Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 24 (2006) 904–913
image AND vision COMPUTING
www.elsevier.com/locate/imavis

# Local approach for face verification in polar frequency domain

Y. Zana [a,*], R.M. Cesar-Jr [a], R. Feris [b], M. Turk [b]

[a] University of São Paulo, Computer Sciences, R. do Matão, 1010 São Paulo, SP, Brazil
[b] University of California, Santa Barbara, CA, USA

Received 31 August 2005; received in revised form 3 February 2006; accepted 16 February 2006

**Abstract**

We present a face verification system inspired by known properties of the human visual system. In the proposed algorithm the face is normalized for geometry and luminance, and Fourier–Bessel (FB) descriptors are extracted from three locations in the eyes region (local analysis). The resulting representations are embedded in a dissimilarity space, where each image is represented by its distance to all the other images, and a Pseudo-Fisher discriminator is built. Using the FERET database, we submitted the system to a battery of tests under a wide variation of imaging conditions, including expression, age, and illumination variations. Results showed that the system outperformed previous state-of-the-art methods in most testing conditions. To deal with partial occlusions, we implemented an occluded region detector that resulted in low performance loss under up to 50% occlusion level. Finally, we automated the registration step by implementing face and eye detection algorithms. We also showed that the local-FB analysis outperforms the global-FB version of the system and an alternative polar frequency representation. In conclusion, the intermediate-scale local analysis approach used in the proposed system resulted in state-of-the-art face verification performance and high robustness to common problems such as expression, age, and illumination variations and to strong occlusions.


*PACS:* 42.30.Sy; 42.66.S

*Keywords:* Face recognition; Human visual perception

## 1. Introduction

Face verification and identification tasks are highly complex due to the many possible variations of the same subject in different conditions, like illumination, facial expression, and age. Many developers of face recognition algorithms adopted a biologically inspired approach in solving these problems (e.g. [1–3]), thus contributing both to understanding human face processing and to building efficient face recognition technologies.

The approach described in the present paper was inspired by developments in neurophysiology and cognitive psychology, and its fundamentals were first described by [4]. It is based on an image representation that may be analogous to those used by the human visual system (HVS). In particular, we evaluated the performance of a face verification algorithm whose primary features are the magnitude of radial and angular components of faces images and representation in a dissimilarity space. We show that using an intermediate-scale local analysis approach of specific face regions, the performance of the proposed system is improved in comparison to a global analysis, and achieves state-of-the-art performance.[1]

The paper is organized as follows: in the next section, we briefly introduce the reader to the primary spatial processing by the HVS, highlight the related literature and summarize the main contributions of this paper. We describe in Section 3 the Fourier–Bessel Transform (FBT) and the proposed system in Section 4. We introduce the face database and testing methods in Section 5. The experimental results are presented in Section 6. In the final two sections we discuss the results and conclusions.

## 2. Background and previous work

Face recognition algorithms, for both verification and identification tasks, were extensively studied in the last two decades. An exhaustive review of the most representative of

* Corresponding author. Tel.: +55 1130916240.
*E-mail addresses:* zana@vision.ime.usp.br (Y. Zana), cesar@vision.ime.usp.br (R.M. Cesar-Jr), feris@cs.ucsb.edu (R. Feris), mturk@cs.ucsb.edu (M. Turk).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.02.014

[1] Partial results have been described in conference papers [5,6].

these can be found in [7]. These methods can be classified into holistic matching, structural-based matching and hybrid methods. Holistic methods use the whole face as the raw input, while structural-based algorithms use local and configural information. Hybrid methods combine both types of information. The system proposed here can be considered hybrid, since it consists in applying a holistic method (FBT) in different regions of the face. Most of the holistic face recognition algorithms, for both verification and identification, are based on feature extraction from a Cartesian perspective, typical to most analog and digital imaging systems. On the other hand, the HVS is known to process visual stimuli by fundamental shapes defined in polar coordinates. In the early stages, the visual image is filtered by neurons tuned to specific spatial frequencies and location in a linear manner [8]. In further stages, these neurons output is processed to extract global and more complex shape information, such as faces [9]. Electrophysiological experiments in monkey's visual cerebral areas showed that the fundamental patterns for global shape analysis are defined in polar and hyperbolic coordinates [10]. Global pooling of orientation information was also shown by psychophysical experiments to be responsible for the detection of angular and radial Glass dot patterns [11]. Further evidence in favor of a polar representation use by the HVS is the log-polar manner in which the retinal image is mapped onto the visual cortex area [12]. Thus, it is evident that information regarding the global polar content of images is effectively extracted by and is available to the HVS.

Global (spatial) log-polar mapping has been previously explored for feature detection [13], face detection [14], and face recognition [15]. One of the disadvantages of this feature extraction method is the rough representation of peripheral regions. The HVS compensates for this effect by eye saccades, moving the fovea from one point to the other in the scene. A similar approach was adopted by the face recognition methods of [15] and [16], who performed fine local analysis. However, these high-resolution polar sampling methods do not provide any information about global patterns.

In [4], we introduced the representation of face images in the polar frequency domain by global two-dimensional FBT features. The novelty of the proposal relied on the transformation of the image from the spatial domain to the polar frequency domain through the FBT and resulted in excellent face recognition performance. However, it still suffered from the deficiencies of the global approach, like low resolution of peripheral regions and sensitivity to partial occlusion.

In this paper, we present a novel approach and integration of new mechanisms that improved significantly the overall performance of the already state-of-the-art performance of the previous system, while making it robust against frequent face verification problems. The main contributions of the paper are:

- *Local approach*: we apply the FBT at three strategic locations [17,18] in the eyes region. This approach assumes that an intermediate working scale is more informative and robust than the completely global or local working scales.
- *Robustness to partial occlusion*: focusing on the eyes region improves stability against partial occlusions. We explored this fact by detecting possibly occluded face regions and excluding them from the luminance normalization pre-processing.
- *Robustness to illumination and age variation*: we present robustness evaluation on a subset with illumination variation and on two subsets with different age variations, besides the previous subset with expression variations.
- *Fully automatic system*: the previous algorithm required ground-truth information for the face normalization process. Now face and eye detection algorithms are integrated, making the verification system completely automatic when faces are not occluded.
- *Comparison with alternative representations*: FBT features are just one form of image representation in the polar frequency domain. Here we show that an alternative representation, the polar Fourier transformation, results in inferior performance.

