Detecting small pedestrians

David Vázquez*, David Gerónimo* and Antonio López*

* *Computer Science Department, Computer Vision Center, Campus UAB, Edifici O, 08193, Bellaterra,, Barcelona, Spain*
E-mail:dvazquez@cvc.uab.es

Abstract

Pedestrian accidents are one of the leading preventable causes of death. In order to reduce the number of accidents, the pedestrian protection systems, a special type of advanced driver assistance systems, have been introduced. Since the appearance of pedestrians varies significantly as a function of distance to the camera. The main aim of this work is to explore if we can detect small pedestrians. We have evaluated the HOG pedestrian detector in two different datasets (INRIA and Daimler09) for two different distances (far and near). The obtained results suggest the use of different detectors for far and near distances.

Keywords: ADAS, Pedestrian detection, HOG.

1 Introduction

Pedestrian run overs represent the second largest source of traffic-related injuries. In order to reduce these road accidents, appear the Pedestrian Protection Systems (PPSs), a special type of Advanced Driver Assistance Systems (ADASs), where an onboard camera explores the road ahead for possible collisions with pedestrians in order to warn the driver or performing braking actions [2, 3]. Pedestrian detection is a challenging task. The main challenges rely on the pedestrian appearance variability due to clothes, poses or sizes and the context where they can be found.

Pedestrian detection produced a vast amount of techniques, models, features and general architectures. Then, it is difficult to compare and study them. Recently, a pedestrian detection survey has been presented [3] which proposes a general module-based architecture and reviews different approaches with respect to the tasks defined in the proposed architecture. This work points out relevant aspects for future research in PPS. Important ones are the lack of good databases and benchmarking protocols as well as exploring the effect of the distance from pedestrians to the camera.

The aim of this work is to explore these two PPSs aspects. This idea of exploring the effect of the distance has been reinforced by Enzweiler and Gavrila's work [2]. We want to answer the following questions: (1) which are the most adequate datasets from the latest ones? (2) for detecting far away people, which is the difference between training a system with actual small pedestrians and training it with scaled pedestrians? (3) which is the performance for each size of the pedestrians and why? (4) how do the optimum parameters of the method vary with respect to the distance?

Given that pedestrians closer to the camera are seen with more detail than the ones which are further away (see Fig. 1 and 2), a study on the benefices of training different classifier models depending on the target distance is of key interest. Since distant pedestrians are smaller and have less details, they tend to be more difficult to classify.

Closer targets present more details and their classification is easier but, the reaction time should be lower. This leads us to think that a system based on multiple classifiers, each specialized on different depths is likely to improve the overall performance with respect to a typical system based on a single general detector.

In this work we train the HOG pedestrian detector in two dataset (INRIA and Daimler09) and for two different distances (far and near) assessing their usefulness by both per-window and per-image evaluation.

The remainder of this paper is organized as follows. In Sect. 2 we overview the benchmark datasets and in Sect. 3 the detection approach needed to study the effect of the distance. The experiments, the evaluation criteria and the results are explained in the Sect. 4 which finalizes with a discussion of the results. Finally, in Sect. 5 we summarize the conclusions of this work, and draw some future work.

2 Pedestrian datasets

Existing datasets can be grouped in two types: (1) *person* datasets that contain non-occluded people in different poses and backgrounds but with a restricted point of view and, (2) *pedestrian* datasets that contain upright or partially occluded pedestrians in an urban environment and usually with motion information and more complete labellings.

Among the *person* datasets we select the INRIA one [1] that remains the most widely used and it is a spread reference in pedestrian detection. Among the *pedestrian* datasets we use the Daimler09 [2] because it is the appropriate dataset for studying the effect of the distance given that it has a lot of examples from different sizes to train and also provides a video sequence fully annotated that allows us to evaluate the experiments.

Figure 1: Images from the INRIA dataset. Top: images with pedestrians. Bottom: cropped pedestrians sorted by distance.

Figure 2: Images from the Daimler09 dataset. Top: images with pedestrians. Bottom: cropped pedestrians sorted by distance.

3 Pedestrian detector

Dalal et al. [1] proposed a pedestrian detector based on Histogram of Oriented Gradients (HOG) features inspired on SIFT and a SVM learning machine. It is the reference in the state-of-the-art of pedestrian detection. These features model the shape and appearance using normalized histograms of the image gradient orientation. The idea is to divide the image with a dense spatial grid in small regions called *cells*. A cell is represented as a histogram of its local gradients binned according to their orientation and weighted by their magnitude. These cells are grouped in larger regions called *blocks*. A block is represented as a feature vector formed by concatenated and normalized histograms of its cells. The final descriptor is a feature vector formed by all the blocks attached and it is classified using a linear SVM. To compute the features we use the parameters suggested by the authors.

