Virtual Scenarios: Achievements and Current Work

Javier Marín, David Vázquez and Antonio M. López

ADAS, Computer Vision Center, Universitat Autonoma de Barcelona, Spain
e-mail: {jmarin, dvazquez, antonio}@cvc.uab.es

Abstract

With this paper, we introduce the work [3] presented at the Conference on Computer Vision and Pattern Recognition (CVPR) 2010 and the one we are currently developing. At the conference, we answered the question of whether or not it was possible for a pedestrian classifier based on 2D appearance and learned in a virtual world to be successfully applied in a real world. The results suggested a positive answer. Afterwards, with the reviews of the conference, we have been extending our work which will be published in a journal article. This extension has involved the use of more real datasets like the INRIA and the CVC ones, as well as new different state-of-the-art algorithms used for pedestrian detection. Besides, a new set of virtual pedestrians has been introduced in order to achieve more variability between pedestrians. In the same way, new textures have been applied and higher graphical definition has been inserted. Finally, our conclusion suggests that the current results are promising.

1 Introduction

Advanced driver assistance systems (ADAS) aim to improve traffic safety by providing warnings and performing counteractive measures in dangerous situations. Pedestrian protection systems are specialised in vehicle-to-pedestrian collisions. They consist in vehicles equipped with a forward facing image acquisition and processing system able to detect pedestrians on the road. Accordingly, research on image-based pedestrian detection for this task has been a very relevant topic for the Computer Vision community [12, 9, 8]. The challenge lies in the fact that pedestrians are very difficult to detect: they are articulated and imaged from a mobile platform in cluttered scenarios and present high variability in clothes, pose, distance to the camera, background and outdoor illumination. Moreover, the nature of the addressed application requires real time and a demanding tradeoff between misdetections and false alarms.

The most promising pedestrian detectors rely on appearance-based pedestrian classifiers learnt from labelled samples, i.e., examples (pedestrians) and counterexamples (background). Having sufficient variability in the sets of examples and counterexamples is decisive to train classifiers able to generalise properly [1]. Unfortunately, obtaining the desired variability in such sets is not easy for pedestrian detection since we cannot control the real world while recording video sequences.

In order to face this problem, Broggi et al. [2] use synthesized examples for pedestrian detection in far infrared images, i.e., images capturing relative temperature. However, the authors admit poor results, since it is difficult to handle variability due to different clothes, person size, more complex background and, in addition, computational time increases with the number of templates to be con-
Enzweiler et al. [7] enlarge the set of examples by transforming the shape of pedestrians (labelled in real images) as well as the texture of pedestrians and background. The pedestrian classifier is learnt by using a discriminative approach (NNs with LRFs and Haar features with SVM are tested). Nonetheless, for applying the different proposed transformations the overall pedestrian silhouette must be traced, which requires a manual labelling much more labour intensive than standard bounding box framing of pedestrians.

Rather than using rough morphological models or synthesized real examples, we proposed to explore the synergies between modern Computer Graphics and Computer Vision in order to close the circle: the Computer Graphics community has modelled real world by building increasingly realistic virtual worlds (e.g., video games). Thus, could we now learn our models of interest in such virtual worlds and use them successfully back in real world? In [3] we focused this question on the visual appearance of pedestrians. In particular, we learned such appearance using virtual samples in order to detect pedestrians in real images.

The experiments we conducted suggested a positive answer to the previous question, which was a new and relevant result for research in pedestrian detection. Currently, we are extending our work by using more pedestrians datasets, features and learning methods.

The remainder of the paper is organized as follows. Section 2 introduces the achievements reached in the last paper. Section 3 details the current work and, finally, Section 4 concludes it.

2 Achievements

In [3] we explored how realistic virtual worlds could help in learning appearance-based models for pedestrian detection in the ADAS area. We used the HOG/linear-SVM technique to learn a pedestrian classifier using only samples from virtual worlds. We plugged-in such classifier in a standard pedestrian detection method and evaluated how this detector worked when applied to real images, i.e., when the pedestrian classifier is used out of its world. The same procedure was followed to obtain an analogous pedestrian detector that only differs from the virtual-world-based one in the sense that the plugged-in classifier was trained using real images (i.e., manually labelled pedestrians). Comparison between virtual and real world based pedestrian detectors revealed a rather close performance. The size and complexity of Daimler’s testing set allowed for this conclusion to be reliable. Therefore, the results presented in [3] were new and relevant for research in pedestrian detection. However, to provide a totally definitive conclusion we suggested testing with more databases, more features and more learning machines.

