Pattern Recognition Letters 29 (2008) 1537–1543

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Face recognition using HOG–EBGM ☆

Alberto Albiol *, David Monzo, Antoine Martin, Jorge Sastre, Antonio Albiol

I-TEAM, Universidad Politecnica de Valencia, Spain

ARTICLE INFO

Article history:
Received 25 May 2007
Received in revised form 10 January 2008
Available online 7 April 2008

Communicated by H.H.S. Ip

Keywords:
Face recognition
EBGM
SIFT
HOG
Local image descriptors

ABSTRACT

This paper presents a new face recognition algorithm based on the well-known EBGM which replaces Gabor features by HOG descriptors. The recognition results show a better performance of our approach compared to other face recognition approaches using public available databases. This better performance is explained by the properties of HOG descriptors which are more robust to changes in illumination, rotation and small displacements, and to the higher accuracy of the face graphs obtained compared to classical Gabor–EBGM ones.

1. Introduction

Face recognition is a research field that has attracted much attention in the past years due to its many applications in public surveillance, video annotation, multimedia and others.

Many algorithms have been proposed for face recognition, we recommend (Zhao et al., 2003) for a complete survey on the topic. Face recognition techniques can be broadly classified into holistic and feature-based. Holistic methods, such as PCA (Turk and Pentland, 1991) or LDA (Belhumeur et al., 1996), project input faces onto a dimensional reduced space where recognition is carried out. These type of methods can be deemed as classical today. However, they are still very popular due to their simplicity and good performance. The greatest problem of eigen-based methods it that in some sense they assume that faces are rigid objects which can be reconstructed with linear combinations of a set of eigenfaces or fisherfaces, which is not true. Also when the illumination conditions change the performance of these methods degrades rapidly. To overcome these difficulties many variations of the original methods have been proposed (del Solar and Navarrete, 2005).

Feature-based methods try to recognize faces using its facial components: eyes, nose, mouth, etc. Active Appearance Model (Cotes et al., 2001) and Elastic Bunch Graph Matching (EBGM) (Wiskott et al., 1996) fall into this category. In EBGM faces are represented as graphs with nodes at facial landmarks (such as eyes, tip of the nose, etc.). Each node contains a set of Gabor wavelet coefficients, known as a jet. To increase robustness of EBGM to changes in expression and illumination, new approaches have been proposed. For example, in (Shin et al., 2007) a graph matching approach which replaces the original Gabor features is presented. Following this line of research, in this paper we present a EBGM algorithm in which Gabor features have been replaced by Histograms of Oriented Gradients (HOG) (Bicego et al., 2006) descriptors, which are inherited from the Scale Invariant Feature Transform (SIFT) proposed by Lowe (2004). SIFT has emerged as a cutting edge technology for extracting distinctive features from images, to be used in algorithms for tasks like matching different views of an object or scene. SIFT achieves invariance to scale changes by extracting keypoints at the local extrema of the scale-space representation of the image, then each keypoint is represented using histograms of image gradients, in the sequel HOG descriptor. HOG descriptors have also been proposed for pedestrian detection (Bay et al., 2006; Mikolajczyk and Schmid, 2005; Dalal and Triggs, 2005). In these approaches objects are assumed to be at a fixed scale and are divided into small connected regions at fixed positions. Then, for each region a HOG descriptor is obtained and the combination of these descriptors is used to represent the object.

SIFT has also been recently proposed for face recognition (Bicego et al., 2006), however this approach totally differs from ours. In the Bicego's algorithm keypoints are located at the local extrema of the scale-space as in the original Lowe's approach (Lowe, 2004). The main problem of this approach is that there is no control on the number, position and scale of the keypoints. However, in our algorithm the keypoints represent specific facial landmarks which are

☆ This work has been supported by the Technical University of Valencia: Programa de apoyo a la investigación y desarrollo PAID-06-06 and the CDTI Hesperia project.

* Corresponding author. Tel.: +34 96 387 73 09.
E-mail address: alalbiol@dcom.upv.es (A. Albiol).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.03.017

detected first as explained below. Once facial landmarks are detected we use HOG descriptors to represent them.

The rest of the paper is organized as follows. First, Section 2 presents in detail how HOG descriptors are built. Next, Section 3 describes our EBGM algorithm that uses HOG descriptors. Sections 4 and 5 show the experimental setup and recognition results. Finally some conclusions and future research are drawn in Section 6.

Henceforth we will use the term Gabor–EBGM, or simply EBGM, when referring to the original EBGM algorithm presented in (Wiskott et al., 1996) while we will use the term HOG–EBGM when referring to the algorithm presented in this paper.

