Transportation Research Part C 25 (2012) 226–237

Contents lists available at SciVerse ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Overview Paper

Stereo regions-of-interest selection for pedestrian protection: A survey

D.F. Llorca [a,*], M.A. Sotelo [a], A.M. Hellín [a], A. Orellana [b], M. Gavilán [a], I.G. Daza [a], A.G. Lorente [a]

[a] *Computer Engineering Department, Polytechnic School, University of Alcalá, Madrid 28871, Spain*
[b] *Institute of Automatics, Faculty of Engineering, National University of San Juan, 1109 San Juan, Argentina*

ARTICLE INFO

Article history:
Received 12 July 2011
Received in revised form 14 June 2012
Accepted 15 June 2012

Keywords:
Pedestrian detection
Regions of interest selection
Intelligent vehicles
Intelligent transportation systems
Stereo
Survey

ABSTRACT

Vision-based pedestrian detection for intelligent vehicles applications is a crucial and active research area due to the essential benefits in terms of reducing the number of accidents involving pedestrians and vehicles. During the last decade a considerable amount of research studies have been proposed, filling the gap between prototypes and commercial implementations. Pedestrian detection systems can be roughly divided into three main different sub-parts: *Region Of Interest – ROI – selection*, *classification* and *tracking*. Previous surveys have covered the literature in a holistic way. An example would be, analyzing all the solutions proposed for all the stages and including higher level analysis, but in most cases they give more emphasis to the classification stage. Due to the difficulty of this detection task, the variety of solutions, sensor configurations (monocular/stereo; visible/infrared) available in the literature, we propose to break down the variability of the problem by providing exhaustive review of one specific stage: *stereo-based ROI selection*. ROI selection is a key component that has to be designed to provide generic obstacles at lowest false negative rate and maintain a low number of false positives. The number of missed pedestrians has to be approximately equal to 0 since a pedestrian missed by the ROI selection stage would not be detected in further stages. In addition, the number of non-pedestrians obstacles should be as low as possible to reduce both the number of false alarms and the computational costs of further stages. In contrast to monocular approaches, stereo ROI selection determines the relative distance between the pedestrian and the vehicle, assuring that the reported candidates are related with real physical objects. The stereo-based ROI selection step is also divided into different components that are independently analyzed, increasing visibility for future proposals and developments. Discussion is finally presented highlighting the current problems for obtaining a global overview of the actual performance of the different approaches and analyzing future trends.

1. Introduction

Pedestrian detection is a fundamental task for a variety of important applications, especially in the context of intelligent vehicles (IVs) and intelligent transportation systems (ITSs), since it clearly enhances the pedestrian safety. Every year, according to the statistics estimated by the World Health Organization (Peden et al., 2004), 1.2 million people are known to die in road accidents worldwide. A majority of the deaths and injuries involve motorcyclists, cyclists and pedestrians. Only in the European Union about 8000 pedestrians and cyclists are killed and about 300,000 injured each year. In North America, approximately 5000 pedestrians are killed and 85,000 injured. In Japan approximately 3300 pedestrians and cyclists are killed and 27,000 injured (UNECE, 2005). Over the last decade, this topic has attracted an extensive amount of interest from

* Corresponding author. Tel.: +3491 885 6641; fax: +3491 885 6682.
E-mail address: llorca@aut.uah.es (D.F. Llorca).

0968-090X/$ - see front matter
http://dx.doi.org/10.1016/j.trc.2012.06.006

national and international authorities, the automotive industry and the scientific community, aiming at improving the safety of the most vulnerable road users.

Active sensors such as acoustic-based, radar-based and laser-based have been proposed for pedestrian detection. Refer to the survey presented by Gandhi and Trivedi (2007), to have a broad overview of these active sensors-based approaches. However, during the last years, passive sensors, and more specifically optical sensors, have attracted most of the attention of the research community as well as the industry, due to two main aspects: inexpensive costs and new potential applications such as *Lane Departure Warning*, *Traffic Signs Recognition*, and *Adaptive Cruise Control*.