## 3. Polar frequency analysis

The Fourier–Bessel (FB) series [19,20] found several applications in analyzing patterns in a circular domain [20,21] and is useful to describe the radial and angular components in images. Let $f(x,y)$ be the region of interest in the image. FBT analysis starts by converting the image coordinates from Cartesian $(x,y)$ to polar $(r,\theta)$. Let $(x_0,y_0)$ be the origin of the Cartesian image. The polar coordinates necessary to obtain the new image representation $f(r,\theta)$ are defined as $\theta=\tan^{-1}(y-y_0/x-x_0)$ and $r=\sqrt{(x-x_0)^2+(y-y_0)^2}$.

For square images, the considered maximum radius was the distance from the center of the region of interest to one of the corners. Radial resolution was fixed at one pixel width, but the angular resolution could be varied by increasing or reducing of the number of sampled radii. The intensity of each point of the $f(r,\theta)$ function was determined by bilinear interpolation [22].

The $f(r,\theta)$ function is represented by the two-dimensional FB series, defined as

$$f(r,\theta)=\sum_{i=1}^{\infty}\sum_{n=0}^{\infty}A_{n,i}J_n(\alpha_{n,i}r)\cos(n\theta)+\sum_{i=1}^{\infty}\sum_{n=0}^{\infty}B_{n,i}J_n(\alpha_{n,i}r)\sin(n\theta) \quad (1)$$

where $J_n$ is the Bessel function of order $n$, $f(R,\theta)=0$ and $0\le r\le R$. $\alpha_{n,i}$ is the $i$th root of the $J_n$ function, i.e. the zero crossing value satisfying $J_n(\alpha_{n,i})=0$ is the radial distance to the edge of the image. The orthogonal coefficients $A_{n,i}$ and $B_{n,i}$ are given by

$$A_{0,i}=\frac{1}{\pi R^2 J_1^2(\alpha_{n,i})}\int_{\theta=0}^{\theta=2\pi}\int_{r=0}^{r=R}f(r,\theta)rJ_n\left(\frac{\alpha_{n,i}}{R}\right)(r)\mathrm{d}r\mathrm{d}\theta \quad (2)$$

if $B_{0,i}=0$ and $n=0$;

$$\begin{bmatrix} A_{n,i} \\ B_{n,i} \end{bmatrix} = \frac{2}{\pi R^2 J_{n+1}^2(\alpha_{n,i})} \times \int_{\theta=0}^{\theta=2\pi} \int_{r=0}^{r=R} f(r,\theta) r J_n\left(\frac{\alpha_{n,i}}{R} r\right) \begin{bmatrix} \cos(n\theta) \\ \sin(n\theta) \end{bmatrix} \mathrm{d}r\mathrm{d}\theta \qquad (3)$$

if $n>0$

However, polar frequency analysis can also be done using other transformations. An alternative method is to represent images by polar Fourier transform descriptors. The polar Fourier transform is a well known mathematical operation where, after converting the image coordinates from Cartesian to polar, as described above, a conventional Fourier transformation is applied. These descriptors are directly related to radial and angular components, but are not identical to the coefficients extracted by the FBT.

## 4. Face verification using FBT

The proposed algorithm (Fig. 1) starts with image registration, occlusion detection, and normalization. After these steps, we extract the FB coefficients from the images, compute the pair-wise Cartesian distance between all the FBT-representations and represent each object by its distance to all other objects. In the last stage, we train a pseudo Fisher classifier. We tested this algorithm on the whole image (global analysis) and on the combination of three facial regions (local analysis).

### *4.1. Face registration, partial occlusion detection and normalization*

Face representation requires prior image registration and usually a spatial and luminance normalization pre-processing. Assuming the sample images contain a single face, we detected the head with a cascade of classifiers [23] and estimated the

