Pattern Recognition Letters 32 (2011) 1598–1603

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Face recognition using Histograms of Oriented Gradients

O. Déniz [a,*], G. Bueno [a], J. Salido [a], F. De la Torre [b]

[a] Universidad de Castilla-La Mancha, E.T.S. Ingenieros Industriales, Avda. Camilo Jose Cela s/n, 13071 Ciudad Real, Spain
[b] Carnegie Mellon University, Robotics Institute, 211 Smith Hall, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 24 September 2009
Available online 20 January 2011
Communicated by H. Wechsler

Keywords:
Face recognition
Histograms of Oriented Gradients
Active Appearance Models

ABSTRACT

Face recognition has been a long standing problem in computer vision. Recently, Histograms of Oriented Gradients (HOGs) have proven to be an effective descriptor for object recognition in general and face recognition in particular. In this paper, we investigate a simple but powerful approach to make robust use of HOG features for face recognition. The three main contributions of this work are: First, in order to compensate for errors in facial feature detection due to occlusions, pose and illumination changes, we propose to extract HOG descriptors from a regular grid. Second, fusion of HOG descriptors at different scales allows to capture important structure for face recognition. Third, we identify the necessity of performing dimensionality reduction to remove noise and make the classification process less prone to overfitting. This is particularly important if HOG features are extracted from overlapping cells. Finally, experimental results on four databases illustrate the benefits of our approach.

1. Introduction

Face recognition has been a long standing problem in computer vision. It has recently attracted significant attention due to the accessibility of inexpensive digital cameras and computers, and its applications in biometrics and surveillance (see Zhao et al. (2003); Chellappa et al. (1995); Samal and Iyengar (1992); Chellappa and Zhao (2005) for recent surveys of face recognition).

Central to the success of face recognition are the feature representation and the classification method. In this paper, we will focus on the former. Broadly speaking, we could classify the features for face recognition as geometric or photometric (view based). The latter seem to have prevailed in the literature (Zhao et al., 2003). There exist a large number of features, starting from the influential Eigenfaces (Principal Component Analysis) (Turk and Pentland, 1991), Gabor wavelets (Amin and Yan, 2009), Local Binary Patterns (Ahonen et al., 2004), Error-Correcting Output Codes (Kittler et al., 2001) and Independent Component Analysis (ICA) (Bartlett et al., 2002) among others.

Histograms of Oriented Gradients (HOGs) (Lowe, 2004) are image descriptors invariant to 2D rotation which have been used in many different problems in computer vision, such as pedestrian detection (Bertozzi et al., 2007; Wang and Lien, 2007; Chuang et al., 2008; Watanabe et al., 2009; Baranda et al., 2008; He et al., 2008; Kobayashi et al., 2008; Suard et al., 2006; Zhu et al., 2006; Perdersoli et al., 2007a,b). Recently, in (Albiol et al., 2008) the authors successfully applied HOG descriptors to the problem of face recognition. In that work, a set of 25 facial landmarks were first localized using the Elastic Bunch Graph Matching framework (see Wiskott et al., 1997). The HOG features extracted from the vicinity in each of these 25 facial landmarks were used for classification, using nearest neighbor and Euclidean distance. In this paper, following (Albiol et al., 2008), we further explore the representational power of HOG features for face recognition, and propose a simple but powerful approach to build robust HOG descriptors. In particular, three are the main novelties: (1) build the HOG descriptor using a regular grid, (2) build and combine the HOG descriptors at different scales, (3) apply a linear dimensionality reduction to remove noise, make the classifier more efficient (i.e. reduce dimensionality) and less prone to overfitting. Our results in four standard face databases support the proposed method.

The rest of the paper is organized as follows. Section 2 describes HOG in detail, as well as our approach. In Section 3 we describe the experimental validation. Section 4 finalizes the paper with the main conclusions.