2. HOG descriptors

As mentioned previously SIFT has emerged as one of the most used detection/description schemes for its ability to handle image transformations like scale changes (zoom), image rotation, and illumination. The major steps of the SIFT algorithm are:

(1) Scale-space extrema detection.
(2) Orientation assignment.
(3) Keypoint descriptor.

The first step is used by SIFT to achieve invariance to scale changes. This is done by extracting SIFT features only at the local extrema of the scale-space representation of the image. The next step aims to obtain image rotation invariance. To that end, at each extrema of the scale-space representation, SIFT finds the dominant orientation using image gradient information and then, all image gradients are made relative to this dominant direction.

While these two techniques have proved to be very useful for images that are arbitrarily scaled or rotated, the fact is that these normalization stages remove information which might be useful for recognition when images are not scaled or rotated. In this paper, we assume that the exact location of both eyes is known a priori. To detect the eyes precisely, we have developed an algorithm that uses a mixed approach of boosted classifiers (Viola and Jones, 2001) and again HOG descriptors. However this problem can be deemed as precise face localization and it is not treated here. Since the exact location of the eyes is used to normalize faces, we do not expect any changes in either scale or rotation. For this reason, we skip the two first steps of the SIFT algorithm and only adopt the last step from Lowe's approach, the keypoint descriptor. This keypoint descriptor is also called HOG in the literature.

The HOG descriptor is a local statistic of the orientations of the image gradients around a keypoint. More formally, each descriptor is a bundle of histograms composed of pixel orientations given by their gradients. The number of possible orientations (histogram bins) is referred to as N_o. Each histogram in the bundle describes a specific area around the keypoint. These areas correspond to the cells of a $N_p \times N_p$ squared grid centered on the keypoint (see Fig. 1). The original paper (Lowe, 2004) sets the parameters of the descriptor to $N_p = 4$ cells for each spatial direction and $N_o = 8$ bins for each histogram in the bundle resulting in a total of $N_p^2 N_o = 128$ elements in a HOG descriptor.

In our work, each spatial cell is a square of 5×5 pixels. This size is chosen accordingly to the distance between eyes of the normalized faces, which in our work is 40 pixels. The results presented in Section 5.1.1 will further justify this selection.

Similar to Lowe's original approach, the contribution of each pixel gradient to the histogram is weighted by the gradient modulus and a Gaussian window. The Gaussian window is centered at the keypoint coordinates and its standard deviation equals to half the extension of the spatial range, which is 10 pixels. Also the pixel contribution is distributed into adjacent spatial cells and orientations bins using trilinear interpolation. This is important to avoid all boundary effects in which the descriptor abruptly changes as a sample shifts smoothly from being within one cell to another or from one orientation to another. Gaussian windowing and trilinear interpolation also increases the robustness of the descriptor against small displacements of the keypoint location.

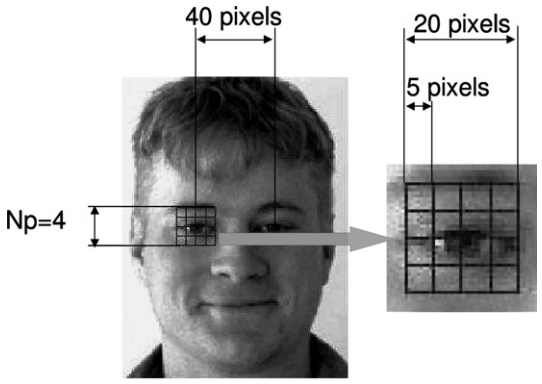

Fig. 1. Normalized face and the spatial cells of the right eye HOG descriptor.

Finally, HOG descriptors are normalized to increase invariance to illumination changes. As in the Lowe's algorithm, first the 128 elements vector are normalized to unit length. This normalization cancels changes in image contrast. Notice that we do not care about changes in brightness, a constant added to pixel values, because they are suppressed by image gradients. Finally, the descriptor is saturated so that no values over 0.2 are allowed and again re-normalized to unit length. This final step is done to reduce non-linear illumination changes.

3. Elastic bunch graph matching

The main idea of EBGM is that a novel face pattern can be recognized by first localizing a set of facial landmarks and then measuring the similarity between these landmarks and those extracted from a set of faces of each individual. Traditionally, EBGM algorithms use Gabor jets as features for both localization and matching of facial landmarks. In our approach, we replace Gabor coefficients by the HOG descriptors presented in Section 2.

Our implementation of EBGM is based on the algorithm developed by Wiskott et al. (1996) which was included by the Colorado State University (CSU) as a baseline algorithm for comparison of face recognition algorithms (Bolme et al., 2003). Basically, this HOG–EBGM algorithm can be decomposed into three steps:

(1) Image normalization.
(2) Creation of face graphs.
(3) Graph matching.

