Using Context to Improve Cascaded Pedestrian Detection

Mohammad Saberian, Zhaowei Cai, Jinhee Lee, Nuno Vasconcelos

University of California San Diego
9500 Gilman Drive
La Jolla, California, USA
saberian@ucsd.edu, zwcai@ucsd.edu, jini92.lee@nextchip.com, nvasconcelos@ucsd.edu

Abstract

The design of a fast and accurate pedestrian detector is considered. A system combining a fast cascaded pedestrian detector and a pedestrian validator is proposed. The detector first scans the image of interest and proposes a set of candidate bounding boxes. The pedestrian validator then decides if each proposed bounding box is consistent with a true pedestrian, based on scene context. Experiments show that the resulting system is faster and more accurate than current approaches to pedestrian detection.

Keywords-component; pedestrin detector; cascade detector; pedestrain validator and boosting

Introduction

In recent years, significant attention has been devoted to driver assistance systems and self-driving vehicles. An important problem for these systems is to detect and localize pedestrians that could be harmed by a vehicle. The design of such pedestrian detectors is challenging because 1) real-time operation requires very fast detectors and 2) the detection must be very accurate. Most notably, these systems must detect all pedestrians in the field of view while guaranteeing a very low false positive rate, so as to avoid false alarms that can ultimately lead drivers to ignore them.

While many methods have been proposed for pedestrian detection, most pedestrian detectors with acceptable accuracy and real-time performance are based on the cascade structure of [1]. In this architecture, a fast pedestrian classifier is trained on a large dataset of carefully cropped pedestrian examples. The detector is then applied in a sliding-window fashion, i.e. evaluated at all possible candidate regions of the image where the detection is to be performed. The main difficulty of this approach is that the decision about each candidate region is independent of the information outside of that region. However, this information can be a very useful aid to the detection process, especially when the image is of a structured scene. For example, Fig 1. shows a street scene along with detection outputs of a typical cascaded pedestrian detector. The detector was able to successfully detect two pedestrians, but also detected two false-positives. However, in the context of the larger scene, these false-positive cannot be valid pedestrians: the left one would corresponds to a 3-meter tall person, the one in the center to a pedestrian that walks 2 meters above ground.

In this paper, we propose a system to include context information in the detector so as to identify invalid

Fig. 1, detection result of a typical pedestrian detector

false-positives. In the proposed system, the pedestrian detector first scans the image and proposes a set of candidate bounding boxes. A pedestrian validator then decides if the proposed bounding boxes are consistent with real pedestrians, in the context of the whole image. Experiments show that the use of an effective pedestrian validator helps to eliminate false positives. In addition it is shown that, by integrating the pedestrian validator within the pedestrian detector, it is possible to shrink the scanning region and speed up the pedestrian detector itself.

Proposed System

The proposed system is illustrated in Fig. 2. The pedestrian detector first scans an image and proposes a set of candidate bounding boxes for the locations most likely to contain a pedestrian. The pedestrian validator then uses information from image context to decide if the proposed bounding boxes are consistent with true pedestrians. For applications like driver assisted systems, where pedestrian locations have high regularity across scenes, the validator can be trained a priori. This results in a set of scale-dependent locations for the bounding box, as shown in Fig. 3. The detector can, in turn, use this information to avoid scanning regions that do not conform to valid pedestrians.

A. System implementation

The cascaded detector is implemented with the algorithm of [6]. To design an accurate yet simple pedestrian validator we used a data driven approach. Rather than hard coding parameters of the scene, such as camera location inside the vehicle, we trained a classifier to discover the properties of valid bounding boxes from a training set. To train this classifier, we collected a set of random bounding boxes as negative examples. For positive examples, we collected the bounding box information for all non-occluded pedestrians in the training set of Caltech Pedestrian dataset [2]. From each bounding box we extract the following feature vector

$$\left[\frac{i}{H}, \frac{j}{W}, \frac{h}{H}, \frac{w}{W}\right], \quad (1)$$

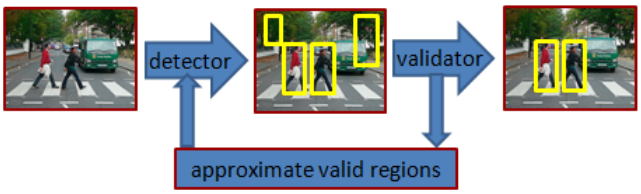

Fig. 2, proposed system for pedestrian detection

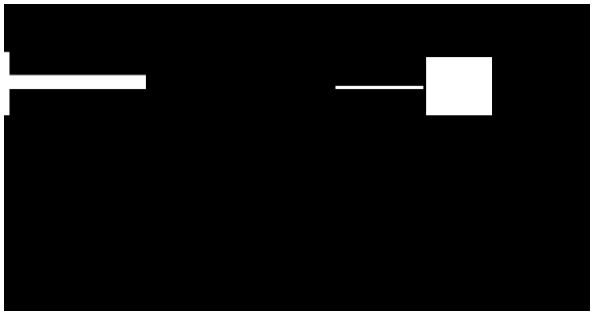

Fig. 3, white pixels show the possible locations for upper-left corner of pedestrians of height 260 pixels in a 640×480 images

where (i,j) are coordinates of the upper left corner of the bounding box, (h, w) its height and width and (H, W) the height and width of the image. This feature vector encodes relative size and location of pedestrians observed by a camera mounted on a vehicle. While the training process should be repeated whenever the camera position changes, this is not expected to be a problem for driver-assistance, where pedestrian detection is based on cameras installed by the manufacturer during vehicle assembly.