2. Building a representation for face recognition using HOGs

The algorithm for extracting HOGs (see Dalal and Triggs, 2005; Lowe, 2004) counts occurrences of edge orientations in a local neighborhood of an image. In our case, the image is first divided into small connected regions, called cells, and for each

* Corresponding author. Fax: +34 926295361.
E-mail addresses: oscar.deniz@uclm.es, odeniz@dis.ulpgc.es (O. Déniz).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.01.004

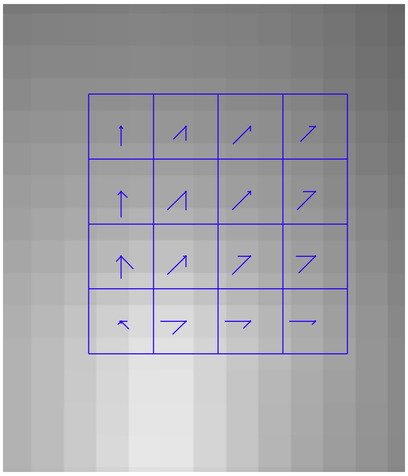

Fig. 1. Example HOG descriptors, patch size=8 × 8. Each cell of the patch shows the orientation of the gradients.

Table 1
Four face databases used in the experiments.

Name	Classes	Total samples	Samples per class (min/avg/max)	Variations present
FERET	1195	3540	2/2.9/32	Facial expression, aging of subjects, illumination
MPIE-2	337	2509	2/7.4/11	Expression, session
AR	132	3236	13/24.5/26	Expression, illumination, occlusions (sunglasses and scarves)
Yale	15	165	11	Expression, illumination, glasses

cell a histogram of edge orientations is computed. The histogram channels are evenly spread over 0–180° or 0–360°, depending on whether the gradient is 'unsigned' or 'signed'. The histogram counts are normalized to compensate for illumination. This can be done by accumulating a measure of local histogram energy over the somewhat larger connected regions and using the results to normalize all cells in the block. The combination of these histograms represents the final HOG descriptor. Invariance to scale and rotation may be also achieved by extracting descriptors from only salient points (keypoints) in the scale space of the image following a rotation normalization. The steps involved are:

(1) Scale-space extrema detection.
(2) Orientation assignment.
(3) Descriptor extraction.

The first step is intended to achieve scale invariance. The second step finds the dominant gradient orientation. All the orientation counts are then made relative to this dominant direction. Fig. 1 shows an example patch with their corresponding HOGs.

In (Albiol et al., 2008) (and the shorter version Monzo et al., 2008) the authors successfully applied HOG descriptors to the problem of face recognition. In that work, faces were previously normalized in scale and orientation, so the steps for scale-space extrema detection and orientation were not necessary. A set of 25 facial landmarks were localized using the Elastic Bunch Graph Matching framework (see Wiskott et al., 1997) with HOG features. The HOG features extracted from the vicinity of each of these 25 facial landmarks were used for classification, using nearest neighbor and Euclidean distance. It is important to note that for a new testing face, the matching procedure makes use of the eye's position (the eyes were assumed to be at a fixed position after normalization).

A potential drawback of the approach taken in (Albiol et al., 2008) is that the final error may crucially depend on the reliability of the landmark localizations. Our hypothesis in this paper is that such approach may not work well when landmarks are not precisely localized due to occlusions, strong illuminations or pose changes. Thus, in this work we propose to first normalize the face and then extract HOG features from a regular grid. The grid is formed by placing equal side patches around a first cell centered in the image, until the whole image is covered.