The objective of the image normalization step is to reduce variability produced by changes in illumination, scale and rotation.

The next step creates a face graph after the detection of facial landmarks. Of course, the success of the recognition algorithm depends on a good selection of facial landmarks. More precisely, facial landmarks need to be very distinctive between different people and also be relatively easy to detect in a fully automatic system. Our face graph follows the structure proposed in the CSU project (Bolme, 2003) with 25 facial landmarks which are shown in Fig. 2, the numbers indicate the search order.

It is important to mention, that not all facial areas contribute equally to face recognition. Several studies (Zhao et al., 2003) have shown that the area around eyes and nose are very important for

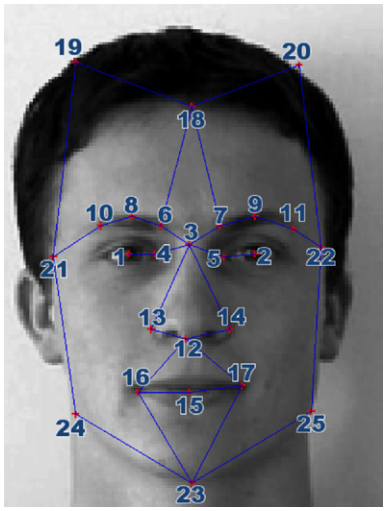

Fig. 2. Face graph and its 25 facial landmarks on a CVL image. The numbers indicate also the search order.

recognition, for this reason more facial landmarks are placed in these two areas.

Next subsections explain in detail the steps of our HOG–EBGM approach.

3.1. Image normalization

As introduced above, in this paper we assume that the exact location of both eyes is known a priori. This simplifies the normalization task since it only deals with scale and rotation variations. The normalized face is a 120×160 pixels image in which left and right eyes are located at (40,80) and (80,80) pixel coordinates, respectively. Normalization in scale is particularly important prior to the computation of image gradients.

3.2. Modeling and location of facial landmarks

In EBGM, each face is described by a face graph (FG), composed of strategically located keypoints (facial landmarks) and their corresponding descriptors. More formally, given a face image, its face graph (FG) is the set of its facial landmark coordinates and their associated HOG descriptors:

$$\mathrm{FG} = \{X_i, J_i = \mathrm{HOG}(X_i), 1 \leqslant i \leqslant 25\}$$

To automatically locate facial landmarks in new faces, a model for each landmark is needed. These models should account for changes of expression, hair styles, illumination, etc. Similar to the original EBGM scheme, we model each facial landmark using a set of HOG descriptors called a *bunch*. These HOG descriptors are manually obtained from a set of N_f normalized training faces at each particular keypoint. In the sequel, $\mathrm{fbg}_i(k)$, $1 < i < 25$ and $1 < k < N_f$, stands for the HOG descriptor of the ith keypoint from the kth training face (remember that face graphs are composed of 25 keypoints). Finally, we call face bunch graph (FBG) to the set of all keypoints models:

$$\mathrm{FBG} = \{\mathrm{fbg}_i(k), 1 \leqslant i \leqslant 25, 1 \leqslant k \leqslant N_f\}$$

The FBG is used to automatically build new FGs in an iterative process that uses the already detected landmarks to reduce the search area. The order in which new facial landmarks are searched is set empirically to produce the best results. The idea is that since we start from eye locations, the points closer to the eyes are detected first.

The process to detect the ith $(i > 2)$ facial landmark is the following:

(1) Initial estimation of the facial landmark location, X_i^s. This estimate is based on the mean of displacements between the ith keypoint and the jth $(j < i)$ keypoints. More in detail:
   (a) Let $d(i,j)$ be the mean displacement between keypoints i and j estimated using the data in the FBG.
   (b) Let X_j $(j < i)$ be the coordinates of the jth keypoint which has been already located.
   (c) For each X_j, we define $X_i(j) = X_j + d(i,j)$ as the initial prediction of the ith keypoint based on the jth keypoint
   (d) The initial estimate of $X_i^s = \frac{1}{i-1}\sum_{j<i} X_i(j)$, i.e. the mean of the estimates of previous keypoints.

(2) Calculate the HOG descriptor on the previous location, $\mathrm{HOG}(X_i^s)$.

(3) Compare $\mathrm{HOG}(X_i^s)$ with the $\mathrm{fbg}_i(k)$, $0 \leqslant k \leqslant N_f$ in the FBG and let:

$$k_{\min} = \min_k \|(\mathrm{HOG}(X_i^s) - \mathrm{fbg}_i(k)\|$$

(4) Define a search area S_i around X_i^s. The extent of the search area depends on the particular keypoint as shown in Fig. 3. We empirically set the search areas considering the dispersion of the location of facial landmarks in the FBG for each keypoint.