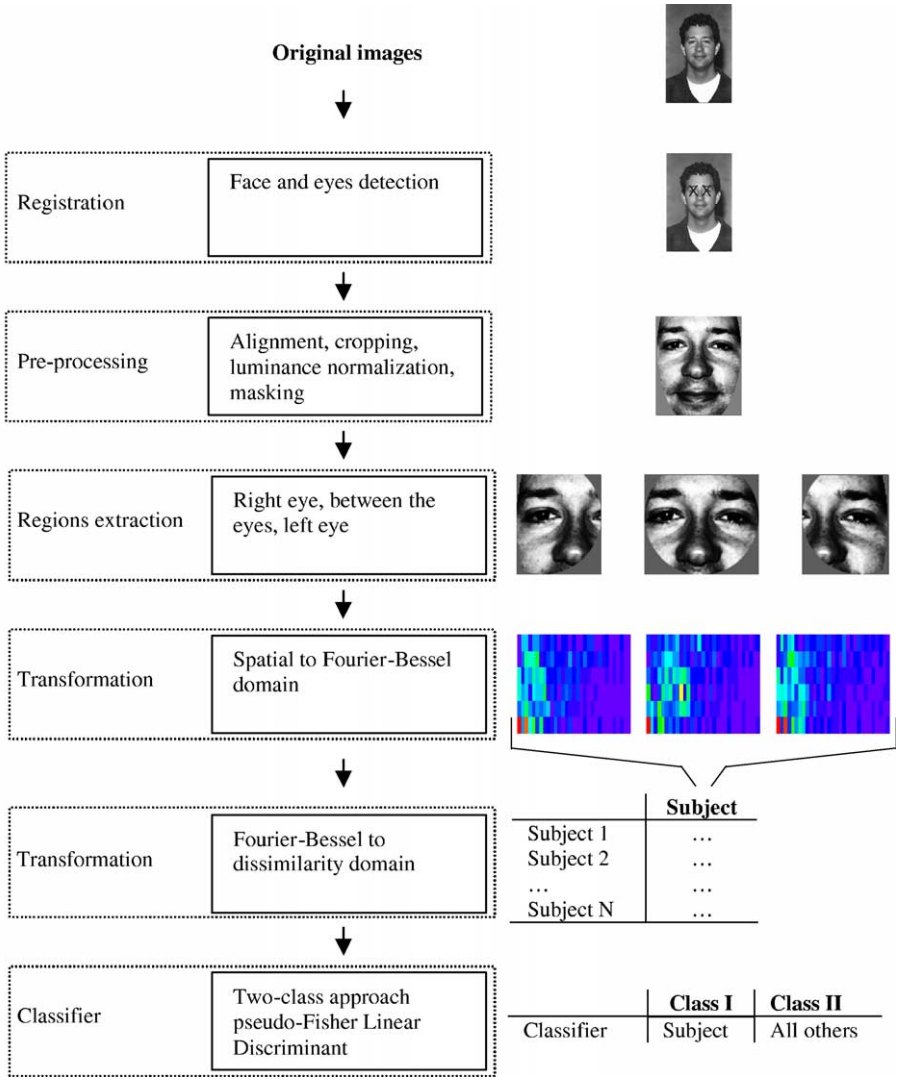


Fig. 1. Block diagram of the training phase in the proposed face verification system (left column) and an example case (right column). The modulus of the FB coefficients is represented in colored levels (red indicates the highest value, blue the lowest). Test images undergo through the same processing sequence, except for the last step that is substituted by a classification stage.

location of the eyes region with an active appearance model algorithm [24]. Within this region we used flow field information [25] to determine the eyes center. Using the eyes coordinates, we translated, rotated, and scaled the images so that the eyes were registered at specific pixels. Next, the images were cropped to a size of 130×150 pixels and a gray mask was applied to remove most of the hair and background. At this point we detected possibly occluded regions, as explained below, and masked them too. The unmasked region was histogram equalized and normalized to zero mean and a unitary standard deviation.

Face occlusion is one of the most difficult problems for face verification. Although our face and eye detection algorithms were not adapted to handle face occlusions, we were interested in preparing our system to deal with this face verification problem, assuming that the face detection problem was solved. Therefore, in the occlusion condition tests, we used ground truth information for face registration. Face occlusion was modeled as a homogeneous–luminance region of a connected area without any 'holes' (small regions within the homogeneous area) in it and of at least 650 pixels, i.e. approximately 3% of the total image area. Thus, we detected occluded regions by applying morphological operators to segment the image into flat zones. If the biggest region exceeded the 650 pixels size criteria and contained no 'holes,' it was considered as an occluded region. In that case, it was masked along with the presumably hair and background regions and not considered in the next luminance normalization step.

### 4.2. Spatial to polar frequency domain

Images were transformed by a FBT up to the 30th Bessel order and 6th root with angular resolution of 3°, thus obtaining 372 coefficients. These coefficients correspond to a frequency range of up to 30 and 3 cycles/image of angular and radial frequency, respectively. This frequency range was selected based on earlier tests [4] with the small-size Olivetti face database [26]. We tested the FBT descriptors of the whole image, as well as a combination of the upper right region, upper middle region, and the upper left region (Fig. 1). When using the polar Fourier transform, the angular sampling was matched and only coefficients related to the same frequency range covered by the FBT were used. Both amplitude and phase information were considered, as is the case of FBT.

### 4.3. Polar frequency to dissimilarity domain

Images were transformed from the FBT domain to a dissimilarity space as follows. Let the representation set $\mathbf{T}=\{\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_n\}$ refer to the $n$ training FBT images. Given an Euclidean distance matrix $\mathbf{D}\in\mathbf{R}^{n\times n}$ between those objects, where $D(t_i,t_j)$ is the Euclidean distance between $\mathbf{t}_i$ and $\mathbf{t}_j$, each image $\mathbf{t}_i$ is mapped as a vector of its disimmilarity to all the training images, i.e. $\mathbf{t}_i \rightarrow d(\mathbf{t}_i,\mathbf{T})=[D(\mathbf{t}_i,\mathbf{t}_1),D(\mathbf{t}_i,\mathbf{t}_2),\ldots,D(\mathbf{t}_i,\mathbf{t}_n)]$. This approach was formulated by [27] and is based on the assumption that the dissimilarities of similar objects to 'other ones' is about the same. Among other advantages of this representation space, by fixing the number of features to the number of objects, it avoids a well known phenomenon, where recognition performance is degraded as a consequence of the small number of training samples as compared to the number of features.

### 4.4. Classifier

We classified test images based on a pseudo Fisher linear discriminant (FLD) using a two-class approach [28]. A FLD is obtained by maximizing the (between subjects variation)/(within subjects variation) ratio [29]. Here we used a minimum-square error classifier implementation [30], which is equivalent to the FLD for two-class problems [29]. In these cases, after shifting the data such that it has zero mean, the FLD can be defined as

$$g(\mathrm{x}) = \left[d(\mathrm{x},\mathrm{T}) - \frac{1}{2}(\mathrm{m}_1-\mathrm{m}_2)\right]^T S^{-1}(\mathrm{m}_1-\mathrm{m}_2) \quad (4)$$

where $\mathbf{x}$ is a FBT probe image, $\mathbf{S}$ is the pooled covariance matrix, and $\mathbf{m}_i$ stands for the mean of class $i$. The probe image $\mathbf{x}$ is classified as corresponding to class-1 if $g(\mathbf{x})\geq 0$ and to class-2 otherwise. However, as the number of training objects and dimensions is the same in the dissimilarity space, the sample estimation of the covariance matrix $\mathbf{S}$ becomes singular, and the classifier cannot be built. One solution to the problem is to use a pseudo-inverse and augmented vectors [30]. Thus, Eq. (4) is replaced by

$$g(\mathrm{x}) = (d(\mathrm{x},\mathrm{T}),1)(d(\mathrm{T},\mathrm{T}),I)^{(-1)} \quad (5)$$

where $(d(\mathbf{x},\mathbf{T}),1)$ is the augmented vector to be classified, $d(\mathbf{T},\mathbf{T})$ is the training set dissimilarity matrix $[d(\mathbf{t}_1,\mathbf{T}),d(\mathbf{t}_2,\mathbf{T}),\ldots,d(\mathbf{t}_n,\mathbf{T})]^T$, and $(d(\mathbf{T},\mathbf{T}),I)$ is the augmented training set dissimilarity matrix. The inverse $(d(\mathbf{T},\mathbf{T}),I)^{(-1)}$ is the Moore–Penrose Pseudo-inverse, which gives the minimum norm solution. The pseudo-inverse relies on the singular value decomposition of the matrix $(d(\mathbf{T},\mathbf{T}),I)$ and it becomes the inverse of $(d(\mathbf{T},\mathbf{T}),I)$ in the subspace spanned by the eigenvectors corresponding to the non-zero eigenvalues. The classifier is found in this subspace.