On the other hand, the size of the patch used to extract the HOG features is important. In (Albiol et al., 2008) the best size for the patch was estimated via cross-validation in the Yale database, prior to using the FERET database for the final experiments. The locality of the extracted features is determined by the patch size. We hypothesize that a better result could be obtained by combining information from different patch sizes. The fusion strategy that will be considered here is the product combination of the classifiers at patch sizes. Note that this combination rule is not optimal since it assumes independence of the classifiers, though empirically has

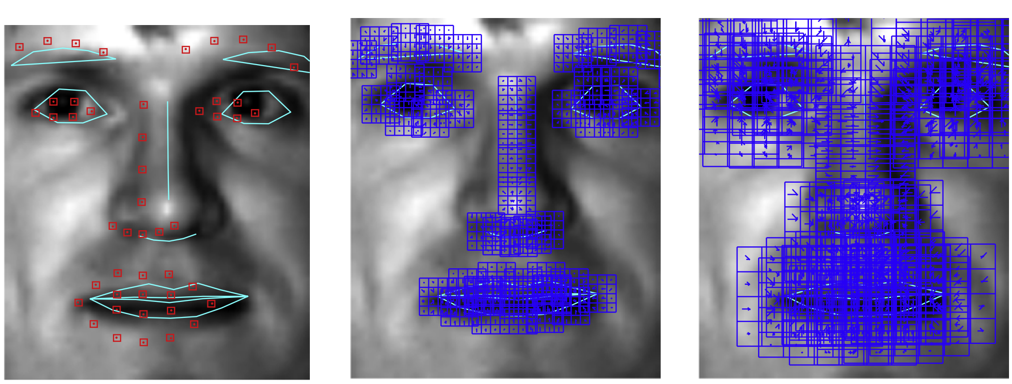

Fig. 2. Left: Initialized landmarks (red boxes) and result for the AAM fitting in an image from the Yale database. Center: extracted HOG descriptors, patch size = 24 × 24. Right: extracted HOG descriptors, patch size = 64 × 64. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

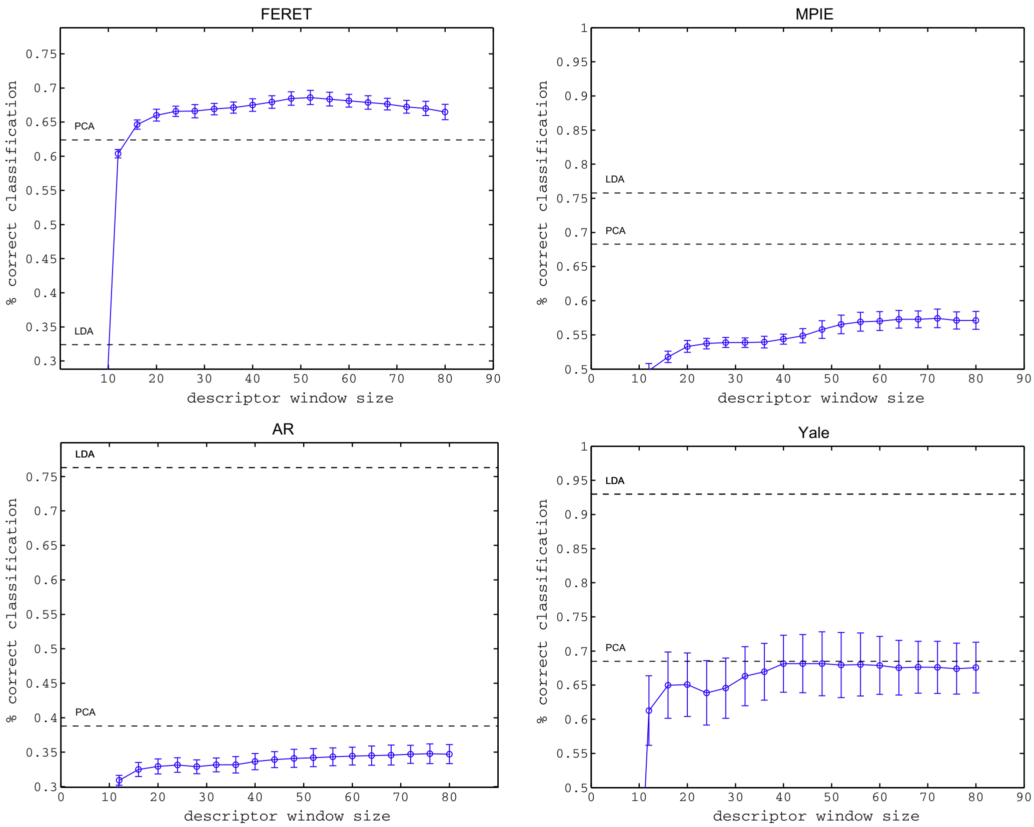

Fig. 3. Recognition rates (with standard deviations) using HOG features extracted from facial landmarks (10 runs with random training-test distribution of samples). The dashed lines show the recognition results using standard PCA and LDA (we use the number of bases that gave best results in the test set).