(5) Refine the initial estimate of the ith keypoint using the descriptor $\mathrm{fbg}_i(k_{\min})$:

$$X_i = \min_{X \in S_i} \|\mathrm{HOG}(X) - \mathrm{fbg}_i(k_{\min})\|$$

Note that in this paper, comparison between HOG descriptors is always done using euclidean distance.

3.3. Creation and distance between face graphs

To compare two face graphs FG^k and FG^l from two different faces, we just sum up the distances between corresponding keypoint descriptors J_i^k and J_i^l:

$$D_{kl} = |FG^k, FG^l| = \sum_{i=1}^{25} \|J_i^k - J_i^l\|$$

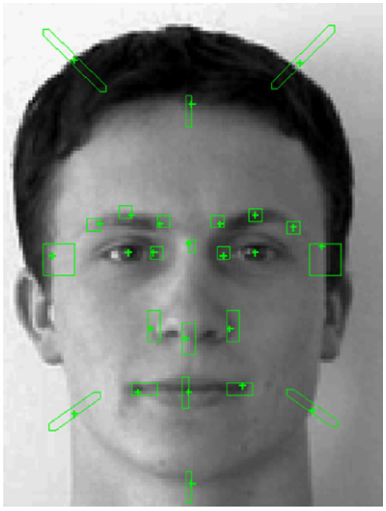

Fig. 3. Search area for each face landmark and detected keypoints.

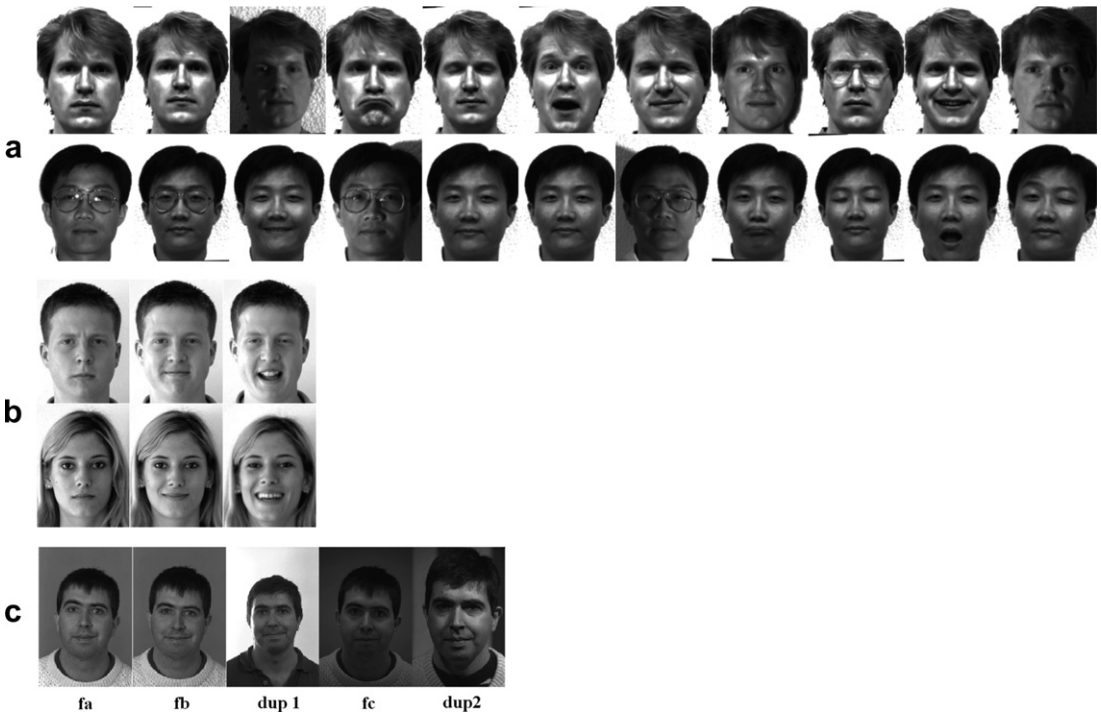

Fig. 4. Examples of subjects from (a) Yale database, (b) CVL database and (c) FERET database.