The current $L$-classes problem can be reduced and solved by the two-classes solution described above. The training set was split into $L$ pairs of subsets, each pair consisting of one subset with images from a single subject and a second subset formed from all the other images. A pseudo-FLD was built for each pair of subsets. A probe image was tested on all $L$ discriminant functions, and a 'posterior probability' score was generated based on the inverse of the Euclidean distance to each subject.

## 5. Database, preprocessing, and testing procedures

We used the FERET database, due to its large number of individuals and strict testing protocols that allow precise performance comparisons between different algorithms [31]. Here, we compare our algorithm performance with a 'baseline' algorithm and with the published results of three successful approaches [32]. As a baseline algorithm we implemented a standard principal component analysis [1]. The principal

components were based on a set of 700 images selected randomly from the gallery subset. Not all 1196 images were used, due to the huge amount of random-access-memory that such operation requires. The first three principal components, which encode basically illumination variations [33], were excluded before projecting of the training and test images. The three other approaches are: Gabor wavelets combined with elastic bunch graph matching (EBGM) [2], localized facial features extraction followed by a Linear Discriminant Analysis (LDA) [34], and a Bayesian generalization of the LDA method [35]. Test results of these methods were obtained from the official FERET database site: www.itl.nist.gov/iad/humanid/feret/perf/eval.html.

In the FERET protocol, a *gallery* set of one frontal view image from 1196 subjects is used to train the algorithm and a different dataset is used as probe. All images are gray-scale $256\times384$ pixels size. We used the four probe sets, termed *FB*, *DupI*, *DupII* and *FC* [32]. The FB dataset is constituted of a single image from 1195 subjects, taken from the same subjects in the gallery set, after an interval of a few seconds, but with a different facial expression. There were no constrains or anotation of the type of facial expressions. The DupI and DupII datasets include 722 and 234 images, respectively. The DupI images were taken immediately or up to 34 months after the gallery images, while the images in DupII were taken at least 18 months after the gallery images. The FC subset contains 194 images of subjects taken with a different camera and different lighting. Fig. 2 shows a few example images with the different variations and the effect of face normalization and FB transformation.

The eyes coordinates were extracted automatically, as described in Section 4.1. Approximately 1% of the faces were not localized, in which cases the eyes region coordinates were set to a fix value derived from the mean of the located faces. The final mean error was $3.6\pm5.1$ pixels. In order to estimate the system performance under minimal localization errors and face occlusion, we executed a second series of experiments in which ground-truth information was used. The face registration was followed by a normalization step, as described in Section 4.1. The same pre-processing procedure was used in previous algorithms, except for the Gabor–EBGM system where a special normalization procedure was used.

The performance of the system was evaluated by verification tests according to the FERET protocol [31]. Given a gallery image $g$ and a probe image $p$, the algorithm verifies the claim that both were taken from the same subject. The verification probability $P_V$ is the probability of the algorithm accepting the claim when it is true, and the false-alarm rate $P_F$ is the probability of incorrectly accepting a false claim. The algorithm decision depends on the posterior probability score $si(k)$ given to each match and on a threshold $c$. Thus, a claim is confirmed if $si(k)\leq c$ and rejected otherwise. A plot of all the combinations of $P_V$ and $P_F$ as a function of $c$ is known as a receiver operating characteristic (ROC). $P_V$ and $P_F$ were calculated as the number of confirmations divided by the number of correct or incorrect matches, respectively. This procedure was repeated for 100 equally spaced threshold levels. Training and tests were done with the PRTools toolbox [36].

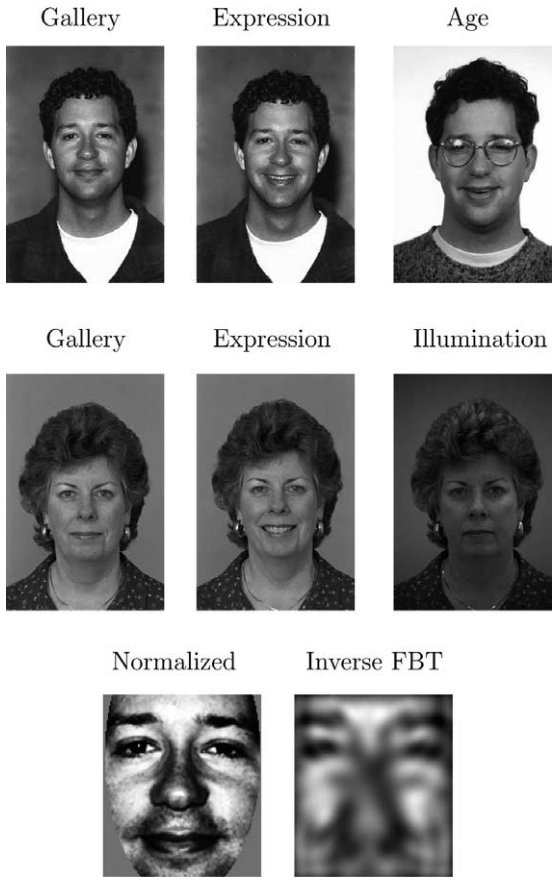


Fig. 2. 1st and 2nd rows: samples from the datasets. 3rd row: normalized face (gallery image from 1st row) and the global FB inverse transformation image.