performed well in our experiments.[1] Suppose R individual classifiers c_k $(k = 1, \ldots, R)$ each one trained using HOG features are extracted from different patch sizes. Each classifier assigns one input sample (represented as x_k) to a label L_k $(L_k = w_1, \ldots, w_m)$. Assume the classifier c_k gives every output a measurement which is represented as a posterior probability vector, $P_k = [p(w_1|x_k), \ldots, p(w_m|x_k)]^t$, where $p(w_i|x_k)$ denotes the probability that the classifier considers that x was labeled with w_i. The product rule consists of fusing the final decision as:

$$j = \arg\max_i p(w_i) \prod_{k=1}^{R} p(w_i|x_k) \tag{1}$$

Note that when several overlapping patches are used, the final feature representation will be highly redundant and if the classifier does not have any mechanism for feature selection it might severely suffer from overfitting. Observe that the human face displays a structure common to all individuals. This implies that some gradient orientations would be very frequent in some specific zones of the face. Other orientations, on the contrary, would never or almost never appear in a given region. This reinforces the idea of the need for dimensionality reduction techniques.

3. Experiments

This section describes two experiments. The first experiment evaluates the impact of facial landmark localization in the face recognition performance. The second experiment evaluates the impact of extracting the HOG features from a regular grid and at multiple scales.

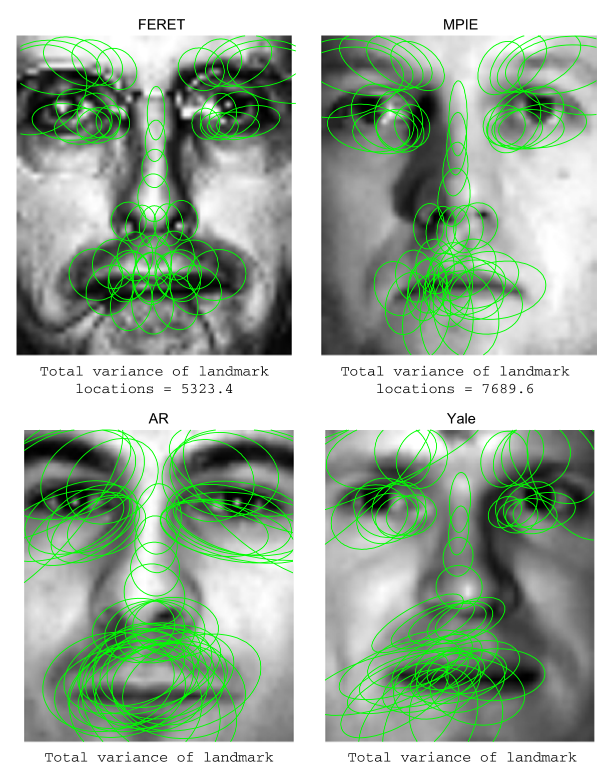

Fig. 4. Dispersion of the coordinates (ellipse fitting) and total sum of variances of the localized landmarks.

[1] Other evaluated fusion rules were majority voting, mean, median and maximum.

