Local Boosted Features for Pedestrian Detection*

Michael Villamizar[1,2], Alberto Sanfeliu[1,2], and Juan Andrade-Cetto[1]

[1] Institut de Robòtica i Informàtica Industrial, CSIC-UPC
[2] Department of Automatic Control, UPC

Abstract. The present paper addresses pedestrian detection using local boosted features that are learned from a small set of training images. Our contribution is to use two boosting steps. The first one learns discriminant local features corresponding to pedestrian parts and the second one selects and combines these boosted features into a robust class classifier. In contrast of other works, our features are based on local differences over Histograms of Oriented Gradients (HoGs). Experiments carried out to a public dataset of pedestrian images show good performance with high classification rates.

1 Introduction

Recently, several techniques based on Histograms of Oriented Gradients (HoG) have been developed showing successful results in object detection and categorization [1,2,3,4,5,6]. These type of features have demonstrated robustness and reliability for representing local image features. The keypoint in using HOG descriptors is to capture or encode feature layout where each histogram cell contains an oriented gradient distribution for pixels within this cell.

Dalal and Triggs [1] proposed to use HOG descriptors for pedestrian detection in static images and in videos. They use an overlapping local contrast normalization in order to improve detection performance giving a certain invariance to illumination and shadows. In Bosch *et al.* [2] pyramidal Histograms of Oriented Gradients (PHoGs) are used for object categorization. These pyramidal descriptors encode features and their spatial layout in several resolution levels, allowing robustness to small feature shifts. Finer histogram levels are weighted more than coarser ones, since finer resolutions have more a detailed feature shape description. This idea is inspired by image pyramid representation of scenes [3]. This spatial pyramidal representation is an extension to the Dalal and Triggs method where Histograms of Oriented Gradients are restricted to finer resolutions. In the same way, SIFT features [4] compute fixed HoG descriptors in a grid of 4*x*4 cells and 8 gradient orientations around interest points.

A cascade of HoG features has been proposed using the Adaboost algorithm to construct a fast and accurate human detector [5]. This idea is similar to addressed by Viola and Jones [7], but using HoGs instead of Haar-like features. The method selects features

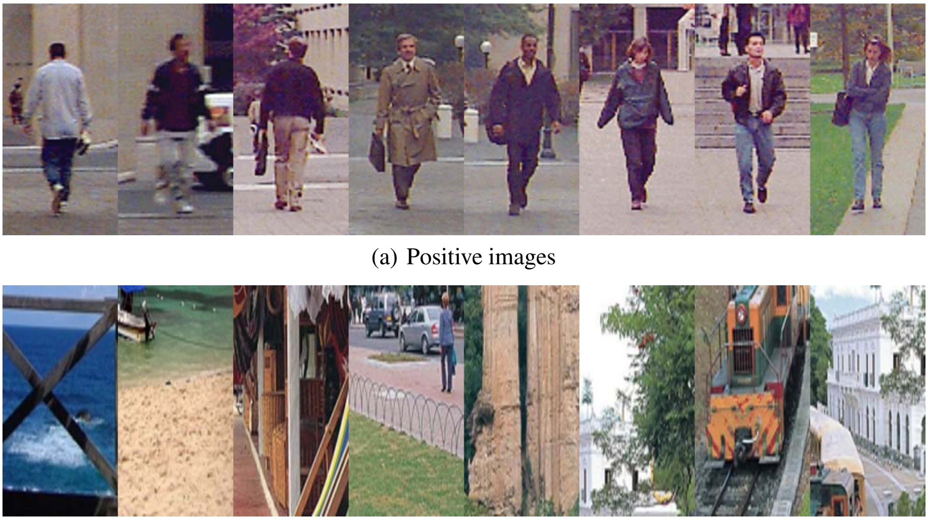
* This research was partially supported by the Spanish Ministry of Innovation and Science under projects Consolider Ingenio 2010 CSD2007-00018, and projects DPI 2007-614452 and DPI 2008-06022; by the URUS project IST-045062 of the European Union; and by the Technical University of Catalonia (UPC) to MV.

of several sizes, finding in early cascade stages larger features. This fact is suitable for rapid detection and for discarding background image locations but it might have problems when pedestrians are partially occluded. A similar method is proposed by Laptev [6] for object class detection. This method aims to find what HoG features to use and where to compute them. For this, training information is used to determine the local HoG features, and HoG feature selection is carried out using random regions and selecting the best boosting over Weighted Fischer linear discriminant weak learners.

