Multi-Feature Walking Pedestrian Detection Using Dense Stereo and Motion

Silviu Bota and Sergiu Nedesvchi
Technical University of Cluj-Napoca
Email: {Silviu.Bota, Sergiu.Nedesvchi}@cs.utcluj.ro

Abstract—This paper deals with a pedestrian pre-crash sensor based on dense stereovision and motion analysis. We aim to provide means to deploy safety measures before a crash occurs. Depending on the concrete selected actuator, it could be interesting to know the type of object (e.g., vehicle, pedestrian, bicyclist, etc.) involved in the collision. Some actuators need the information about the collision partner type, because they have to be triggered only in specific situations, or in a situation dependent manner. We present a pedestrian detection system based on dense stereo and motion, acting as pre-crash sensor. In order to detect pedestrians, we use simple features such as object size, speed, and also more complex, motion based features. All these features are combined into a flexible, Bayesian framework.

I. INTRODUCTION

Traditionally, vehicle safety is mainly defined by passive safety measures. Passive safety covers highly sophisticated constructions in the vehicle body. The occupant cell has become a more stiff structure to mitigate deformations. The frontal part of vehicles has been improved as well, e.g. it became specially designed "soft" areas to reduce the impact in case of a collision with a pedestrian. In the recent decades a lot of improvements have been done in this field. Nowadays the integrated safety approach becomes more and more attractive. Research activities focus on the combination of environmental sensors (like radar and cameras) together with deployable or conditionable safety measures. These so-called pre-crash applications are dedicated to trigger safety measures some milliseconds before the collision occurs. Two different strategies are currently under consideration: reversible and non-reversible deployment. The reversible activation can take place depending on the concrete actuator or warning in an earlier stage. The non-reversible actuators can be triggered at earliest when the situation becomes so critical that the collision is, from a driving physical point of view, unavoidable. Examples of such pre-crash systems can be found in [1], [2], [3].

A. Related work

There are various methods for object detection and pedestrian hypothesis generation. Some are based on 2-D information, for example the detection of image regions where there are a significant number of vertical edges [4]. Other methods are based on some type of additional information like IR images [4], or depth information [5],[4]. Most of pedestrian detection methods that use depth information rely on the disparity map and make some kind of segmentation on this map to detect objects [5], or use a v-disparity approach [4].

Object detection and object classification based on pattern matching are traditionally limited to 2-D image intensity information [6]. The disadvantage of using only 2-D image information is that it does not have any additional spatial information about where the objects are and what their general size is.

The 3-D information generated by a stereo reconstruction system provides depth for the objects which make up the scene. There are some classification methods that use this information directly [7], [8]. The classification based solely on 3-D information is difficult in the context of the current 3-D reconstruction systems because the information is not accurate enough to allow the reliable extraction of 3-D shapes and surfaces.

The use of pattern matching in conjunction with 3-D information has not been extensively explored, mainly because real time dense stereo reconstruction systems have not been available until recently. Some approaches aimed at pedestrian detection have used of dense 3-D information, but only as a validation method [9].

Another important feature for walking pedestrians detection is their walking pattern. There are a number of works, e.g. [10], [11] that have used motion cues in pedestrian detection.

B. Contributions

We present a novel, multi-feature pedestrian detector, useful as a pre-crash sensor. Our detector uses a stereo–
camera pair, coupled with a high performance hardware dense stereo reconstruction machine. This configuration is capable of supplying intensity (left and right) images and a range (depth) image associated with the left camera. We use an object detection algorithm, for detecting 3–D objects present in the scene. Objects are tracked across multiple frames in order to obtain stable detection, and to be able to compute motion based features for pedestrian detection. We compute a number of simple features (object width, height, lateral and longitudinal speed) and two complex features, the motion signature (the variance of the 3–D motion field generated by each object) and periodicity of this motion signature. These features are combined into a flexible Bayesian framework. Using manual labeling of a large set of objects and a custom made tool, we obtain automatically an optimal naive Bayesian classifier, which has a high rate of correct pedestrian detection.

C. Paper Structure

In section II we present our stereovision sensor. In section III we present our pedestrian detector. In section IV the experimental results obtained by using our sensor are presented, and in section V concludes our paper.

II. THE DENSE STEREO CAMERA SENSOR

Our sensor is a stereo camera system which is developed by Technical University of Cluj-Napoca/ Romania. It has an opening angle of 72 (horizontal) and detects targets up to 50 m. The stereo approach offers the opportunity to detect targets by triangulation techniques. Detection using stereo vision is, from an algorithmic point of view, less demanding as detection in pure mono vision images. After evaluating edge-based stereo algorithms in the recent years [12], the current system is based on dense stereo.

The stereo camera is designed specifically for urban environment. In inner-city conditions the density of objects is much higher than in extra-urban situations. There are many objects positioned or moving to the right or to the left of the ego’s lane. Crossing pedestrians are often seen. Narrow and curvy roads are the typical structure inside cities. These requirements are answered by our chosen sensor, especially because of its wide opening angle.