## 6. Results

### 6.1. Semi-automatic system

Fig. 3 shows the performance of the proposed algorithm in the verification test with ground-truth information. This test is important for evaluation of the system performance without the interference of face localization errors and accurate comparison with previous algorithms. On the expression dataset the global and local FBT versions performed at about the same level as the best and second-best algorithms, respectively. On both age datasets the FBT algorithms outperformed the previous algorithms, with the local version being slightly superior. On the illumination dataset the global and local FBT algorithms were equal or better than the second-best previous algorithm (PCA + LDA). When the global and local algorithms were based on polar Fourier transform descriptors, instead of on FBT coefficients, a major loss of performance was observed in tests on all datasets, with the exception of the local version on the expression dataset. These results indicate that the PFT representation can achieve good performance, but it is very sensitive to age and illumination variations. We also computed the equal error rate of the proposed algorithms (Table 1). The equal error rate occurs at a threshold level where the incorrect rejection and false alarm rates are equals ($1-P_V=P_F$). Lower values indicate better performance. The equal error rate results

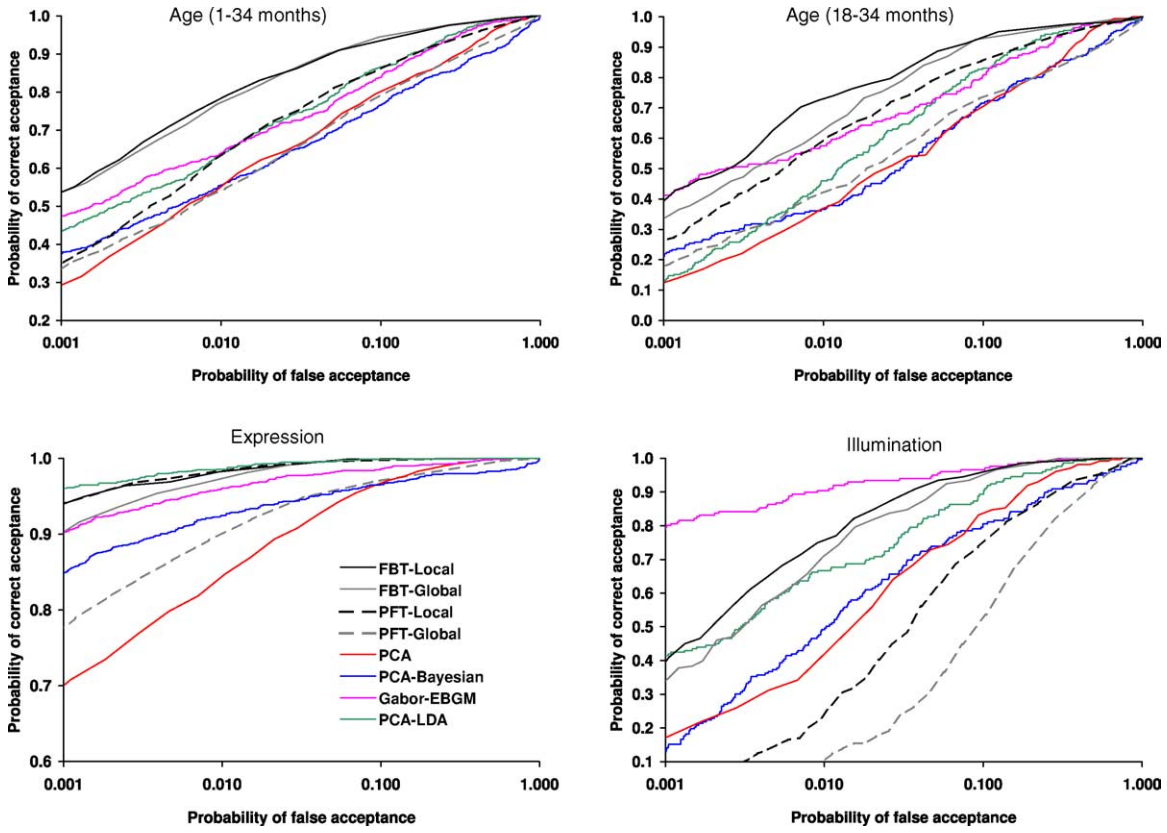


Fig. 3. ROC functions of the FBT, polar Fourier transform (PFT), principal component analysis (PCA), and previous algorithms on the age, expression and illumination subsets.

confirm the performance superiority of the FBT algorithms as shown by the ROC functions.

### 6.2. Partial occlusion

Local approaches for face recognition are in general more robust for occlusions (e.g. [37,38]) than global ones. We integrate a simple occluded regions detector in order to exclude these regions from the normalization pre-processing procedures, as described in Section 4.1. One can have an idea of the FBT features occlusion robustness potential by examining the effect of occlusion on a few sample images. We used two images from one subject and one image from another subject. One of the images from the former subject was partially occluded. Next, we extracted FB features from the three eye regions of the three images and computed the Euclidean distance between the occluded image and the other two images. Fig. 4 shows the results for one representative example. Without any occlusion, the two images from the same subject are closer than the images from different subjects. Occlusion linearly increases the distance between the images of the same subjects, but does not affect significantly the distance between images of different subjects.

Table 1
Equal error rates (%) of the FBT, polar Fourier transform (PFT), principal component analysis (PCA), and previous algorithms on the age, expression and illumination subsets

| Algorithm | Age (1–34 months) | Age (18–34 months) | Expression | Illumination |
|---|---|---|---|---|
| FBT-global | 7.7 | 8.3 | 1.7 | 7.4 |
| FBT-local | 7.3 | 8.3 | 1.4 | 5.9 |
| PFT-global | 16 | 20 | 4.2 | 23 |
| PFT-local | 12 | 12 | 1.4 | 16 |
| PCA | 16 | 20 | 5.6 | 14 |
| PCA + Bayesian | 18 | 21 | 4.9 | 18 |
| PCA + LDA | 13 | 13 | 1.2 | 10 |
| Gabor–EBGM | 13 | 14 | 2.5 | 5.1 |

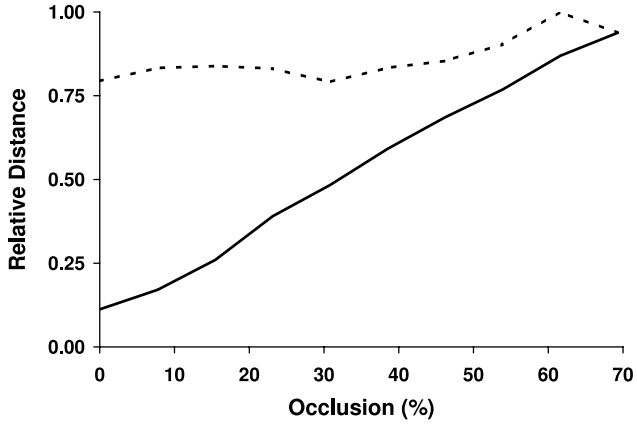


Fig. 4. Effect of occlusion (% of columns starting from the right-eye side) on the Euclidean distance between FBT occluded test image and a training image from the same (continuous line) or different (dashed line) subject.