Our contribution is to learn reliable features inside HoG instead of using the whole HoG descriptor to detect human parts. In contrast to previous methods that use whole local HoG descriptors, we propose to use the training information to seek out the most discriminant HoG-based features using a boosting algorithm. These boosted features focus on histomgram's bins with high occurence and discard those whose contribution to object detection is lower. This learning process is carried out in all image locations in order to determine which are the most relevant pedestrian parts. Once the boosted features are learned a final boosting step is performed to select and combine the most discriminant.

2 Approach

We present boosted features computed on Histograms of Oriented Gradients in order to have faster and robust features with which to describe human parts and to face up intraclass variations present in class images (Figure 1). These local features are learned in a first training step and combined in a second one. This last step rejects some initial boosted features because they are not discriminant enough and tend to favor the background. Unlike Laptev's work, our boosted features are not computed in random



(a) Positive images

(b) Negative images

Fig. 1. Positive and negative images

locations but computed exhaustively over the whole image with the aim to determine which image locations are human parts and which ones are background. The training is carried out using the well-known Adaboost algorithm that has successful results in object detection [6,7,8].

Given that pedestrian HoGs are corrupted by background, we propose HoG-based features instead of whole HoG descriptors in order to concentrate on HoG parts with high reliability. Although these type of features have been used for keypoint classification and segmentation in intensity images [9,10], we use them over histograms of oriented gradients, combining the simplicity of these features with the robustness of HoGs.

The paper is organized as follows, Section 3 explains the local boosted feature computation using training data and simple features over histograms of oriented gradients. In Section 4 the boosted feature selection is described. The implementation details and experiments performed on a public pedestrian dataset are shown in Sections 5 and 6, respectively.

3 Boosted Features

Histograms of oriented gradients are used to capture local feature layout over images. Each histogram cell has a local distribution of features (gradient orientations). The proposed work aims to seek out the image locations, corresponding to human parts, that have high similarities across positive images and are discriminant to negative ones. These image locations are selected via a local boosting step where HoG-based features are computed.

Given that the input images have pedestrians segmented and aligned, it is possible to learn a classifier for each image location that allows to recognize it. This classifier is called boosted feature because of it consists of a weighted combination of simple HoG-based features. Boosted features measure similarity across positive and negative images. Positive histograms have strong similarities across images. Such similarities are learned using boosting of simple features over histograms. Figure 2 shows a local HoG

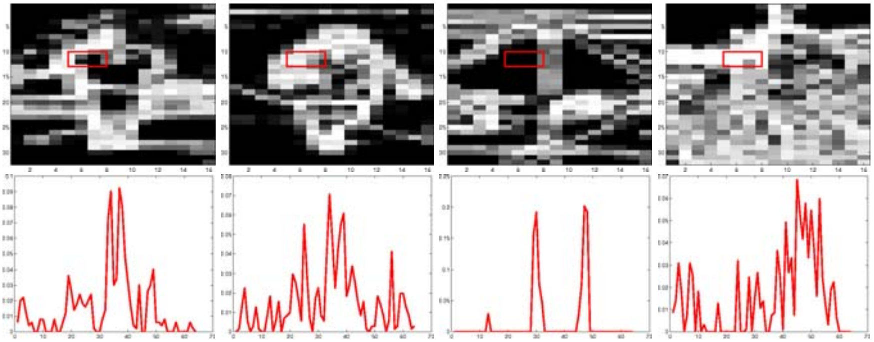
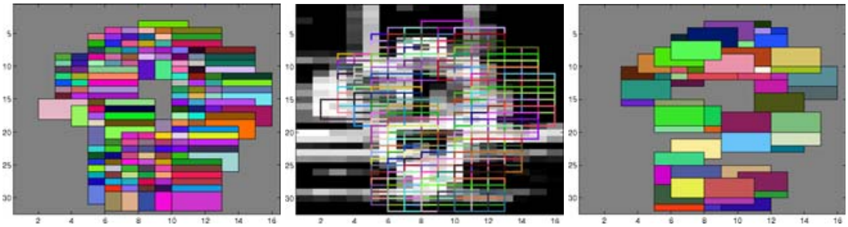


Fig. 2. Local HoG similarities. Positive and negative gradient images (first row), and their local HoGs (second row).