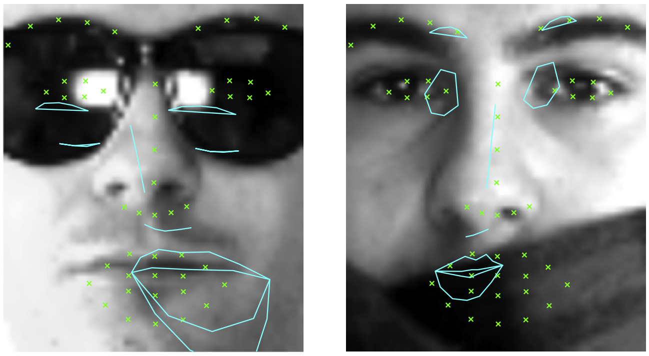

Fig. 5. Bad landmark localization examples in the AR database. The green crosses denote the initial search points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The experimental results were evaluated using four face databases: FERET (Phillips et al., 2000), AR (Martinez and Benavente, 1998), CMU Multi-PIE (Sim et al., 2001) and Yale (Yale face database, 2009). All these data sets cover a wide range of variations and scenarios for face recognition, see Table 1. All the images were previously normalized to 58×50 pixels.

3.1. Experiment 1: robustness to facial feature location

In the first experiment we evaluated the effect of the facial feature localization error on the final recognition performance. We compared the recognition performance using a holistic representation versus a HOG-based representation on facial landmarks. Forty-nine facial landmarks were localized using Active Appearance Models (AAMs) (De la Torre et al., 2007). Fig. 2 on the left shows the initialized set of landmarks and the landmarks found by the AAM in an image from the Yale database. Fig. 2 on the center and right illustrates the spatial domain for the HOG descriptors.

Fig. 3 shows the classification accuracy when using an holistic gray level representation with PCA or LDA (dotted lines), or when automatically detecting the landmarks with AMM and extracting HOG features (continuous line). When we consider absolute recognition performances, we see that the recognition rates when using the HOG representations extracted from the detected facial landmarks is lower when compared to the holistic PCA and LDA (except for FERET). The AAM was not trained under strong illumination and pose changes, and cannot deal with strong occlusions. This leads to large errors in the localization of facial features biasing the recognition performance. Observe that in the FERET database the classification performance of the landmark-based method is better because the landmark location is more accurate (no occlusion, and strong illumination changes). To check this hypothesis we computed a measure of landmark localization dispersion, see Fig. 4. The FERET database has the lower variance in the detected landmarks, whereas the AR has the highest variance (the AR database is the only one that includes major occlusions, like sunglasses and scarves, see Fig. 5). In this case, the precision of landmark detection correlates with the recognition performance using HOG