3 Current work

The future experiments we introduce above are part of our current work. The specific datasets, features and learning machines are detailed below.

3.1 Pedestrian datasets

The pedestrian datasets we are currently working with are the following ones:

Daimler09. During the development of [3], Enzweiler et al. [12] published their pedestrian dataset. The test set contains a sequence with more than 21,790 images with 56,492 pedestrian labels (fully visible or partially occluded), captured from a vehicle during a 27 min drive through urban traffic, at VGA resolution (640x480, uncompressed).

INRIA. Currently, the INRIA person dataset [4] is the most widely used in pedestrian detection. In these images people are usually standing, but appear in any orientation and on a wide variety of
background image including crowds. The data set contains images from several different sources, images from pedestrian datasets, images from personal digital image collection and some images from the web.

**CVC-02: Classification-and-System.** CVC-02 consists of three subsets, each one focuses on a different task of pedestrian detection: candidate generation, classification and system evaluation. The imagery has been recorded in urban scenarios around Barcelona (Spain), using a Bumblebee colour stereo camera with resolution 640x480 pixels and 6mm focal length. The annotated pedestrians are in the range from 0 to 50 m from the camera, which corresponds to a smallest pedestrian of 12x24 pixels. The main features of each subset are the following:

- **CVC-02-Classification:** training (1016 cropped positives and mirrors, 7650 cropped negatives and pedestrian-free images), testing set (window-based, with 570 cropped positives and mirrors, 7500 cropped negatives and pedestrian-free images) and testing set (image-based, with 250 frames containing 587 annotated pedestrians).

- **CVC-02-System:** 15 sequences of 4364 frames, 7983 pedestrian annotations. All the images are provided in lossless PNG format, both in color and depth versions, and in their original size and 64x128 pixels rescaled. Regarding the annotations, we label them as obligatory or optional (very young children, significantly occluded or partially out of the image).

### 3.2 Pedestrian detection methods

In the state of the art of Object Classification a wealth of methods have been proposed for PPS. Thanks to the recent surveys published by M. Enzweiler and D. Gavrila [6] in 2008, by Gerónimo et al. [9] in 2009 and the paper published by Dollár et al. [5] in 2009, the wearisome task of summarizing, comparing and evaluating the different methodologies and datasets present in literature has been made easier.

For [3] we just used the pedestrian detection method presented in 2005 by N. Dalal and B. Triggs [4]. In such work, they demonstrated that using a linear SVM based on grids of Histograms of Oriented Gradient (HOG) as descriptors, significantly outperformed existing feature sets for human detection.

The new pedestrian detection methods we are now using rely on appearance-based pedestrian classifiers learnt from labelled samples:

- **Haar wavelets and edge orientation gradients.** This classifier, which consists of Haar wavelets and Edge Orientation Histograms (HaarrEOH) as features and Real AdaBoost as learning machine, have been originally proposed by Levi and Weiss to perform face detection in [10].

- **HOG with non-linear SVM kernels.** These classifiers, which consist of Histogram of Oriented Gradients (HOG) as features and Support Vector Machines with non-linear kernels (RBF, Intersection, Chi-square and Hellinger) as learning machine, have been originally proposed by Dalal but used only the Radial Basis Function (RBF) kernel. Then in [11] Maji proposed to use the intersection kernel and Vedaldi proposed to use the Intersection, Chi square and the hellinger kernels.

### 3.3 Virtual scenario improvements

A new set of virtual pedestrians have been introduced in order to achieve more variability between pedestrians. In the same way, new textures have been applied and higher graphical definition has been inserted.
4 Conclusions

In this paper we have discussed the work we presented in June 2010 at the CVPR conference, in which the comparison between virtual and real world based pedestrian detectors revealed a rather close performance. Therefore, the results presented in such conference were new and relevant for research in pedestrian detection. Continuing our work, we have introduced more datasets, more features and more learning machines. The current results are promising, providing more validity to the final conclusion exposed at the CVPR.

References