FACE RECOGNITION WITH HISTOGRAMS OF ORIENTED GRADIENTS

Oscar Déniz, Gloria Bueno
E.T.S. Ingenieros Industriales, Avda. Camilo José Cela s/n, 13071 Ciudad Real, Spain

Jesus Salido
E. Superior de Informática, Paseo de la Universidad 4, 13071 Ciudad Real, Spain

Fernando de la Torre
Robotics Institute, Carnegie Mellon University
211 Smith Hall, 5000 Forbes Ave. Pittsburgh PA 15213, U.S.A.

Keywords: Face recognition, Histogram of oriented gradients.

Abstract: Histograms of Oriented Gradients have been recently used as discriminating features for face recognition. In this work we improve on that work in a number of aspects. As a first contribution, it identifies the necessity of performing feature selection or transformation, especially if HOG features are extracted from overlapping cells. Second, the use of four different face databases allowed us to conclude that, if HOG features are extracted from facial landmarks, the error of landmark localization plays a crucial role in the absolute recognition rates achievable. This implies that the recognition rates can be lower for easier databases if landmark localization is not well adapted to them. This prompted us to extract the features from a regular grid covering the whole image. Overall, these considerations allow to obtain a significant recognition rate increase (up to 10% in some subsets) on the standard FERET database with respect to previous work.

1 INTRODUCTION

Face recognition is becoming one of the most actively researched problems in Computer Vision. The available literature is increasing at a significant rate, and even the number of conferences and special issues entirely devoted to face recognition is growing. Access to inexpensive cameras and computational resources has allowed researchers to explore the problem from many different perspectives, see the surveys (Zhao et al., 2003; Chellappa et al., 1995; Samal and Iyengar, 1992; Chellappa and Zhao, 2005).

One central aspect in the face recognition problem is the kind of features to use. From the early distinction between geometric and photometric (view based) features, the latter seem to have prevailed in the literature. In any case the proposed features seem endless: Eigenfaces, Gabor wavelets, LBP, error-correcting output coding, PCA, ICA, infrared, 3D, etc. The fact is that researchers continue to propose new features that prove more and more powerful. One of the recent contenders is Histograms of Oriented Gradients (HOG) (Dalal and Triggs, 2005). Originally used for object detection, it has been recently applied to face recognition with promising results. In this paper, the use of HOGs for face recognition is further studied. We improve on previous work by adopting a different approach in the extraction of the features and by identifying the necessity of some kind of posterior feature transformation. Our analysis allows to gain some insights about the feature extraction method, whereby significant improvements can be obtained in standard face databases.

This paper is organized as follows. Section 2 describes HOG in detail, as well as our approach. In Section 3 we describe the experiments carried out. Finally in Section 4 the main conclusions of the work are outlined.

Déniz O., Bueno G., Salido J. and de la Torre F. (2010).
FACE RECOGNITION WITH HISTOGRAMS OF ORIENTED GRADIENTS.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 339-344
DOI: 10.5220/0002820503390344

2 HISTOGRAMS OF ORIENTED GRADIENTS FOR FACE RECOGNITION

The algorithm for extracting HOGs (see (Dalal and Triggs, 2005)) begins by counting occurrences of gradient orientation in localized portions of an image. Basically, the image is divided into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The histogram counts are normalized so as to compensate for illumination. The combination of these histograms then represents the descriptor. Invariance to scale and rotation is also achieved by extracting descriptors only from salient points (keypoints) in the scale space of the image. 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. Figure 1 shows an example patch with their corresponding HOGs.

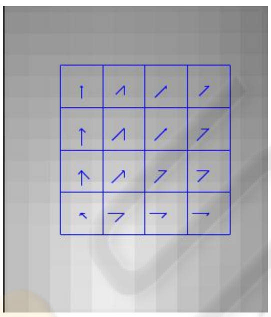

Figure 1: Example HOG descriptors, patch size=8x8. Each cell of the patch shows the gradient orientations present.

Since its introduction, HOG features have been used almost exclusively for person 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; Perdersoli et al., 2007b). To the best of our knowledge, the only work that studies the application of HOGs to face recognition is the recent (Albiol et al., 2008) (and the shorter version (Monzo et al., 2008)). In that work, the faces used were previously normalized so the steps of scale-space extrema detection and orientation assignment 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 the 25 facial landmarks localized were used for classification, using nearest neighbor and Euclidean distance. It is important to note that for each new face, the matching stage of landmark localization had the advantage of starting from the known positions of the eyes.