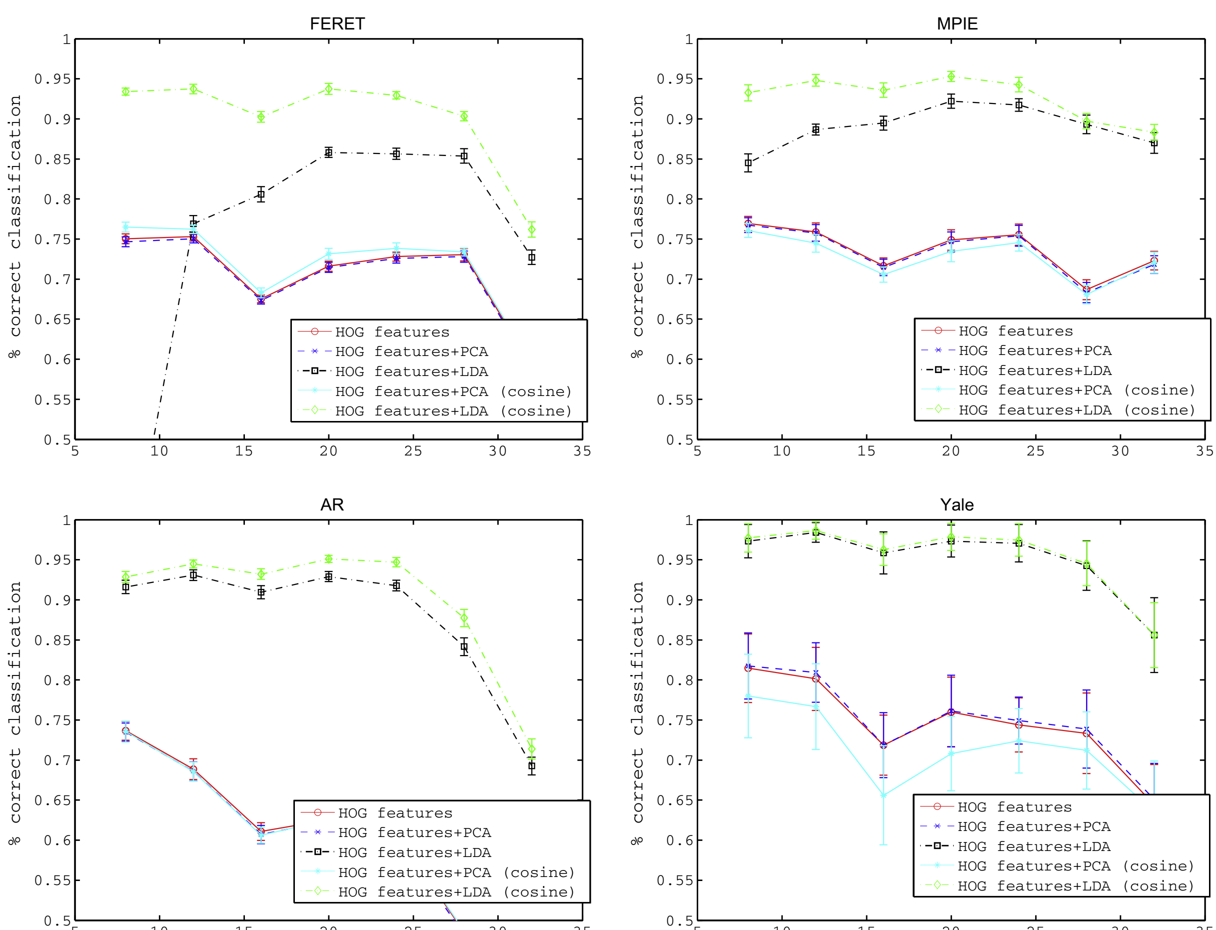

Fig. 6. Classification performance in four face databases. We show the mean and standard deviation over 10 runs, using 50% of the data for training and 50% testing.

Table 2
Best recognition rates in the FERET standard tests. HOG-EBGM refers to the previous HOG-based approach of (Albiol et al., 2008). The results of the last 7 rows were obtained using LDA for feature extraction (full feature space) and cosine distance. The best results for each set are in bold.

	fb (%)	fc (%)	dup1 (%)	dup2 (%)
PCA Euclidean	74.3	5.6	33.8	14.1
PCA Mahal. cosine	85.3	65.5	44.3	21.8
LDA	72.1	41.8	41.3	15.4
Bayesian	81.7	35.0	50.8	29.9
Bayesian map	81.7	34.5	51.5	31.2
Gabor ML	87.3	38.7	42.8	22.7
HOG-EBGM	**95.5**	81.9	60.1	55.6
8×8 patch	91.4	83.0	70.2	62.0
12×12 patch	93.0	82.0	70.8	63.3
16×16 patch	88.4	68.0	68.7	60.7
20×20 patch	93.7	75.3	70.2	60.3
24×24 patch	94.2	70.1	66.8	56.8
28×28 patch	91.6	42.8	60.0	56.0
Combination 8×8–28×28 (product rule)	95.4	**84.0**	**74.6**	**69.2**

Table 3
Time spent computing HOG features (in s). The table does not include the computational cost of locating the landmarks with AAMs.

Patch size	From landmarks	From regular grid
8×8	203.8	388.3
12×12	210.2	200.7
16×16	205.6	73.7
20×20	211.2	75.1
24×24	212.9	75.9
28×28	218.2	27.7
Total	1261.9	841.4

as compared with holistic PCA–LDA. The largest difference between PCA–LDA rates and HOG rates occur in the AR database that has largest variance in facial landmark localization.

