Pedestrian detection to dynamically populate the map of a crossroad

Ujjwal

September 2015

1 Introduction

The problem of "Pedestrian Detection" or "People Detection", henceforth referred to as "**PD**" problem, involves detecting human beings in real-life images and videos. While this is a just another instance of object detection, its implications have a huge impact on a wide variety of technologies especially surveillance. Being a heavily explored area, supported by availability of a wide variety of datasets, there exist a number of popular efforts which address this problem. While this document is not a detailed exposition of literature on PD, we examine the need of undertaking a new approach towards addressing this problem. Thus in this document, we look at the current state of readiness in which PD systems exist to address practical problems, and examine some scenarios of practical importance which require a more detailed and further examination.

2 Current state of Pedestrian detection systems

A wide variety of literature summarizing PD systems exist. [1] is a recent survey, which examines various aspects of the current PD systems. According to [1], the present state-of-art in PD systems are able to achive an impressive performance with a miss-rate of 22.5% on the Caltech dataset. This state-of-art approach uses a random forest classfier and combines a number of published strategies for PD. The false negatives in the present systems primarily stem from contextual reasons such as heavy illumination variations, high density crowd contexts where individual delineation of people is non-trivial and uncommon poses and clothing which have little representation in datasets. At the same time it is not very easy to comment on reasons owing to false positives because that requires a detailed analysis of feature representation and actual information that is encoded by features and classifiers. [1] also outlines that major improvements in PD systems have resulted from the use of better features and to some extent from the use of contextual information. The capacity of trained models to transfer expertise across datasets is however still in need of improvements.

It is important to mention that depending on the nature of datasets and specific settings of cameras, a multitude of different methods have been proposed for PD. While a lot of work has been focused on detecting pedestrians from single-camera static images, a number of methods take advantage of temporal information (e.g - optical flow) and stereo information (if available) to get more successful performance figures. A number of such approaches have led to a rich ensemble of datasets that are widely used today for PD. In the next section we take a brief look at the variety of datasets that are available for the task of PD.

3 Datasets

With respect to examining the current state-of-art in PD systems, it is necesary to take a look at datasets of PD, for PD systems are benchmarked with respect to certain datasets. A wide variety of datasets are publicly available for the task of PD. Table 1 lists a few popular public datasets for pedestrian detection. These datasets exhibit a lot variability in terms of image size and content.With respect to surveillance applications, a greater variety of datasets is required. A wide variety of illumination settings and weather conditions are common examples of surveillance conditions. As an example, all the present public datasets focus on pedestrian detection with respect to daylight settings. Moreover in surveillance applications it is quite often to encounter pedestrian images taken from a very high viewpoint. The task of high-capacity machine learning is inherently tied to the amount of data available. This points to the need of making a greater amount of data available which realistically potray the data as would appear in a practical system incorporating PD.

We have also collected a large repository of pedestrian data in real-life scenarios that can serve as suitable benchmarks for training and testing of PD systems. We call these datasets as Vanneheim and Supermarket datasets. Our datasets contain videos taken from single as well as multiple cameras. These videos are taken over several hours at different times and thus provide an excellent way to train and test PD systems under a wide array of conditions. Our videos depict real depiction of activities in public subways, supermarkets and airports. Some of these images are fully annotated while some require annotations. We have also collected datasets for re-identification that can be used for detection as well as for advanced surveillance applications for re-identification. With a n ovel and systematic use of these datasets it is possible to envision a high-performance PD system. There are other datasets such as Cipebus and Caviar that can also be used for this purpose.

4 A Novel approach

With the proliferation in the number and size of datasets, as well as an increase in the demands of modern surveillance systems, it is important to increase the robustness of PD systems. The idea of robustness becomes very important when

Dataset	**Channels**	**Temporal Information**	**View-point**	**Calibration Information**	**Stereo**	**Segmentation**
Caltech	Color	Yes	Car-View	No	No	No
Daimler Mono Pedestrian	Grayscale	No	Car-View	Yes	No	No
Daimler Stereo-Pedestrian	Grayscale	No	Car-View	Yes	Yes	No
TUD-Brussels	Color	Yes	Car-View	No	Yes	No
ETH-Zurich	Color	Yes	Car-View	Yes (with odometry)	Yes	No
INRIA	Color	No	Head-Level View	No	No	No
PETS	Color	Yes	Top-Level View	Yes	No	No
Viper	Color	No	Head-Level View	No	No	No
Daimler Pedestrian segmentation	No	Car-View	No	No	No	Yes

Table 1: A categorical listing of some popular pedestrian detection datasets. The listing is in decreasing order of dataset size.

it is realized that pedestrians can appear in an instance with a very wide variety of poses, clothing and orientation. This is in addition to the large number of extrinsic factors like illumination and weather.

With availability of large amount of data, it is important to organize the data effectively and use it systematically for machine learning. An effective machine learning system for PD has to take into account, the encoding of all real-time intrinsicc and extrinsic factors in such a way that simple to complicated instances of pedestrians can be successfully detected. Towards this, we envision exploring the concept of using the concept of boosting by incorporating a large number of specialist classifiers and using them together, for a successful PD. While boosting implies combining multiple classifiers, this combination can be done in multiple ways and exploring an optimal combination is a relevant and open problem that has to be explored. It is important to properly explore large datasets and effectively design effective classifiers, so that the final learning has a semantic sense to it. This is important owing to the fact that a successful PD system is expected to be able to encode the semantic understanding of a pedestrian. With a proper semantic understanding of pedestrians, a PD system can then be used to populate a dynamic map in an environment which can be used for autonomous driving and surveillance applications

5 Conclusion

While pedestrian detection has come a long way, and excellent PD systems which work well in conventional instances are available, in order to push the boundary towards more intense scenarios work is still required. With better PD systems, it becomes possible to envision high-fidelity autonomous driving and better surveillance systems for real-time monitoring in a wide variety of conditions.

References

[1] Rodrigo Benenson, Mohamed Omran, Jan Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? In *Computer Vision-ECCV 2014 Workshops*, pages 613–627. Springer, 2014.