The problem of the approach taken in (Albiol et al., 2008) is that the final error may crucially depend on the reliability of the landmark localization stage. Our hypothesis is that such approach may not work well when landmarks are not precisely localized either because occlusions, strong illumination gradients or other reasons. For many facial zones there would be no point in trying to localize landmarks when the face image has been already normalized. Besides, in (Albiol et al., 2008) the authors do not mention that the patch sizes and number of facial landmarks used imply a high degree of overlap between patches, and do not take this fact into account, as no feature selection or extraction is carried out after extracting the HOG features. It is important to note that HOG features are sparse for structured objects. The human face displays some structure that is common to all individuals. This means 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. For these reasons it seems reasonable to think that some sort of feature selection or transformation must be applied to the HOG features.

In this work we propose to extract HOG features from a regular grid covering the whole normalized image of the face, followed by feature extraction. The grid is formed by placing equal side patches around a first cell centered in the image, until the whole image is covered. The next Section shows the experimental results of the proposed modifications.

3 EXPERIMENTS

In order to provide robust results we studied HOGs with four different face databases: FERET (Phillips et al., 2000), AR (Martinez and R.Benavente, 1998), CMU Multi-PIE (Sim et al., 2001) and Yale (Yale face database, 2009). These data sets together cover a wide range of variations and scenarios, see Table 1. All the images were previously normalized to 58x50 pixels.

In the first experiment we tried to test how well the approach of (Albiol et al., 2008) worked. 49 landmark positions were automatically extracted from

Table 1: Facial 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

each face image. The landmark localization method is based on Active Appearance Models (AAM) and is described in detail in (Nguyen and De la Torre, 2008). There was a single set of initialization points for each database, obtained by manually adjusting a standard template both in scale and translation. Figure 2 shows the initialization points (in red) and the localized landmarks for a sample Yale image.

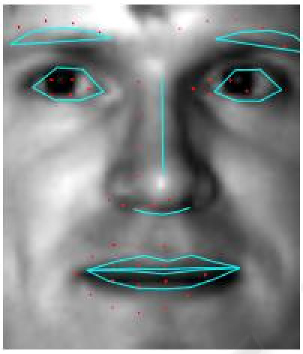

Figure 2: Initialization points (in red) and localized landmarks for a sample Yale image.

Figure 3 shows examples of the HOG features extracted.

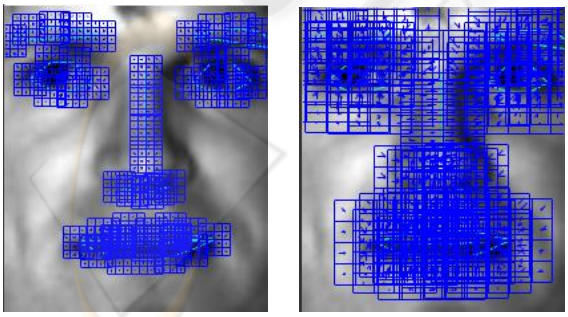

Figure 3: Left: extracted HOG descriptors, patch size=24x24. Right: extracted HOG descriptors, patch size=64x64.

Figure 4 shows the recognition rates using HOG features, along with baseline performances obtained with PCA, LDA and nearest neighbor classifier (Euclidean distance). When we consider absolute performances we see that the recognition rates are too low when compared to PCA and LDA (except for FERET). Our explanation for this is the following: face landmark localization plays a role in absolute performance. For the FERET database landmark localization turns out to be relatively good, but not for other databases. We checked this by considering the landmark localization dispersion. In terms of landmark localization FERET appears to be the best database, the AR[1] being the worst (the total variances of landmark localizations were 5323 for FERET, 7680 for MPIE, 10283 for Yale and 22053 for AR). This fits with the recognition rates using HOG as compared with PCA-LDA: the largest difference between PCA-LDA rates and HOG rates is that of the AR database, while the best relative performance between the two is that of FERET. Hence we conclude that when HOG features are extracted from landmarks, landmark localization plays a role in absolute performances achievable.

The authors of (Albiol et al., 2008) only used the FERET and Yale databases in their experiments. For the Yale database the HOG features allowed them to get a recognition rate as high as with LDA (around 97%). They could not infer the fact that it was not only the HOG features but especially their landmark localization technique what made that good result possible. Our absolute recognition results are comparatively poorer for the Yale database. That can be due to a worse landmark localization, to a worse normalization of the images in the MPIE, AR and Yale databases, or to the fact that we are not using the correct eye positions as initialization for the landmark search as done in (Albiol et al., 2008).

The second experiment considered the proposed regular grid HOG followed by feature extraction. In this case, no landmark localization is performed. HOGs are extracted from a regular grid of non-overlapped patches covering the whole normalized image. HOG features are then processed by PCA

[1] the AR database is the only one that includes major occlusions, like sunglasses and scarves.