The main difficulty in training the classifier to discriminate between the sets of random and true bounding boxes is that the two sets have a significant overlap. In fact, the second set is almost completely inside the first, as most pedestrian locations are also possible locations for pedestrian absence. Hence, training a validator with 100% accuracy is impossible. To overcome this problem, we propose to design a classifier that, instead of minimizing the error rate, accepts all positive examples and minimizes the false-positive rate. In this way, the combined pedestrian+validator system accepts all pedestrians and rejects as many false positives as possible. Since this is a special case of cost-sensitive classification, we have used the cost-sensitive boosting approach of [7] to train the validator. The white pixels of Fig. 3 show the learned locations for the upper-left corner of pedestrians with height of 260 pixels in a 640×480 image. This is roughly the size of the left false-positive in Fig 1. As the map of Fig.3 suggests, to correspond to a true pedestrian, the bounding box should have a much higher upper-left corner. Hence, the validator rejects the bounding box. In addition, the regions of acceptance by the validator can be approximated by rectangles and the detector itself applied only inside these rectangles. This shrinks the scanning region and speeds up the overall detection.

Experiments

We compared the proposed system to a set of state-of-the-art pedestrian detectors on the Caltech Pedestrian dataset [2]. Similarly to Dollar *et al.* [3] we adopted an image representation based on a 10 channel decomposition. This included 3 color channels (LUV color space), 6 gradient orientation channels, and a gradient magnitude channel. The performance of the proposed system was evaluated with the toolbox of [2]. The comparison is based on detecting

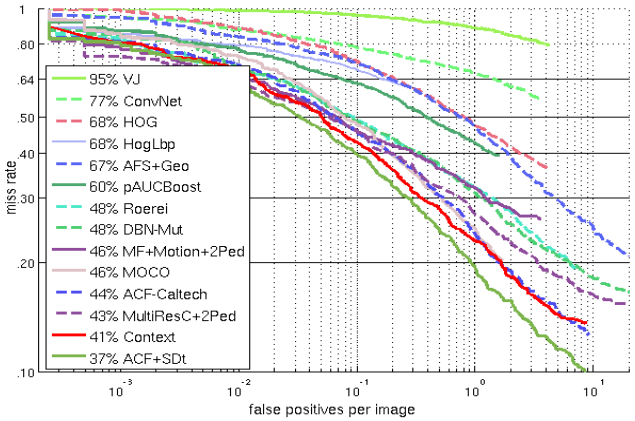

Fig. 4 miss rate vs. FPPI rate for of various pedestrian detectors. The number on the left of each legend is the log-average miss-rate.

pedestrians of height at least 50 pixels in 640×480 images. This is equivalent to detecting pedestrians about 40m away from the vehicle. Fig. 4 presents the results of this comparison. The numbers shown on the left of the legend summarize the detection performance by the log-average miss-detection rate. The set of benchmark detectors includes popular architectures, such as HOG [4] or the deformable part model of [5]. The proposed method outperforms all detectors other than ACF+SDt [6], which (unlike ours) uses motion information. The most direct comparison is ACF-Caltech [6], which uses a similar detector but ignores context information (no validator). The proposed system has better accuracy, i.e. 44% vs. 41% log-average miss-detection. In addition, by using the context information inside the detector and shrinking the scanning area, we were able to speed up the detector by 25%.

Acknowledgment

This work was supported by NSF grant (NSF IIS-1208522) and the Technology Development Program for Commercializing System Semiconductor funded By the Ministry of Trade, industry \& Energy(MOTIE, Korea), [No. 10041126, Title: International Collaborative R\&BD Project for System Semiconductor].

References

[1] P. Viola and M. Jones. “Robust real-time object detection”. Workshop on Statistical and Computational Theories of Vision, 2001

[2] P. Doll´ar, C. Wojek, B. Schiele, and P. Perona. “Pedestrian detection: An evaluation of the state of the art”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):743–761, 2012.

[3] P. Doll´ar, Z. Tu, P. Perona, and S. Belongie. “Integral channel features”. In Proceedings of British Machine Vision Conference, 2009.

[4] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection”. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 886–893, 2005.

[5] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. “Object detection with discriminatively trained part-based models”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.

[6] P. Doll´ar, R. Appel, S. Belongie, P. Perona. “Fast Feature pyramids for Object Detection”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014.

[7] P. Viola and M. Jones. “Fast and robust classification using asymmetric adaboost and a detector cascade”. In Proceedings of the Neural Information Processing Systems Conference, pages 1311–1318, 2002.