Vision-Based Pedestrian Detection for Driving Assistance

Literature Survey

Multidimensional DSP Project, Spring 2005

Marco Perez

Abstract

This survey focuses on some of the most important and recent algorithms for pedestrian detection in video sequences taken from a camera mounted on a car. The long-term goal of this line of research is to develop vision-based systems that could be successfully implemented in cars for driver support in urban scenes or for autonomous driving. Some of the surveyed systems have already proved their effectiveness in on-road tests.

I. Introduction

More than 39,000 pedestrians are killed and 430,000 injured in traffic around the world every year. Pedestrian accidents represent the second largest source of traffic-related injuries and fatalities, just after accidents involving car passengers, see Table 1 [4].

	Killed	Injured	Total
Car passengers	75,615	3’751,024	3’826,639
Pedestrians	**39,670**	**436,422**	**476,092**
Bicycles	6,872	236,027	242,899
Mopeds	3,151	163,854	167,005
Motor Cycles	10,972	227,946	238,918
Other	28,397	1’303,571	1’331,968
Total	161,677	6’118,844	6’283,521

Table 1: Road Traffic Accidents 1997 – Figures for UN-ECE Countries (Accident Source: UN-ECE).

Despite improvements in pedestrian safeguards, many accidents still occur due to poor driving conditions (e.g. low light or fog) or a momentary distraction of either the driver or pedestrian. An automatic system to detect pedestrians in the surroundings of a vehicle is highly desirable and is today one of the main concerns of both private car manufacturers and public entities.

An open issue is which sensors are best suited for pedestrian detection in an urban environment. At present, radar or laser range finders and video sensors seem the most attractive solutions. Besides, there is a growing interest towards infrared vision systems, which have recently belonged only to military applications, but promise to be very appealing thanks to new cheaper technologies.

The use of video sensors comes quite naturally. Texture information at a fine angular resolution allows the use of discriminative pattern recognition techniques. The human visual system is probably the best example of what performance might be achieved with these sensors, if only the appropriate processing is added. Moreover, video cameras are cheap and because they do not emit signals, there are no issues regarding interference with the environment [5].

The vision-based pedestrian detection problem is challenging for a number of reasons. Pedestrians appear in highly cluttered backgrounds and have a wide range of appearances, due to different poses, body sizes, clothing and outdoor lighting conditions. They also may stand relatively far away from the camera and appear rather small in the images. Furthermore, it is not possible to use background subtraction methods to obtain foreground regions containing pedestrians because of the moving camera. There are also real-time requirements which rule out any brute-force approach [5].

II. Background

Various methods implemented for the detection of a candidate pedestrian and the subsequent recognition of a person among the collected candidates can be found in the literature. Motion detection, range thresholding and shape-based methods are the most common methods for the detection phase. Gait analysis and shape models are the most common methods for the recognition phase.

Motion is a cue for detecting interesting regions, possibly containing pedestrians, in a scene and it heavily relies on temporal information. It also needs the analysis of a sequence of a few frames and it does not detect standing pedestrians.

Range information is another powerful cue for foreground/background segmentation. Range-based segmentation is less affected by light conditions, shadows and occlusion. It is also less expensive computationally and suitable for real time implementation. But it may be affected by noise, especially at longer ranges [10].

The shape-based approach performs a direct search over the entire image based on a priori human shape information. This approach bypasses any problem related to non-stationary background and rarely integrates temporal information in the form of a sequence analysis. These algorithms must necessarily consider prior information about human shape or otherwise they may result in a very intensive computation.

Periodicity of the human gait is a strong cue for the recognition of walking pedestrians, especially when they walk laterally to the viewing direction. A number of frames must be analyzed before the system can actually give a response, and a reliable tracking of the pedestrian is needed. Methods implying gait recognition depend strongly on a good lateral view of a pedestrian. Stationary pedestrians are not detected by this approach.

Systems that implement shape models for pedestrian recognition do not need any temporal information and may perform validation of stationary pedestrian candidates. This approach is more sensible to false positives and need a good detection phase.

III. Pedestrian Detection Using Stereo Vision

Stereo is a technique that is commonly used to segment foreground objects from the background. Compared with intensity-based approaches, stereo-based segmentation is less affected by lighting conditions, shadows and occlusions. Fig. 1 displays the stereo geometry.

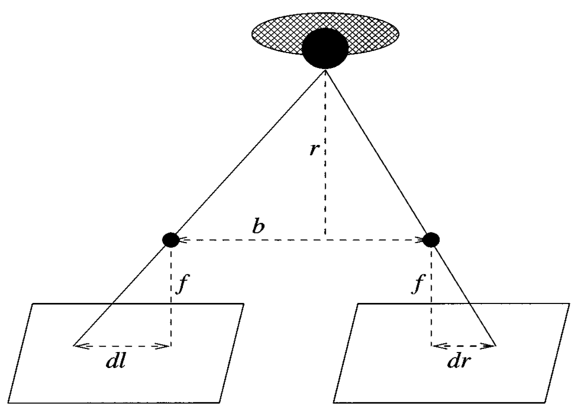

Figure 1. Stereo geometry.