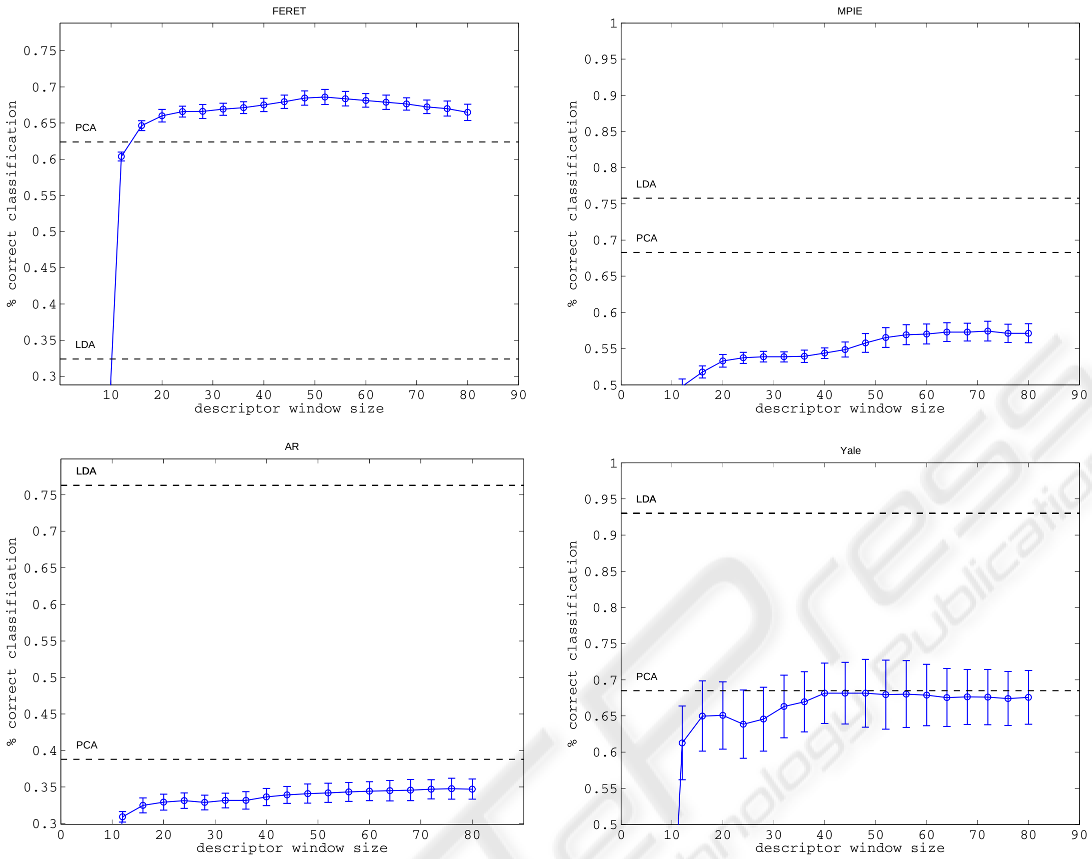

Figure 4: Recognition rates using HOG features extracted from facial landmarks.

or LDA. Nearest neighbor (Euclidean and cosine distances) is used for classifying (no other classifier was used since we wanted to compare results with (Albiol et al., 2008)). Figure 5 shows the results.

Table 2 shows the results within the context of the FERET standard test and compared with the algorithms provided by the CSU Face Identification Evaluation System (Beveridge et al., 2005). In this test, 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 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. As can be seen on the table, recognition rates are significantly higher than in the previous reference work.

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 6 rows were obtained using LDA for feature extraction (full feature space) and cosine distance.

	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%
8x8 patch	91.4%	83.0%	70.2%	62.0%
12x12 patch	93.0%	82.0%	70.8%	63.3%
16x16 patch	88.4%	68.0%	68.7%	60.7%
20x20 patch	93.7%	75.3%	70.2%	60.3%
24x24 patch	94.2%	70.1%	66.8%	56.8%
28x28 patch	91.6%	42.8%	60.0%	56.0%

4 CONCLUSIONS

This work shows the results of a study of HOG features in face recognition, improving on recent pub-

