proceedings IEEE ICIP 2005*, pp. 962–965 (2005)
5. Shen, L., Bai, L.: A review on Gabor wavelets for Face Recognition. *Pattern Analysis and Applications* 9, 273–292 (2006)
6. López, A.M., Lumbreras, F., Serrat, J., Villanueva, J.J.: Evaluation of Methods for Ridge and Valley Detection. *IEEE Transactions on PAMI* 21(4), 327–335 (1999)
7. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. *IEEE Transactions on PAMI* 24(24), 509–522 (2002)
8. Bailly-Bailliere, 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)
9. Messer, K., et al.: Face Authentication Test on the BANCA Database. In: Zhang, D., Jain, A.K. (eds.) *ICBA 2004*. LNCS, vol. 3072, pp. 8–15. Springer, Heidelberg (2004)

10. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the Use of SIFT Features for Face Authentication. In: Proc. of IEEE CVPR Workshop on Biometrics, p. 35 (2006)
11. Kohir, V.V., Desai, U.B.: Face Recognition Using DCT-HMM Approach. In: Proc. Workshop on Advances in Facial Image Analysis and Recognition Technology (AFI-ART), Freiburg, Germany (1998)
12. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision* 60(2), 91–110 (2004)
13. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(12), 2037–2041.
14. Jiao, F., Gao, W., Shan, S.: A Face Recognition Method Based on Local Feature Analysis. In: ACCV 2002, Melbourne, Australia, pp. 188–192 (2002)
15. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey. *ACM Computing Surveys* 35, 399–458 (2003)
16. González-Jiménez, D., Pérez-González, F., Comesaña-Alfaro, P., Pérez-Freire, L., Alba-Castro, J.L.: Modeling Gabor Coefficients Via Generalized Gaussian Distributions for Face Recognition. In: IEEE International Conference on Image Processing 2007 (accepted)