Pedestrian detection is a difficult task from the computer vision perspective. Large variations in pedestrian appearance, e.g. pose, clothing, size, etc., and environmental conditions, e.g. lighting, moving background, etc., make this problem particularly challenging. Vision-based detection systems can be classified with respect to the number of cameras, monocular or stereo, as well as depending on the spectrum, visible or infrared. In addition, vision-based pedestrian detection systems can be roughly divided into three main stages (see Fig. 1). The first stage consists of identifying generic obstacles as regions of interest, ROI selection, using prior scene knowledge: camera calibration, stereo information, ground plane constraint, etc. Subsequently, a more expensive pattern recognition step is applied: classification or verification. The lack of explicit models leads to the use of machine learning techniques, where an implicit representation is learned from features obtained from thousands, or millions, of samples. Finally, temporal integration or tracking stage is applied to improve single-frame detection performance and smooth the relative vehicle-to-pedestrian trajectory.

As depicted in Fig. 1, the different sub-parts are sequentially linked, that is, ROI selection outputs are fed to the classifier and classifier outputs are used as inputs for the tracking step. Accordingly, the performance of each stage is related to some extent with the performance of previous stages. For example, if the classifier fails when recognizing pedestrians, tracking stage would not be able to follow them. However, if a pedestrian has been detected and tracked during a considerable number of frames, tracking can absorb spurious classification errors. Thus, tracking performance has to be evaluated in the context of a ROI selection and classification ensemble. Additionally, the classifier results are strongly correlated with the type of samples provided by the ROI selection module, not only in terms of computational costs, i.e. the higher (lower) the number of samples to classify, the greater (lesser) the time needed, but in terms of both detection rate and false positive rate (Alonso et al., 2007). Actually, it is recommended to train the classifier with samples generated by the specific ROI selection mechanism in order to optimize the detection performance (Alonso et al., 2007). If the ROI selection algorithm usually provides a specific set of false positives (e.g., poles, trees, etc.), classifier should be boosted using these samples as negative samples. Single-frame analysis is usually carried out by using specific training and test databases to obtain information about the classifier performance and to define the working point of the classifier. However, the actual performance of the classifier can only be measured in real applications working in parallel with the ROI selection algorithm.

The early stage of a pedestrian detection system (ROI selection), does not depend on previous stages (see Fig. 1) and it is probably the key component due to one of the most critical requirements: the number of false negatives has to be approximately equal to 0. If the ROI selection does not detect a pedestrian as a candidate, this one would be neither classified nor tracked by further stages. The number of false positives provided by this stage is not as critical as the number of false negatives since non-pedestrian samples can be rejected by the classifier. However the classification computational cost defines an upper bound for this number. In addition, one of the desirable features of this stage is to provide both pedestrian and non-pedestrian candidates that correspond to real physical objects, that is, to avoid ghost targets that mainly appear due to reflections and shadows.

Among the surveys in the context of pedestrian detection available in the literature we remark (Gandhi and Trivedi, 2007; Gavrila, 1999; Moeslund and Granum, 2006; Poppe, 2007; Enzweiler and Gavrila, 2009; Gerónimo et al., 2010a). Most of the work concerning human motion has been summarized in (Gavrila, 1999; Moeslund and Granum, 2006; Poppe, 2007). Focusing on the pedestrian protection application in the context of intelligent vehicles, we have found three main surveys in the literature (Gandhi and Trivedi, 2007; Enzweiler and Gavrila, 2009; Gerónimo et al., 2010a). Gandhi and Trivedi (2007),

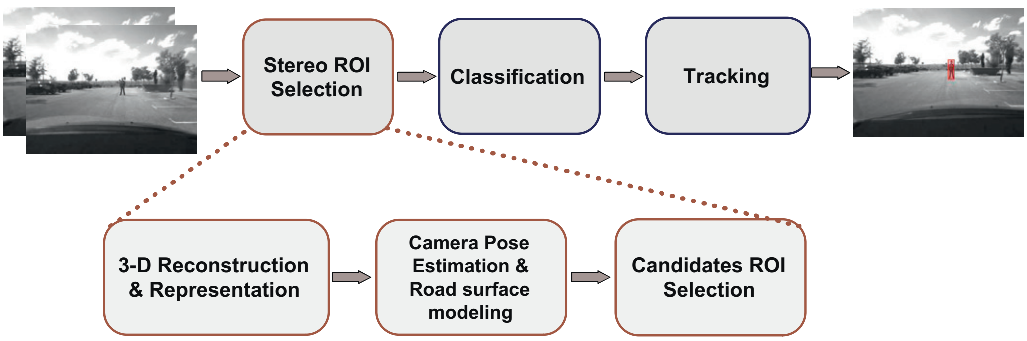

Fig. 1. Overview of the stages of a stereo vision-based pedestrian detection system. The presented survey covers the ROI selection module by means of stereo vision-based algorithms.