(a) Initial boosted features (b) Initial boosted features (c) Final boosted features

Fig. 3. Boosted features. Color boxes correpond to locations where boosted features have a high similarity across positive images.

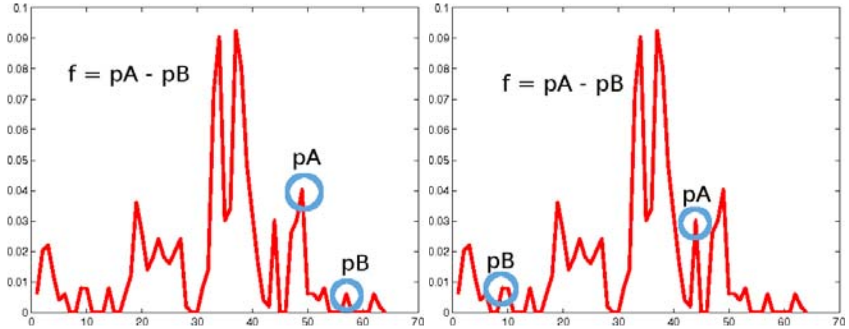


Fig. 4. HoG-based feature instances

and its similarities across images. Regions where it is possible to get a boosted feature classifier are selected as initial human parts. As background locations do not have strong similarities, the method can not find a classifier for those locations and therefore such locations are discarded. Figures 3a and 3b show initial boosted features. As we can see, the boosted features are localized on human contours and some background regions that present certain similarities such as ground edges.

The boosted features consist of combinations of simple and rapid features. These simple HoG-based features are defined as the difference between values of histogram bins

$$f_i(HoG) = HoG(pA) - HoG(pB), \tag{1}$$

where pA and pB are bin indexes. The boosted combination of these features gives a boosted feature classifier that represents those features which better classify the training data for current image location :

$$bof(HoG) = \sum_{i=1}^{n} \alpha_i f_i(HoG) \tag{2}$$

Parameter α_i is the weight associated to each feature and is obtained with the Adaboost algorithm, according to its classification error. Some illustrative feature instances are shown in Figure 4. These simple features resemble to features used for keypoint classification and for image categorization and segmentation. However, in this work they are computed over HoG descriptors. To increase robustness to image and intraclass variations, multiple boosted features are computed for the same image location and their outputs combined. This idea is inspired from randomized decision forests where each tree is trained on small random subsets from training images [10]. This technique speeds up the training time and reduces over-fitting problems. Therefore our local boosted feature is a combination of simple boosting iterations :

$$bof(HoG) = \sum_{j=1}^{m} \sum_{i=1}^{n} \alpha_{i,j} f_{i,j}(HoG) \tag{3}$$

Although each boosted feature has to be computed for each image location, by a window convolution over the image, the process is rapid lasting few seconds per region (in our case 2 seconds using 50 positive and 500 negative training images).

4 Boosted Feature Selection

In this step a second boosting process is applied to select the most accurate and discriminant boosted features over a validation set of images. The aim is to choose those boosted features that correspond to human parts in order to form the whole pedestrian detector while rejecting boosted features with low discriminant power against negative or background images. Furthermore, this step reduces the amount of boosted features that is suitable for a rapid detection. In spite of the reduced number of features and their simplicity, the proposed method achieves high detection rates in our experiments. This is because the method is focused on selecting what features to use and where to compute them. Figure 3c shows the final boosted features after the boosting selection step. The boosted features are localized on pedestrian contours.

5 Implementation Details

5.1 Histogram of Oriented Gradients

Gradient computation is applied to input images to form HoG images. The input image size is $256x128$ pixels both for positive and negative images. 4 gradient orientations are chosen to this work. A spatial binning of $8x8$ pixels is done to obtain a HoG image of $32x16x4$ cells. The boosted features are computed in small regions of $4x4x4$ cells using a sliding window over whole HoG image.