III. PEDESTRIAN DETECTION USING DENSE STEREO

Pedestrian detection is not an easy task, especially in cluttered urban environments [13] and using moving cameras. The shape of pedestrians is highly variable, affected by pose, clothing, handbags, backpacks, walking speed etc. It is also hard to segment out pedestrians from background, because of the moving camera, the fact that pedestrians may move in groups, parts of their body may be invisible due to occlusions and other parts may be spuriously added.

The high interest in human recognition systems, both for automotive and surveillance applications, can be seen from the large number of papers published in this domain.

An excellent survey on the analysis of human motion is [14]. In [15] the authors use far infrared cameras, hyper permutation networks, hierarchical contour matching and a cascaded classifier approach. In [9] a method using the “chamfer system”, texture classification and stereo verification is presented. [13] describes a top down segmentation approach which aggregates evidence in several stages in order to detect pedestrians in crowded scenes using a fixed camera. An approach related to our detector is presented in [16], which uses various 2–D motion patterns in order to detect pedestrians.

This section presents a novel approach for pedestrian detection, using multiple features. The features used are the vertical and horizontal dimensions of the pedestrian, the pedestrian’s speed, a motion signature, namely the variance of the motion field caused by the different parts of the pedestrian’s body and the periodicity of this motion signature. Assuming that these features are independent, a naive Bayesian framework is used to combine them. A learning tool has been developed for training the Bayesian classifier.

The architecture of our system is described in fig. 1. Our system uses 2 gray-scale cameras and a hardware stereovision machine named TYZX [17], which is capable of real-time dense stereo reconstruction. The TYZX system supplies a range image, associated with the left camera image. The next processing step is to group reconstructed points into 3–D objects, having a box shape. Object grouping is able to segment pedestrians, pedestrian groups and other types of objects reasonably well. In order to stabilize detection across multiple frames, detected coarse objects are tracked, using Kalman filters.

Next, various features are extracted, and are combined into a naive Bayesian classifier. A large set 1000 of manually labeled pedestrians and non-pedestrians objects were used to train the classifier. The classifier is automatically generated using a custom made tool.

In the following sub-sections we describe the features used by the classifier.
A. Simple features

A number of easy to compute features are used in our classifier. These features use only the information given by the grouping and tracking modules.

The objects are represented by their bounding 3-D box. We extract the simple features directly from this box. The simple features are: the object’s height, base radius, longitudinal speed (in the direction of the ego vehicle) and lateral speed (the direction perpendicular to the direction of the ego vehicle). Although these simple features are not powerful enough by themselves to detect pedestrians, adding them improves the final detection. Figure 2 shows some plots of these features’ relevance to classification. The objects having a box height of 1.5–2m and a box radius of about 0.5m positively influence the overall pedestrian probability, while other values have a negative influence. Also, the probability of pedestrians drops greatly with the increase of both lateral and longitudinal speed. Some small lateral speed seems to increase the probability of pedestrians, because most pedestrians in the training set are crossing the street in front of the ego vehicle. By testing our system with and without using these simple features, we came to the conclusion that, while the influence of these simple features on the overall detection is not very large, they prove useful in some occasions. As these simple features are already computed, using them does not increase the complexity of our algorithm.

B. Motion Signature

A powerful feature for walking pedestrian detection is the fact that, while walking, the pedestrian’s legs and arms swing back and forth. Therefore, the motion vectors of points located on the pedestrian’s body have a large variance, as opposed to the motion vectors of a rigid object (whose parts move rigidly). We call this variance in 3-D a motion signature.

Unfortunately, the motion signature is difficult to compute. This is usually due to substantially large dynamic range of the 2-D motion in the image plane caused by projection of the 3-D motion vectors corresponding to the motion of the objects in the scene and the motion of the ego vehicle.

To deal with the difficult problem of extracting a motion signature we use a hierarchical approach, starting with low level image features and gradually focusing to
pedestrian detection. The list of steps performed in order to detect pedestrians is (see Fig. 1):

a) Dense Stereo: We use a stereo camera configuration and a hardware dense stereo reconstruction system, “TYZX” [17] which provides us with the basic features: the gray-scale left and right images and a depth image, associated with the left gray-scale image.

b) Points Classification: The points for which the hardware “TYZX” is able to recover the correct depth are further classified into road points, object points and other points, based on various features. Only object points, i.e. points that could possibly belong to interesting objects in the scene (vehicles, pedestrians, road-side objects) are considered in the following steps.

c) Grouping: Object points are grouped into meaningful cuboid objects, based on their 3-D coordinates distribution. The result of this step is list of coarse objects, characterized by their associated coordinates and dimensions.

d) Tracking: Coarse objects are tracked across multiple frames using a Kalman filter-based multi-tracker. This step ensures stable detection and easier optical field computation (because of the inherent motion compensation resulting from the tracking). All tracked objects are considered as possible pedestrian candidates.