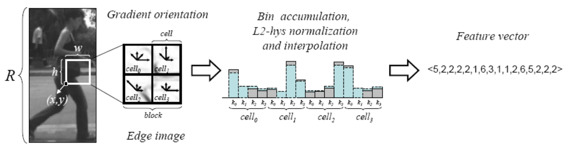

Figure 3: histogram of oriented gradients. (Reproduced from [4])

4 Experiments

Our experiments consist in evaluating the HOG classifier in the selected databases (INRIA and Daimler09) and for two different distances (near and far).

4.1 Evaluation methodology

To evaluate the experiments, there exist two established methodologies: *per-window* and *per-image* evaluation. In the per-window approach (Fig. 4), the detector is evaluated by classifying cropped pedestrians versus non-pedestrians crops and the performance is shown in the Detection Error Trade off (DET) curve that plots the miss-rate versus the False Positive Per analyzed Window (FPPW). In the per-image approach (Fig. 5), an image is given to the detector and it returns a list of Bounding Boxes (BB) with a given confidence. In this case, the detector scans the image by a sliding window approach and clusterize the detections with a Non-Maximum Suppression (NMS). The evaluation consists in performing a correspondence between the BB detections, namely BB_{dt} and the BB groundtruth, BB_{gt}. To compare methods we employ the False Positives Per Image (FPPI).

Usually in the per-window evaluation to compare the results of two methods we look the missrate value at FPPW of 10^{-4} over the DET curves and in the per-image evaluation we compare the missrate values at 1 FPPI.

4.2 Results

To compare our results with the obtained by other authors we use the INRIA dataset. Figure 1a shows that our implementation of HOG gives the same results than the original HOG of Dalal and how the bootstrapping iterations improve the results. In Figure 1b we can see some images of detections and in Figure 6 we can see the obtained models for the detector.

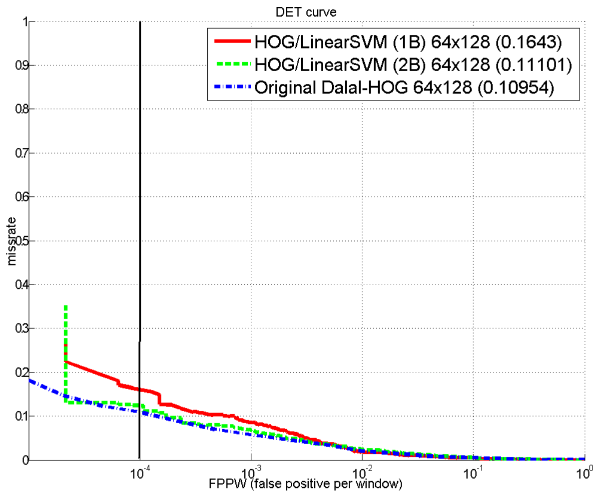

Figure 4: Per-window evaluation of the HOG and in the INRIA dataset. The missrate at the interesting point of false positives is the number inside the parenthesis.

To obtain small pedestrian samples we downscale the images of the two datasets to a smaller size: 32x64 for INRIA and 24x48 for Daimler09. Then we train and test in the per-window evaluation over the scaled images. And in the per-image evaluation we use the classifiers trained with the

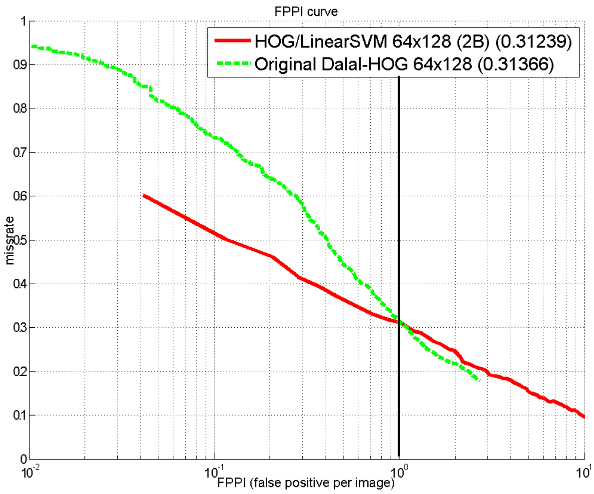

Figure 5: Per-image evaluation of the HOG and in the INRIA dataset. The missrate at the interesting point of false positives is the number inside the parenthesis.

small pedestrians to detect the big ones in order to compare the classifiers over the same set of images. Analyzing Table 1 it can be appreciated that for the HOG detector the bigger the image is the better the detector performs.