4. Experimental setup

All the experiments carried out in this paper have been done using three sets of images namely: the Yale database (Yale database, 1997), a subset of the CVL face database (Peer, 1999) and the FERET face database (Phillips et al., 2000). The Yale database contains 165 frontal images of 15 subjects (11 images/subject) showing different expressions and illumination conditions as illustrated in Fig. 4a. The CVL database contains seven images from 114 individuals under uniform illumination and different orientations around the vertical axis. In this work, we only use the frontal faces of the CVL (3 images/subject). Fig. 4b shows the selected images of the faces from two representative subjects of the CVL database. The FERET database contains 3365 full frontal facial images of nearly 1000 subjects. FERET database images are organized into a gallery set (*fa*) and four probe sets (*fb*,*fc*,*dup1*,*dup2*) as illustrated in Fig. 4c. Using the FERET terminology (Phillips et al., 2000) the gallery is the set of known facial images and the probe is the set of faces to be identified. The images in sets *fa* and *fb* were taken in the same session with the same camera and illumination conditions but with different facial expressions. The *fc* images were also taken in the same session but using a different camera and different lighting. Finally sets *dup1*, *dup2* are by far the most challenging sets. These images were taken on a later date, sometimes years apart, and the photographers sometimes asked the subjects to put on their glasses and/or pull their hair back. The reader can see (Phillips et al., 2000) for further details.

In this paper, we have used the Yale and CVL databases for tuning the parameters of our HOG–EBGM algorithm whereas the FERET database is used as a common framework to compare and evaluate our approach respect to other face recognition approaches.

Finally, it should be said that as a preparation task, we manually marked the 25 facial landmarks in all images of the Yale and CVL databases. This data is used to build the FBG as described in Section 3.2 and in the experiment of Section 5.1.1.

5. Experiments and results

Two kinds of experiments have been carried out:

Table 1
Influence of the feature window size on the recognition rate

	Size of the HOG window				
	12 × 12	16 × 16	20 × 20	24 × 24	28 × 28
Yale database	98.2%	97.6%	97.0%	95.8%	94.6%
CVL database	96.2%	98.5%	99.1%	98.8%	98.0%

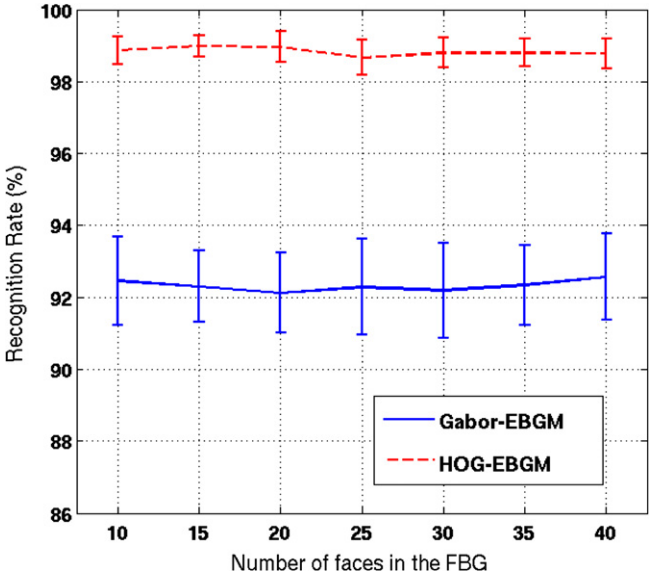

Fig. 5. Performance of Gabor–EBGM and HOG–EBGM for different number of images in the FBG on the CVL database.

(1) Study the influence and tuning of the HOG–EBGM parameters.
(2) Comparison of HOG–EBGM against the algorithms provided by the CSU Face Identification Evaluation System (Bolme et al., 2003).

5.1. Tuning of HOG–EBGM parameters

In this section, we study the influence on the recognition rate of the window size of HOG descriptors and the influence of the number of training faces N_f of the FBG. The study has been performed using a leave-one-out methodology on the CVL and Yale databases. In each experiment round we take one face image which is used as a probe set and the remaining images of the database are used as the gallery set. Then, the image is recognized using a nearest neighbor classifier. This process is repeated for all faces in the database and the average is the recognition rate.

5.1.1. Influence of the window size of HOG descriptors

In this experiment, we study the influence of the size of the HOG descriptor window. To that end we obtain recognition rates using the Yale and CVL databases. To make this study independent of the building process of face graphs, all facial landmarks in this experiments are manually marked. The window sizes range from 12×12 to 28×28. This range takes into consideration that the