5.2 Boosted Features

The initial HoG-based features are selected randomly. In our experiments 150 features per location have been chosen. Boosting selects the 8 (n) more discriminant HoG-based features and uses 5 (m) boosting iterations to have finally a boosted Feature formed by 40 difference features. These parameters are chosen by the user according to object class. The number of images per set is 50% of all training images.

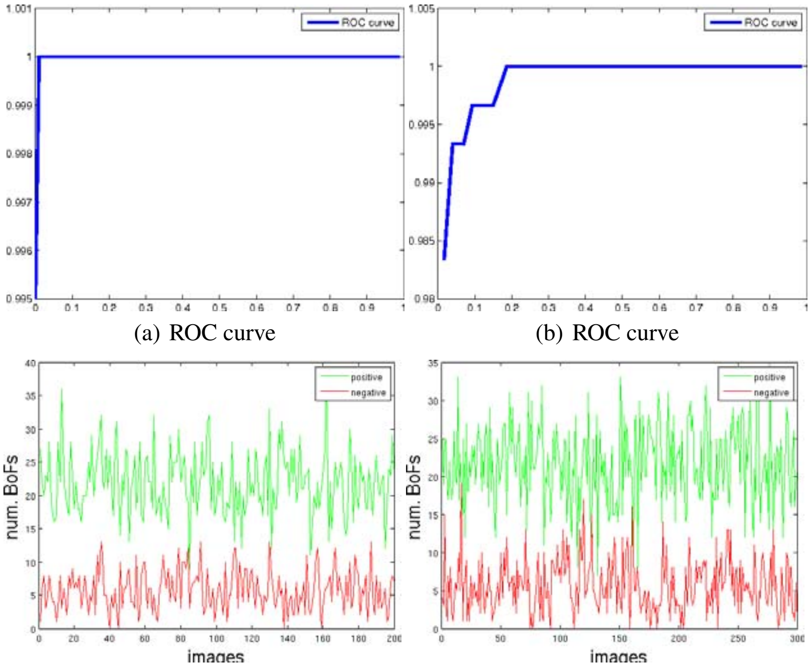
5.3 Detection Images

For each detection image an image pyramid is built, and for each level a HoG image is computed using the previous implementation details. The number of pyramid levels is 6 for an input image size of 640*x*854 pixels.

6 Experiments

The proposed method has been tested in a public pedestrian dataset [13]. This dataset contains people segmented but with a large amount of background. Moreover, people in images have a high class varitions because of different clothes and illumination conditions, see Figure 1. Negative images are extracted from several images which do not have people and have a high gradient content. Since the proposed method requires training, validation and test sets of images, the pedestrian dataset is split up randomly. For training, 50 positive and 500 negative images are chosen. In the same way, 200 and 200 images for validation and 300 and 300 for test, respectively.

In spite of the small set of positive training images (50) the proposed method performs well in image classification both in validation and test sets. These results can be seen in the ROC curve (Figure 5a).These results show how our approach using the learned boosted features generalizes the pedestrian class classification. Figures 5b and