INRIA	Per-window		Per-Image	
	32x64	64x128	32x64	64x128
HOG	0.30	0.11	0.60	0.31

Daimler09	Per-window		Per-Image	
	24x48	48x96	24x48	48x96
HOG	0.60	0.23	0.32	0.23

Table 1: Performance evaluation of the HOG method in the INRIA and Daimlero9 datasets.

To evaluate if there is any difference between training with these scaled pedestrians and training with actual small pedestrians, we split the Daimler09 set into a training and a testing sets and we train and test a HOG. The obtained results with the actual small images is slightly better in the per-image evaluation: a 3% of improvement.

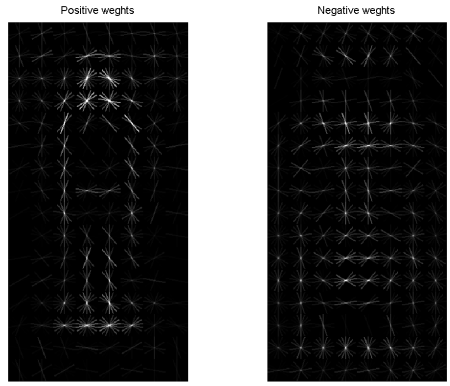

Figure 6: Models learned by the HOG method in the INRIA dataset.

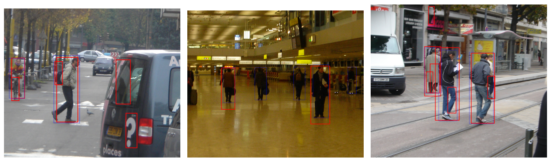

Figure 7: Some detections of the HOG method in the INRIA dataset. The blue BBs are the groundtruth and the red ones are the detections.

To find the method optimum parameters we tuned the HOG parameters in the same way as Dalal et al. suggest in [1]. Among the parameters that can be optimized we optimize only the size of the cells and blocks and the orientation bins. From the results it seems that a fine binning (9 orientation) and large scale features (blocks of 2x2 cells of 8x8 pixels) are the best parameters for INRIA and Daimler09 datasets.

4.3 Discussion

At the beginning of the work raised some questions and after the experiments, some other questions appeared. Let's answer them.

- ***Should a pedestrian detection system take into account the pedestrian distance?:*** Experi-

Figure 8: Some detections of the HOG method in the Daimler09 dataset.

ments suggest to use multiple detectors specialized in distances is better than to have only one detector for all the distances.

• ***How does affect the size of the training images in the optimum parameters?:*** After the study of the HOG parameters we can conclude that there is a set of canonical parameters that performs well for all the cases and they are not affected by the pedestrian size.

• ***In order to learn a classifier for far away pedestrians, do we need samples with small pedestrians for training or is it enough to downscale the big ones?:*** Results show that the classifier trained on the actual pedestrians is slightly better than the other. Thus, we expect that if we get a training set with more actual small pedestrians this performance difference could be higher.

• ***What are the differences between the performances obtained by the per-window and the per-image evaluation?:*** We have seen that the per-window performance could be not very realistic as it does not take into account the errors caused by the sliding window and the NMS.

• ***Which are the most suitable datasets for PPSs?:*** After working and analyzing the datasets we have seen that the Daimler dataset is more suitable for our purposes: it has many more examples, well labeled and at several scales. The problem of this dataset it is that the original frames from which the training samples were extracted are not available.

5 Conclusions

We have explored two interesting aspects to develop a pedestrian detector: the datasets used for learning the classifiers and the effect of the distance in the detection. To study this effect we have used two datasets (INRIA and Daimler09) and the HOG pedestrian detection method. The evaluation has been done following two different approaches (per-window and per-image) for two different distances (far and near). After the discussion we have realized that the distance is of key importance in PPSs. Therefore, we propose to build a system that combines specialized classifiers optimized for different ranges of distance to improve its performance.

References

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893, 2005.

[2] Markus Enzweiler and Dariu M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(1), 2008.

[3] David Gerónimo, Antonio M. López, Angel D. Sappa, and Thorsten Graf. Survey on pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[4] David Geronimo, Angel D. Sappa, Antonio Lopez, and Daniel Ponsa. Adaptive image sampling and windows classification for on-board pedestrian detection. *Proceedings of the International Conference on Computer Vision Systems*, 2007.