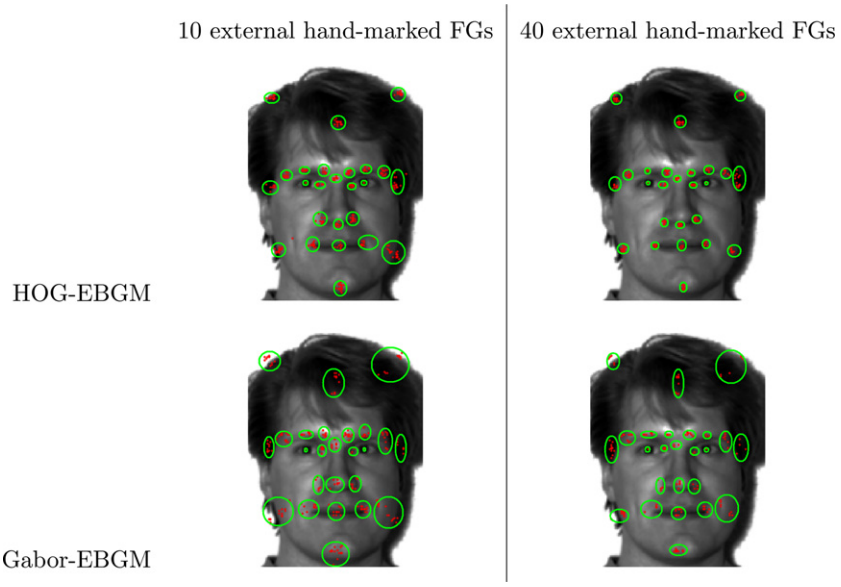

Fig. 6. Dispersion of automatically detected facial landmarks for 10 and 40 images in the FBG. Dispersion is obtained after 20 trials using random sets of images in the FBG.

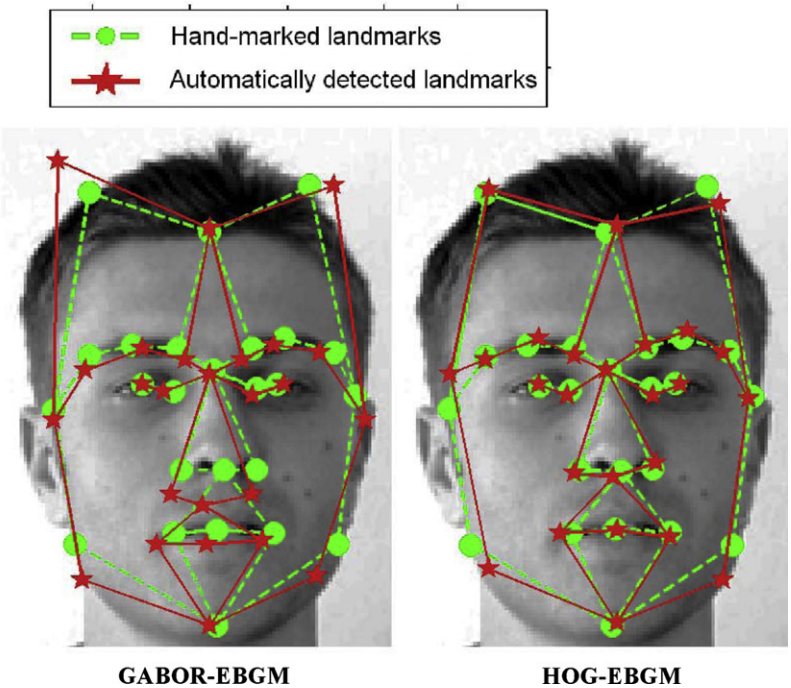

Fig. 7. Comparison of hand-marked and automatic face graphs for both EBGM algorithms.

distance between eyes on the normalized face is always 40 pixels. As the reader will notice, window sizes are multiple of 4 because our HOG descriptors use a 4×4 grid (Fig. 1).

The recognition rates are presented in Table 1. It can be seen that 20×20 gives a maximum when the CVL data is used while the performance for the Yale database tends to increase as the window size decreases. This can be explained because smaller window sizes make the features more robust to illumination changes (which are more noticeable in the Yale database). For this reason, we have selected a 20×20 window as a trade off value in the sequel experiments.

5.1.2. Study of the influence of the number of images on the FBG

As described in Section 3.2, our HOG–EBGM approach uses a FBG to automatically detect facial landmarks. The FBG is composed of a set of N_f training faces which are used to model the facial landmarks. In this experiment we study the influence of the parameter N_f on the recognition rate. This recognition rate is also compared with the one obtained using standard Gabor–EBGM. Here, the FBG is built using a random set of N_f training images from the Yale database and the recognition rate is evaluated on the CVL database.

Fig. 5 shows the recognition rate when N_f changes from 10 to 40. Also to check if the algorithm is dependent on the particular set of N_f images, we make 20 trials with different sets of random images for each value of N_f. The curves represent the mean recognition rate along with their corresponding standard deviation.