3.2. Experiment 2: extracting regular grids and patch size combination

In the second experiment, we tested the effect of extracting HOG descriptors from a regular grid, at multiple scales, and the use of dimensionality reduction techniques. In this experiment, no landmark localization is used. HOGs are extracted from a regular grid of non-overlapped patches covering the whole normalized image. HOG features are then processed by PCA or LDA. A nearest neighbor classifier with Euclidean and cosine distances is used as classifier (no other classifier was used to be able compare results with (Albiol et al., 2008)). Fig. 6 illustrates the results on four databases.

Table 2 shows the results for the FERET standard tests and its comparison with the algorithms provided by the CSU Face Identification Evaluation System (Beveridge et al., 2005). In this test, the database images are organized into a gallery set (fa) and four probe sets (fb, fc,dup1,dup2). Using the FERET terminology, the gallery is the set of facial images with known identity 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 images in fc set were also taken in the same session but using a different camera and different lighting. Finally, the images in the sets dup1, dup2 were taken months apart (sometimes years), and there are changes in hair style as well as other iconic changes (e.g. glasses on), that makes these sets the most challenging ones. As can be seen on the table, recognition rates outperform previous work (Albiol et al., 2008).

Table 3 shows the computational cost of extracting HOG features for all 3540 FERET images both from landmarks and from a regular grid. As expected, the larger the patch size the less time is spent computing the HOG features. Note that for patch sizes larger than 12×12 the regular grid approach is less costly. As can be seen in the table, the average computational cost of the regular grid approach is lower. Moreover, if we consider the time spent in locating landmarks, about 2.5 s per image (AAM implemented in Matlab using non-optimized code), the proposed approach is computationally more efficient. All these measures were taken using interpreted code.

4. Conclusions

This paper explores the use of HOG features for face recognition. The contributions are threefold: (1) to provide robustness to facial feature detection, we propose to uniformly sampling the HOG features, (2) to remove redundancy in the data, improve computational efficiency and avoid overfitting, we propose to use dimensionality reduction in the HOG representation, (3) we show that a decision-level combination of results using HOG features extracted from different image patch sizes significantly improves on choosing a single best patch size. Taking into account these considerations we were able to obtain a significant increase (up to 13%) in recognition performance on the standard FERET database.

Acknowledgments

This work was partially funded by project PII2I09-0043-3364 of Castilla-La Mancha Regional Government and the Spanish Ministry of Science and Innovation Funds (project TIN2008-06068).

References

Ahonen, T., Hadid, A., Pietikainen, M., 2004. Face recognition with local binary patterns. In: Proc. Eighth European Conf. Computer Vision, pp. 469–481. <http://www.springerlink.com/content/p5d9xp9gfkex5gk9>.
Albiol, A., Monzo, D., Martin, A., Sastre, J., Albiol, A., 2008. Face recognition using HOG-EBGM. Pattern Recognition Lett. 29 (10), 1537–1543.
Amin, M.A., Yan, H., 2009. An empirical study on the characteristics of gabor representations for face recognition. IJPRAI 23 (3), 401–431.
Baranda, J., Jeanne, V., Braspenning, R., 2008. Efficiency improvement of human body detection with histograms of oriented gradients. In: Proc. ICDSC08, pp. 1–9.
Bartlett, M.S., Movellan, J.R., Sejnowski, T.J., 2002. Face recognition by independent component analysis. IEEE Trans. Neural Networks 13 (6), 1450–1464 <http://dx.doi.org/10.1109/TNN.2002.804287>.
Bertozzi, M., Broggi, A., Rose, M.D., Felisa, M., Rakotomamonjy, A., Suard, F., 2007. A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. In: Proc. Intelligent Transportation Systems Conf., pp. 143–148.
Beveridge, J., Bolme, D., Draper, B., Teixeira, M., 2005. The CSU face identification evaluation system: Its purpose, features, and structure. MVA 16 (2), 128–138.
Chellappa, R., Wilson, C., Sirohey, S., 1995. Human and machine recognition of faces: A survey. Proc. IEEE 83 (5), 705–740.
Chellappa, R., Zhao, W. (Eds.), 2005. Face Processing: Advanced Modeling and Methods. Elsevier.
Chuang, C., Huang, S., Fu, L., Hsiao, P., 2008. Monocular multi-human detection using augmented histograms of oriented gradients. In: Proc. ICPR08, pp. 1–4.
Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proc. CVPR 2005, vol. 1, pp. 886–893. <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1%467360>.
De la Torre, F., Collet, A., Cohn, J.F., Kanade, T., 2007. Filtered component analysis to increase robustness to local minima in appearance models. In: IEEE Conf. on Computer Vision and Pattern Recognition.
He, N., Cao, J., Song, L., 2008. Scale space histogram of oriented gradients for human detection. In: Internat. Symposium on Information Science and Engineering, 2008 (ISISE'08), pp. 167–170.
Kittler, J., Ghaderi, R., Windeatt, T., Matas, J., 2001. Face verification using error correcting output codes. In: Computer Vision and Pattern Recognition (CVPR01). IEEE Press, pp. 755–760.