e) Depth Masks: Object masks are computed for all tracked objects. Only points for which their 3-D coordinates lie inside the tracked cuboid are considered. This step is important as it eliminates spurious points and deals with partial occlusions.

f) Optical Flow: A pyramidal, corner-based optical flow detection algorithm [18] is used to compute optical flow in all corner points belonging to tracked objects. Only optical flow vectors starting and ending on non-masked points are considered.

g) 3-D Velocity: The true 3-D velocity of the considered points is computed, using the 2-D optical flow, stereo depth and frame time-stamps.

h) PCA: Principal component analysis is used to find the principal direction of the 3-D velocity field variation for each individual object. Variance is smoothed across frames, to increase its stability.

Objects with a high value of the motion signature are most probably pedestrians, while objects with a low motion signature value are most probably rigid objects such as cars. Occasionally, because of errors in optical flow computation, some objects may generate a spurious motion field. These situations are dealt with in the next section.

C. Motion Signature Periodicity

Another powerful feature, related to the motion signature is its periodicity. Although non-pedestrian objects display occasionally some spike noise in their motion signature, this noise is not periodic.

Welch’s averaged modified periodogram method of spectral estimation [19] is used to compute the frequency spectrum. Frequency spectra corresponding to periodic pedestrian motion are typically band limited, while for other types of objects, the spectrum never gets close to zero. Figure 4 shows some examples of the motion signature variation and their spectra. The classifier analyzes the frequency spectrum of motion signature in order to identify the periodical motion of the pedestrian’s arms and legs. The cutoff frequency is used as feature.

D. The Bayesian Classifier

In order to combine the six features described in the previous section into a powerful classifier we used a Bayesian approach. Because computing the full joined probability of 6 features requires an immense data set, we made the simplifying assumption that the features we considered are independent. The resulting classifier is therefore a naive Bayesian classifier.

Our using of a Bayesian classifier is very flexible, as it permits adding any number of new features. It is also the optimal classifier (if the independence assumption is
made), and it generates better results than an empiric, rule-based system. The classifier is also very fast.

IV. EXPERIMENTAL RESULTS

Our system was tested in many different, crowded urban traffic scenarios, using cameras mounted on a moving road vehicle. The detection rate is high, and our system proves to be reliable. There are few false positives, and pedestrians are detected early enough to permit taking active measures for collision avoidance. The use of multiple features combined in a Bayesian framework greatly increases the precision of the pedestrian detector. A number of 600 test objects were manually labeled, and the classifier was tested on them. Table I shows the results. Figure 5 shows some correctly detected pedestrians while fig. 6 shows some correctly detected non-pedestrians. The system achieves real-time performance, because the features used are easy to compute.

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>600</td>
<td>100%</td>
</tr>
<tr>
<td>Pedestrians</td>
<td>243</td>
<td>40.5%</td>
</tr>
<tr>
<td>Other Objects</td>
<td>357</td>
<td>59.5%</td>
</tr>
<tr>
<td>False Positives</td>
<td>49</td>
<td>8.1%</td>
</tr>
<tr>
<td>False Negatives</td>
<td>39</td>
<td>6.5%</td>
</tr>
<tr>
<td>Correct Detection</td>
<td>512</td>
<td>85.3%</td>
</tr>
</tbody>
</table>

Fig. 4. Left: objects, middle: variation with time of the motion signature, right: motion signature frequency spectrum

Fig. 5. Correctly classified pedestrians (depth masked. a) shows a bicyclist classified as a pedestrian because of leg movement. d) and e) are examples of occluded, but correctly detected pedestrians. g) and h) show that even a pedestrian with a baby carriage is correctly detected

Fig. 6. Correctly classified non-pedestrians
V. CONCLUSIONS AND FUTURE WORK

We managed to develop a walking pedestrian detector, based on dense stereo and motion analysis. Our method has a high rate of correct detection and is less sensitive to incorrect grouping and partial occlusions than methods based solely on pedestrian shape. Because our classification is based on 3–D motion field, we can detect pedestrians walking in various directions and at various distances. Our system achieves a high rate of correct detection, above 85%. Because the optical flow computation is restricted to only a few points (a few hundreds), our detector is fast. It is able to run in real-time on a Intel Core 2 Duo processor at 2.4 GHz, achieving 20 frames per second.

There are a number of possible improvements for our detector:

1) Considering the occluded area of each object and normalizing the magnitude of optical flow variation with the size of non-occluded object area.
2) Optimizations of the corner detection, Gaussian pyramids generation and corner tracking, using SIMD instructions, in order to make the detector as fast as possible.
3) Using an adaptive number of Gaussian pyramid levels based on each object’s distance and velocity.
4) Perhaps the most obvious improvement will be using other features, like texture, shape etc. However, we need to find ways to improve the object grouping and tracking, to make these other feature usable.

REFERENCES