From the results, we can notice that the recognition rate is quite independent of N_f. This fact is also illustrated in Fig. 6. In this figure we show an example of the dispersion of facial landmarks after the 20 trials. We can see that this dispersion is only slightly decreased

Table 2
Recognition rate for PCA Euclidean, PCA Mahalanobis Cosine, LDA, Bayesian MAP, Bayesian ML, Gabor–EBGM and HOG–EBGM using the FERET database

	PCA Euclidean	PCA Mahalanobis Cosine	LDA	Bayesian MAP	Bayesian ML	Gabor–EBGM	HOG–EBGM
fafb	74.3%	85.3%	72.1%	81.7%	81.7%	87.3%	95.5%
fafc	5.6%	65.5%	41.8%	35.0%	34.5%	38.7%	81.9%
dup1	33.8%	44.3%	41.3%	50.8%	51.5%	42.8%	60.1%
dup2	14.1%	21.8%	15.4%	29.9%	31.2%	22.7%	55.6%

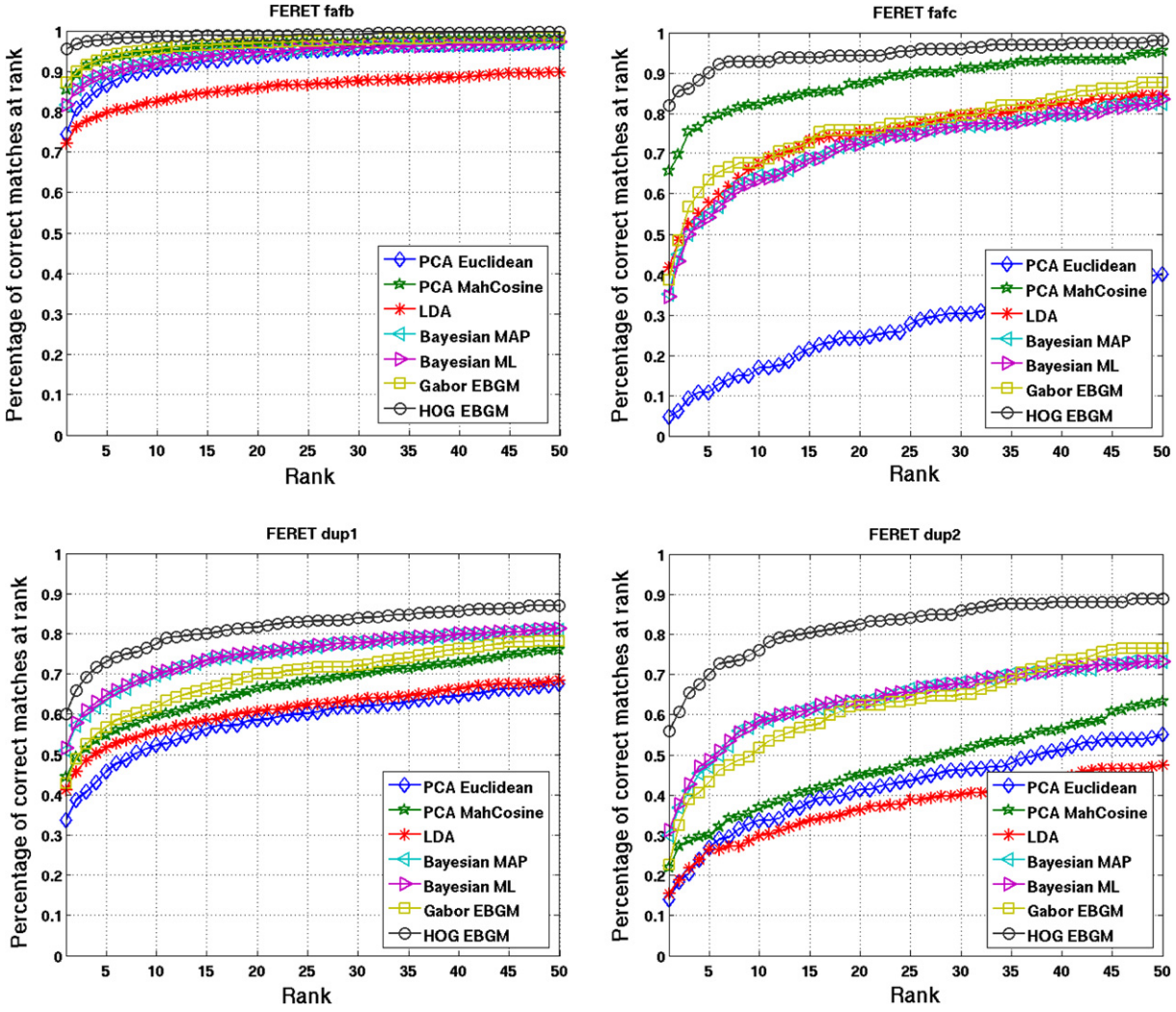

Fig. 8. Rank curves on the FERET database for *fafb*, *fafc*, *dup1* and *dup2* experiments.

when the number of faces is set to 40. It can also be seen that the dispersion for the HOG–EBGM is lower than for the Gabor–EBGM which can explain part of the better recognition rate of our approach.