Fig. 5. Examples of image occlusion: no occlusion, mouth+right eye (MO+RE), and mouth+nose (MO+NO).

As a result, only after occlusions of more than 40% does it starts to become difficult to identify the test image.

We evaluated the system robustness by occluding all the test images (but not the training images) by a gray mask that covered at least 50% of the 130×150 pixels face image. We tested two masking options: (1) masking of the mouth and right-eye regions and (2) masking of the mouth and nose regions (Fig. 5). Although this type of occlusion is not as realistic as using objects such as scarfs and sunglasses, it is efficient in simulating absence of spatial information. Fig. 6 shows the effect of occlusion on the performance of the global and local versions of the FBT algorithms. The local version was significantly affected when illumination variation was combined with occlusion, but was quite stable under expression and age variations. In contrast, the global version performed much worse under occlusion conditions on all subsets. These results confirm the advantage of the local over the global approach, and demonstrate the high robustness of the local-FBT under strong occlusion conditions combined with expression and age variations.

### 6.3. Fully automatic system

Fig. 7 shows the performance of the FBT algorithms with ground-truth information and when the eyes were detected automatically. The localization errors introduced in the latter case reduced the performance of the FBT algorithms up to 20%, approximately as it affected the PCA algorithm, which is known to be sensitive to this type of error [39]. The localization sensitivity of the proposed system is expected, considering the variance property of the FBT to translation [40]. It is interesting to notice, however, that under such conditions the advantage of the local over the global approach was reduced. Unfortunately, it is hard to compare the sensitivity of the current system to registration error with the sensitivity of previously published methods, since the authors do not make available their specific implementation, and when they perform such tests, they do not report the localization error levels of their face detector algorithm.

## 7. Discussion

We introduced a fully automated biologically-motivated local-based system for face verification tasks. The main empirical result of this study is the demonstration of the high performance of a verification system based on FBT descriptors,

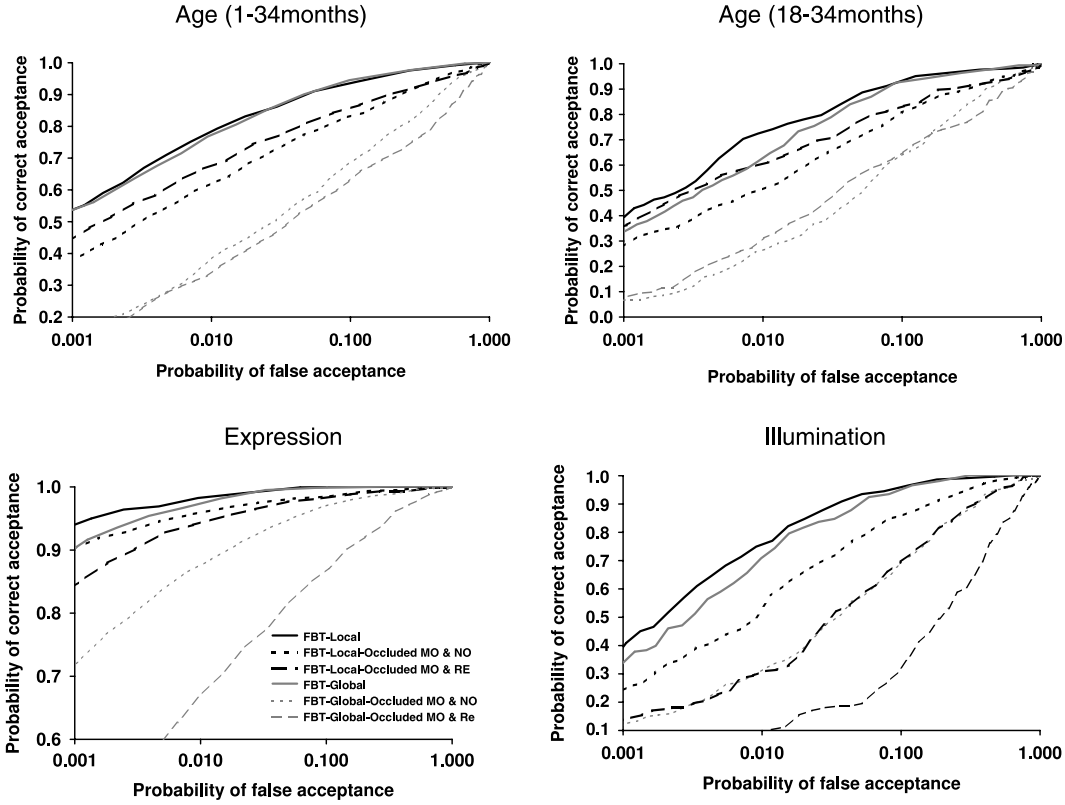


Fig. 6. ROC functions of the FBT on the occluded and not occluded age, expression and illumination subsets. Types of occlusion: mouth+right eye (MO+RE), and mouth+nose (MO+NO).