Kobayashi, T., Hidaka, A., Kurita, T., 2008. Selection of histograms of oriented gradients features for pedestrian detection. In: Proc. ICONIP 2007, Revised Selected Papers, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 598–607.

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

Martinez, A., Benavente, R., 1998. The AR face database. Tech. Rep. 24, CVC.

Monzo, D., Albiol, A., Sastre, J., Albiol, A., 2008. HOG-EBGM vs. Gabor-EBGM. In: Proc. Internat. Conf. on Image Processing, San Diego, USA.

Perdersoli, M., Gonzalez, J., Chakraborty, B., Villanueva, J., 2007a. Boosting histograms of oriented gradients for human detection. In: Proc. 2nd Computer Vision: Advances in Research and Development (CVCRD), pp. 1–6.

Perdersoli, M., Gonzalez, J., Chakraborty, B., Villanueva, J., 2007b. Enhancing real-time human detection based on histograms of oriented gradients. In: 5th Internat. Conf. on Computer Recognition Systems (CORES'2007).

Phillips, P., Moon, H., Rizvi, S., Rauss, P., 2000. The FERET evaluation methodology for face-recognition algorithms. TPAMI 22 (10), 1090–1104.

Samal, A., Iyengar, P.A., 1992. Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recognition 25 (1).

Sim, T., Baker, S., Bsat, M., 2001. The CMU pose, illumination, and expression (PIE) database of human faces. Tech. Rep. CMU-RI-TR-01-02, Robotics Institute, Pittsburgh, PA.

Suard, F., Rakotomamonjy, A., Bensrhair, A., Broggi, A., 2006. Pedestrian detection using infrared images and histograms of oriented gradients. In: Intelligent Vehicles Symposium, Tokyo, Japan, pp. 206–212.

Turk, M.A., Pentland, A., 1991. Eigenfaces for recognition. Cognitive Neurosci. 3 (1), 71–86 <ftp://whitechapel.media.mit.edu/pub/images/>.

Wang, C., Lien, J., 2007. Adaboost learning for human detection based on histograms of oriented gradients. In: Proc. ACCV07, pp. 885–895.

Watanabe, T., Ito, S., Yokoi, K., 2009. Co-occurrence histograms of oriented gradients for pedestrian detection. In: Proc. PSIVT09, pp. 37–47.

Wiskott, L., Fellous, J., Krüger, N., von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. In: Sommer, G., Daniilidis, K., Pauli, J. (Eds.), Proc. 7th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'97, Kiel. No. 1296. Springer-Verlag, Heidelberg, pp. 456–463 <citeseer.ist.psu.edu/wiskott99face.html>.

Yale face database, 2009. Yale face database. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html> (last accessed April 2009).

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

Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S., 2006. Fast human detection using a cascade of histograms of oriented gradients. In: Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2006, pp. 1491–1498.