Also from the example in Fig. 7 we can see that the FGs located with HOG–EBGM adjust better with their corresponding ideal graphs than the FGs located with Gabor–EBGM. This partly explains the better results in Section 5.2.

From the previous results we set to 10 the number of model images in the FBG for subsequent experiments.

5.2. Comparison to other algorithms

In order to evaluate the performance rates of our approach, we have used the CSU Face Identification Evaluation System (Bolme et al., 2003), which also includes other face recognition algorithms for comparison. The evaluation has been performed on the FERET database using the *fa* set as a gallery and the *fb*, *fc*, *dup1* and *dup2* as probe sets. The algorithms compared in this experiment are:

- Principal Component Analysis (PCA) (Turk and Pentland, 1991), considering Euclidean and Mahalanobis Cosine distances.
- Linear Discriminant Analysis (LDA) (Belhumeur et al., 1996).
- Bayesian algorithm with variants MAP and ML (Moghaddam et al., 1996).
- Original Gabor–EBGM algorithm (Wiskott et al., 1996).
- Our HOG–EBGM algorithm.

The reader can see the references for implementation details of these algorithms.

We can see in Table 2 the recognition rates obtained for all the experiments. FERET also uses rank curves to compare face recognition algorithms. A rank curve shows for each rank k the probability that the test face is between the first k nearest faces. Fig. 8 shows the rank curves for the *fafb*, *fafc*, *dup1* and *dup2* experiments.

The results presented in Table 2 and Fig. 8 show that our HOG–EBGM performs better than the other recognition algorithms with all probe sets. This is particularly true for the most difficult probe sets *dup1* and *dup2*.

6. Conclusions and future research

This paper presents a new face recognition algorithm based on the well-known EBGM method which replaces Gabor features by HOG descriptors. The recognition results show a better performance of our approach compared to other face recognition approaches using public available databases. This better performance is explained by the properties of HOG descriptors which are more robust to changes in illumination, rotation and small displacements, and to the higher accuracy of the face graphs obtained compared to classical Gabor–EBGM ones.

Future research is focused on improving the classification algorithm (in this work we use nearest neighbor). Also the influence of each particular landmark on the recognition rate will be studied so that landmarks can be properly weighted or resized.

References

Bay, H., Tuytelaars, T., Gool, L.V., 2006. SURF: Speeded up robust features. In: Proc. 9th European Conf. on Computer Vision, Graz, Austria.

Belhumeur, P., Hespanh, J., Kriegman, D., Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In: Proc. 4th European Conf. on Computer Vision, Cambridge, UK, 1996, pp. 45–58.

Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M., 2006. On the use of SIFT features for face authentication, in: Proc. Internat. Conf. on Computer Vision and Pattern Recognition Workshop, New York.

Bolme, D., 2003. Elastic bunch graph matching, Ph.D. thesis. Colorado State University, Fort Collins, Colorado.

Bolme, D.S., Beveridge, J.R., Teixeira, M., Draper, B., 2003. The CSU face identification evaluation system: Its purpose, features, and structure. In: Proc. Internat. Conf. on Computer Vision Systems, Graz, Austria, pp. 304–313.

Cotes, T., Edwards, G.J., Taylor, C.J., 2001. Active appreance models. IEEE Trans. Pattern Anal. Machine Intell. 23 (6), 681–685.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proc. 9th European Conf. on Computer Vision, San Diego, CA.

del Solar, J.R., Navarrete, P., 2005. Eigenspace-based face recognition: A comparative study of different approaches. IEEE Trans. Systems Man Cybernet. 35 (3), 315–325.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60 (2), 91–110.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27 (10), 1615–1630.

Moghaddam, B., Nastar, C., Pentland, A., 1996. A bayesian similarity measure for direct image matching. In: Proc. Internat. Conf. on Pattern Recognition, Vol. 2, Vienna, Austria, pp. 350–358.

Peer, P., 1999. CVL Face database, University of Ljubjana. <http://www.fri.uni-lj.si/en>.

Phillips, J.P., Moon, H., Rizv, S., Rauss, P.J., 2000. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Machine Intell. 22 (10), 1090–1104.

Shin, H., Kim, S.D., Choi, H.C., 2007. Generalized elastic graph matching for face recognition. Pattern Recognition Lett. 28 (9), 1077–1082.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cognitive Neurosci. 3 (1), 71–86.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: Proc. Internat. Conf. on Computer Vision and Pattern Recognition, Hawaii.

Wiskott, L., Fellous, J.M., Kruger, N., Malsburg, C., 1996. Face recognition by elastic bunch graph matching, Tech. Rep., Ruhr-Universitat Bochum.

Yale database, 1997. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P., 2003. Face recognition: A literature survey. ACM Comput. Surv. 35 (4), 399–458.