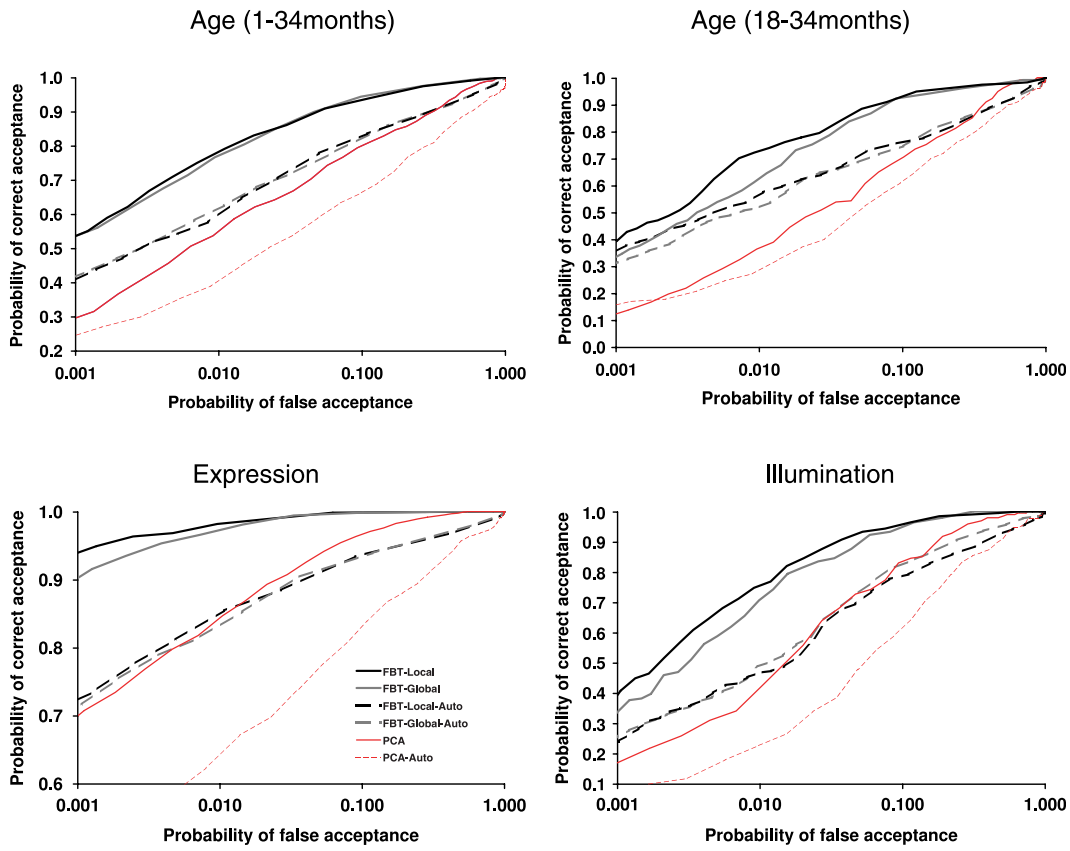


Fig. 7. ROC functions of the semi-automatic and automatic FBT and principal component analysis (PCA) algorithms on the age, expression and illumination subsets.

especially when these are extracted locally. Moreover, it was shown that a system based on an alternative polar frequency representation, namely the polar Fourier transform, is significantly less robust to common face variations. The significant advantage of the FBT approach in the experiments is an indication of the robustness of the polar features in realistic situations of face variations that exceeds simple facial expression, like illumination and age.

The superior behavior of the local approach was especially strong w.r.t. robustness to occlusion. In the local version, the mouth region is completely ignored, thus its occlusion or variation (e.g. due to a new beard or a scarf) does not affect performance at all. However, the local-FBT outperformed the global-FBT even when the occluded regions included face regions that were analyzed by the local version. Pilot tests indicated that the local-FBT algorithm fails only when more than 50% of the face image is occluded or when both eyes where covered, for example, by sunglasses (data not shown).

The property of robustness to occlusion of local analysis was explored by others. Local principal component analysis was used in [18,37,38] to detect occluded regions in face images. Test images were classified by comparing the unoccluded regions to corresponding regions in training images. However, the combination of FBT features and local approach has several advantages over that method besides performance. In our proposal, there is no need for special training strategies [37] or for training of specific classifiers for each testing image depending on the occluded region [38]. Finally, there is no need for any classification rule for the combination of the local features, as the FBT features form a single vector.

In the current study, occlusion was performed on the raw images. In contrast, [37] and [38] occluded normalized face images. This is an unrealistic condition, as their luminance normalization pre-processing considered information from occluded regions. It could be argued on similar ground that relying on ground-truth information in the occlusion condition is equally unrealistic. However, this is a qualitatively different situation, since even a partially occluded face still can be correctly registered, although at the moment that was shown only when a small region ($<13\%$) was occluded [37].

It is hard to compare our performance results with those obtained by [37,38], since their tests were performed on subsets of less than 100 images. The training and test images also did not included variations of expression, illumination or age. The algorithm of [37] was adapted in order to deal with expression variation by weighting differently local areas and assuming that the facial expression of the training images is known. In contrast, here we show that the proposed system can deal simultaneously with expression, illumination, and age variations, in addition to large scale occlusions.

The results indicate that significant performance gain of the automatic FBT method can be achieved by improving the eyes

localization algorithm. For example, [37] learned the subspace that represents localization errors within eigenfaces. This method can be adopted for the FBT subspace, with the advantage of the option to exclude from the final classification face regions that give high localization errors.

The relation of the present algorithm to human face recognition was not directly evaluated here, but a few associations can be done. As discussed in the introduction, there is clear evidence that the HVS extracts global radial and angular shape information, a fact that might look incompatible with the informative advantage of the local information pooling showed here. However, only little is known about the size of the global pooling area. A 1.2 visual degrees pooling area was suggested for the detection of Glass patterns [11], but the spatial locations and scale regarding face images remain open questions.

In the proposed system, the classifier operates in non-domain-specific metric space whose coordinates are similarity relations. The high performance achieved by this representation indicates that the 'real-world' proximity relations between face images are preserved to a good extent in the constructed internal space. It is possible that humans also use an analogous space to represent visual objects. This hypothesis was studied by correlating the distance between different shape objects by objective and perceptual parameters (see [41] for a review). Comparison of the two measurements is usually done by a multi-dimensional scaling analysis, which projects objects as points in a two-dimensional space where the distance between the points approximate the Euclidean distance between the original objects. For example, in one study [42] objective and perceptual sorting of face images were highly correlated, especially when the objective sorting used global features, such as age and weight of the persons in the images. Similar results were obtained in a neurophysiological study [43] in which monkeys were presented with face images. It was found that the multidimensional scaling analysis maps obtained from the original images and from the response patterns of neurons in the inferotemporal cortex had similar patterns. These results indicate that representing images in a dissimilarity space can be analogous to human representation mechanisms.

## 8. Conclusions

The proposed system combines several methods to achieve state-of-the-art face verification performance for expression and age variations, and robustness to wide occlusion. However, its robustness to registration errors and illumination variation should be still improved. The proposed method can also be useful for the understanding of human face processing and we are currently developing psychophysical experiments to establish the level of its relation to biological systems. Preliminary results [44] indicate a similar pattern of sensitivity of the human and FBT-based algorithm to polar frequency filtering.

## Acknowledgements

Y. Zana is grateful to FAPESP (03/07519-0). R. Cesar-Jr. is grateful to FAPESP (99/12765-2) and to CNPq (300722/98-2, 474596/2004-4).