(a) ROC curve

(b) ROC curve

(c) validation

(d) test

Fig. 5. Detection curves

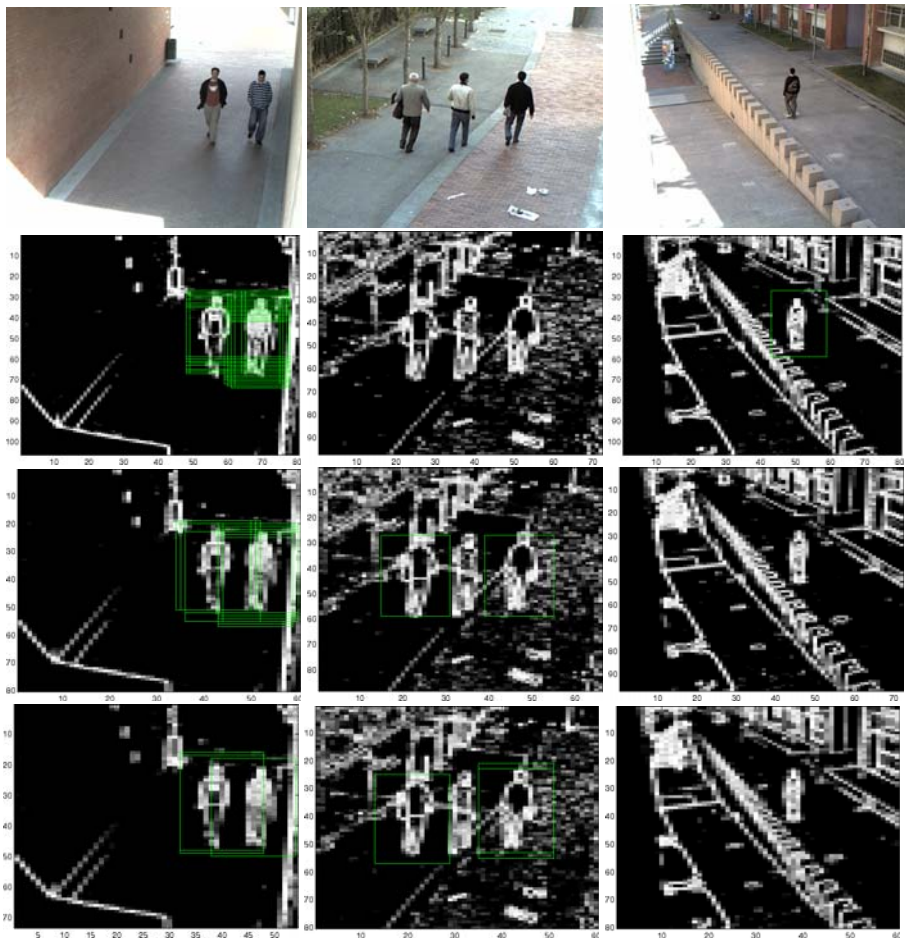


Fig. 6. Pedestrian detection

5c show boosted feature detection in validation and test images. We can see that positive images have more detected features than negative ones. However, negative images contain a considerable amount of features because they have local image representations like edges, corners, etc. One second test was to perform the pedestrian detector over outdoor scenes (Figure 6). The top images are input images and in each column appears detections in some HoG pyramid levels for each image. The first column shows how pedestrians are detected across several levels unlike image 3. The second column shows one false negative.

7 Conclusions

This paper proposes a detection method using learned boosted features that can be used for object categorization under high intraclass variation. The approach finds out the most significant local parts while rejecting background regions. These HoG-based features allow a robust and rapid image detection with high classification rates. The approach generalizes the pedestrian class using a reduced number of training images.

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, California, vol. 1, pp. 886–893 (2005)
2. Bosch, A., Zisserman, A., Muoz, X.: Image Classification Using ROIs and Multiple Kernel Learning. International Journal of Computer Vision (2008)
3. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, vol. 2, pp. 2169–2178 (2006)
4. Lowe, D.: Distinctive image features from scale invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
5. Zhu, Q., Avidan, S., Ye, M., Cheng, K.-T.: Fast human detection using a cascade of Histograms of Oriented Gradients. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, vol. 2, pp. 1491–1498 (2006)
6. Laptev, I.: Improvements of object detection using boosted histograms. In: British Machine Vision Conference, vol. 3, pp. 949–958 (2006)
7. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, vol. 1, pp. 511–518 (2001)
8. Villamizar, M., Sanfeliu, A., Andrade-Cetto, J.: Computation of rotation local invariant features using the integral image for real time object detection. In: IAPR International Conference on Pattern Recognition, Hong Kong, pp. 81–85 (2006)
9. Ozuysal, M., Fua, P., Lepetit, V.: Fast Keypoint Recognition in Ten Lines of Code. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, pp. 1–8 (2007)
10. Shotton, J., Johnson, M., Cipolla, R.: Semantic Texton Forests for Image Categorization and Segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, pp. 1–8 (2008)
11. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: International Conference on Computer Vision, Beijing, vol. 2, pp. 1458–1465 (2005)
12. Chum, O., Zisserman, A.: An exemplar model for learning object classes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, pp. 1–8 (2007)
13. Papageorgiou, C., Poggio, T.: A trainable system for object detection. International Journal of Computer Vision 38(1), 15–33 (2000)