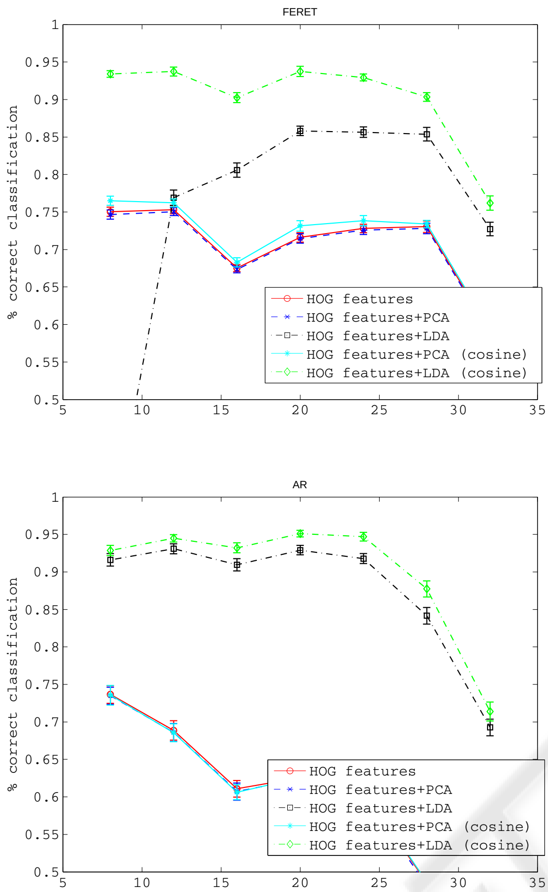

Figure 5: Results using the proposed approach. Averages of 10 runs, 50%-50% partition of the samples between train and test.

lished work in a number of aspects. As a first contribution, it identifies the necessity of performing feature selection, especially if HOG features are extracted from overlapping cells. Second, the use of four different face databases allowed us to conclude that, if HOG features are extracted from facial landmarks, the error in landmark localization plays a crucial role in the absolute recognition rates attainable. This implies that the recognition rates can be lower for easier databases if landmark localization is not well adapted to them. This prompted us to extract the features from a regular grid covering the whole image. Overall, these considerations allow to obtain a significant increase (up to 10% in some subsets) in recognition ratios on the standard FERET database.

ACKNOWLEDGEMENTS

This work was supported by funds from project PII2I09-0043-3364 from Gobierno de Castilla-La Mancha and from José Castillejo scholarships JC2008-00099 and JC2008-00057 from the Spanish Ministry of Science.

REFERENCES

Albiol, A., Monzo, D., Martin, A., Sastre, J., and Albiol, A. (2008). Face recognition using HOG-EBGM. *Pattern Recognition Letters*, 29(10):1537–1543.

Baranda, J., Jeanne, V., and Braspenning, R. (2008). Efficiency improvement of human body detection with histograms of oriented gradients. In *ICDSC08*, pages 1–9.

Bertozzi, M., Broggi, A., Rose, M. D., Felisa, M., Rakotomamonjy, A., and Suard, F. (2007). A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. In *Proc. Intelligent Transportation Systems Conference*, pages 143–148.

Beveridge, J., Bolme, D., Draper, B., and Teixeira, M. (2005). The CSU face identification evaluation system: Its purpose, features, and structure. *MVA*, 16(2):128–138.

Chellappa, R., Wilson, C., and Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings IEEE*, 83(5):705–740.

Chellappa, R. and Zhao, W., editors (2005). *Face Processing: Advanced Modeling and Methods*. Elsevier.

Chuang, C., Huang, S., Fu, L., and Hsiao, P. (2008). Monocular multi-human detection using augmented histograms of oriented gradients. In *ICPR08*, pages 1–4.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. volume 1, pages 886–893.

He, N., Cao, J., and Song, L. (2008). Scale space histogram of oriented gradients for human detection. In *International Symposium on Information Science and Engieering, 2008. ISISE '08*, pages 167–170.

Kobayashi, T., Hidaka, A., and Kurita, T. (2008). Selection of histograms of oriented gradients features for pedestrian detection. pages 598–607.

Martinez, A. and R.Benavente (1998). The AR face database. Technical Report 24, CVC.

Monzo, D., Albiol, A., Sastre, J., and Albiol, A. (2008). HOG-EBGM vs. Gabor-EBGM. In *Proc. Internacional Conference on Image Processing*, San Diego, USA.

Nguyen, M. and De la Torre, F. (2008). Local minima free parameterized appearance models. In *IEEE Conference on Computer Vision and Pattern Recognition*.

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

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

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

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

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

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

Wang, C. and Lien, J. (2007). Adaboost learning for human detection based on histograms of oriented gradients. In *ACCV07*, pages I: 885–895.

Watanabe, T., Ito, S., and Yokoi, K. (2009). Co-occurrence histograms of oriented gradients for pedestrian detection. In *PSIVT09*, pages 37–47.

Wiskott, L., Fellous, J., Krüger, N., and von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. In Sommer, G., Daniilidis, K., and Pauli, J., editors, *Proc. 7th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'97, Kiel*, number 1296, pages 456–463, Heidelberg. Springer-Verlag.

Yale face database (2009). Yale face database. Last accessed: April 2009.

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

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