## References

[1] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.
[2] L. Wiskott, J. Fellous, N. Kruger, C. VonDerMalsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 775–779.
[3] T. Sim, R. Sukthankar, M. Mullin, S. Baluja, Memory-based face recognition for visitor identification, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 214–220.
[4] Y. Zana, R.M. Cesar-Jr, Face recognition based on polar frequency features, ACM Transactions on Applied Perception 3 (1) (2006) 1–21.
[5] Y. Zana, R. Cesar-Jr, R.A. Barbosa, Automatic face recognition system based on local fourier-bessel features, in: XVIII Brazilian Symposium on Computer Graphics and Image Processing 2005, pp. 233–240.
[6] Y. Zana, R. Cesar-Jr, R. Feris, M. Turk, Face verification in polar frequency domain: a biologically motivated approach, in: Lecture Notes in Computer Science, vol. 3804, 2005, pp. 138–190.
[7] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (2003) 399–458.
[8] R.L. DeValois, K.K. DeValois, Spatial Vision, Oxford Sciences Publication., Oxford, 1990.
[9] E. Perret, D.I. Rolls, W. Caan, Visual neurons responsive to faces in the monkey temporal cortex, Experimental Brain Research 47 (1982) 329–342.
[10] J. Gallant, J.L. Braun, D. VanEssen, Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex, Science 259 (1993) 100–103.
[11] H. Wilson, F. Wilkinson, Detection of global structure in glass patterns: implications for form vision, Vision Research 38 (1998) 2933–2947.
[12] E. Schwartz, Spatial mapping in primate sensory projection: analytic structure and relevance to perception, Biological Cybernetics 25 (1977) 181–194.
[13] T. Grove, R. Fisher, Attention in iconic object matching, in: Proceedings of the British Machine Vision Conference, vol. 1, 1996, pp. 293–302.
[14] T. Hotta, K. Kurita, T. Mishima, Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image, in: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 70–75.
[15] M. Tistarelli, E. Grosso, in: H. Wechsler et al. (Ed.), NatoAsi Advanced Study on Face Recognition, F-163, Springer, Berlin, 1998, pp. 262–286. Ch. Active vision-based face recognition issues, applications and techniques.
[16] F. Smeraldi, J. Bigun, Retinal vision applied to facial features detection and face authentication, Pattern Recognition Letters 23 (2002) 463–475.
[17] B. Heisele, T. Koshizen, Components for face recognition, in: Sixth IEEE International Conference on Automatic Face and Gesture Recognition 2004 pp. 153–158.
[18] T.E. Campos, R.S. Feris, R.M. Cesar-Jr, Improved face × non-face discrimination using Fourier descriptors through feature selection, in: Proceedings of the 13th SIBGRAPI, IEEE Computer Society Press, 2000, pp. 28–35.
[19] F. Bowman, Introduction to Bessel functions, Dover Pub., New York, 1958.
[20] P. Fox, J. Cheng, J. Lu, Theory and experiment of fourier-bessel field calculation and tuning of a pulsed wave annular array, Journal of the Acoustical Society of America 113 (2003) 2412–2423.
[21] S. Guan, C. Lai, G. Wei, Fourier-bessel analysis of patterns in a circular domain, Physica D 151 (2001) 83–98.

[22] W.K. Pratt, Digital Image Processing, Wiley, London, 1991.
[23] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511–518.
[24] T. Cootes, G. Edwards, C. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 681–685.
[25] R. Kothari, J. Mitchell, Detection of eye locations in unconstrained visual images, in: IEEE International Conference on Image Processing, 1996, pp. 519–522.
[26] F. Samaria, A. Harter, Parametrization of a stochastic model for human face identification, in: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.
[27] R. Duin, D. DeRidder, D. Tax, Experiments with a featureless approach to pattern recognition, Pattern Recognition Letters 18 (1997) 1159–1166.
[28] D. Tax, R. Duin, Using two-class classifiers for multiclass classification in: C.S.R. Kasturi, D. Laurendeau (Eds.), Proceedings of the 16th International Conference on Pattern Recognition (ICPR), 2, IEEE Computer Society Press, Silver Spring, MD, 2002, pp. 124–127.
[29] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1990.
[30] M. Scurichina, R. Duin, Stabilizing classifiers for very small sample sizes, in: Proceedings of the 13th International Conference on Pattern Recognition, Track B, vol. 2 1996, pp. 891–896.
[31] P. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing Journal 16 (1998) 295–306.
[32] S. Rizvi, P. Phillips, H. Moon, The FERET verification testing protocol for face recognition algorithms, Tech. rep., NIST, technical report NISTIR 6281, 1998.
[33] P.W. Hallinan, A low-dimensional representation of human faces for arbitrarylighting conditions, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 1994, pp. 995–999.
[34] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, Journal of the Optical Society of America Am-Optics Image Science and Vision 14 (1997) 1724–1733.
[35] B. Moghaddam, T. Jebara, A. Pentland, Bayesian face recognition, Pattern Recognition 33 (2000) 1771–1782.
[36] R. Duin, PRTools3: A Matlab Toolbox for Pattern Recognition, Delft University of Technology, Delft, 2000.
[37] A. Martínez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 748–762.
[38] A. Lanitis, Person identification from heavily occluded face images., in: Proceedings of the 2004 ACM Symposium on Applied Computing (SAC), 2004, pp. 5–9.
[39] A. Lemieux, M. Parizeau, Experiments on eigenfaces robustness, in: Proceedings of the International Conference on Pattern Recognition, 2002, pp. 421–424.
[40] J. Cabrera, A. Falcón, F. Hernández, J. Méndez, A systematic method for exploring contour segment descriptions, Cybernetics and Systems 23 (1992) 241–270.
[41] S. Edelman, Representation and Recognition in Vision, MIT Press, Cambridge, 1999.
[42] G. Rhodes, Looking at faces: first-order and second-order features as determinants of facial appearance, Perception 17 (1988) 43–63.
[43] M. Young, S. Yamane, Sparse population coding of faces in the inferotemporal cortex, Science 256 (1992) 1327–1331.
[44] Y. Zana, R.M. Cesar-Jr. J.P. Mena-Chalco, Human and machine recognition of Fourier-Bessel filtered face images. Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition, 2006, pp. 299–304.