Two images of the same object are taken from different viewpoints. The distance between the viewpoints is called the *baseline (b)*. The focal length of the lenses is *f*. The horizontal distance from the image center to the object image is *dl* for the left image, and *dr* for the right image. The stereo cameras are set up so that their image planes are embedded within the same plane. The difference between *dl* and *dr* is called the disparity.

In [10], background objects are eliminated from the disparity map by range thresholding. Then, a morphological closing operator is employed in order to remove the noise and to smooth the foreground regions. A connected-component grouping operator is applied to find the foreground regions with smoothly varying range. Finally, small regions are eliminated through size thresholding. Fig. 2 illustrates the segmentation results. Overlapping objects are successfully separated due to their different distances from the camera. After the segmentation step, a three-layer feed forward network trained by the back-propagation algorithm is used for pedestrian recognition.

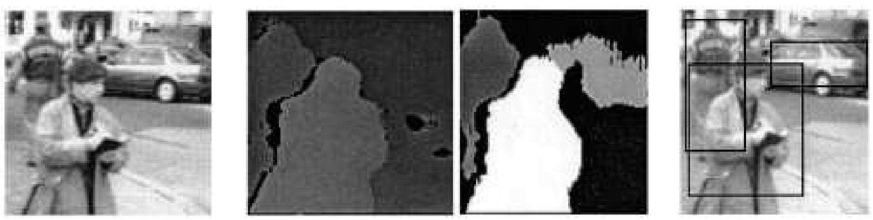

Figure 2. Left image from stereo cameras, disparity map, segmentation result and detected objects.

In [5], a two-step approach for object detection is used. In the first step, contour features are used in a hierarchical template matching approach to find candidate solutions. Shape matching is based on Distance Transforms. This method achieves very large speed-ups compared to brute force methods. The second step utilizes the set of intensity features in a pattern classification approach to verify the candidate solutions. See Figs. 3 and 4.

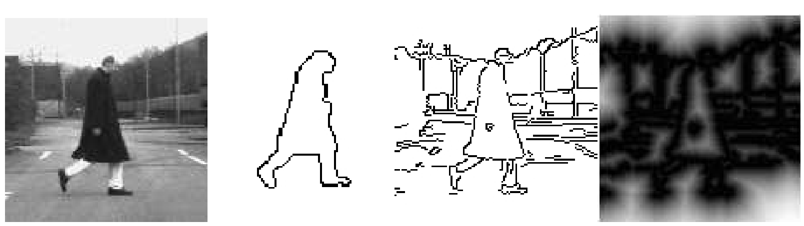

Figure 3. Original image, template, edge image and DT image.

Figure 4. Pedestrian detection results.

In [7], a method for pedestrian detection is presented, using a stereo night vision system. This method detects moving objects that have motions not consistent with the movement of the background. This technique makes use of characteristics of night vision video data, in which humans appear as hotspots. The method works well in cases where the camera motion has a dominant translational motion with a small amount of rotational motion, which is suitable for the camera on the vehicle.

IV. Shape-Based Pedestrian Detection

Shape-based approaches try to solve the difficult problem of recognizing pedestrians in single images, taking care of moving and stationary pedestrians. The biggest challenge that this problem offers is to model the huge amount of variations in shapes, poses, sizes and appearances of humans.

In [1], the knowledge of the vision system's extrinsic parameters and the flat scene assumption allows to reduce the search for candidates to one limited part of the image. After this low level processing, an analysis of vertical symmetry maps derived from gray-level and horizontal gradient image values is performed. Then, an identification of regions that can be characterized as human shapes takes place. This stage produces boxes with a high probability to fit a pedestrian. Finally, a number of filters eliminate some of the candidates that do not actually represent a human shape, see Fig. 5.

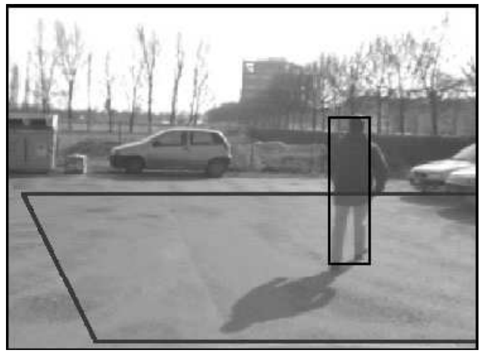

Figure 5. Algorithm result for an example outdoor image acquired from a moving vehicle.

In [8], a real-time pedestrian detection system that works on infrared videos is presented. First, a pixel-based representation is used. The raw intensity values at each pixel are used to classify target objects as pedestrians or non-pedestrians. The target objects are extracted from the raw video by using simple intensity thresholding. The pixels that belong to the background are given value 0 and the pixels that correspond to objects emitting heat are replaced by 1's. A training data set is used for developing a probabilistic template which consists of 1000 128x48 rectangular images all of which contain humans. Each template is translated so that the centroid of the non-zero pixels matches the geometrical center of the image. For each pixel of the template, the probability $p(x,y)$ of it being pedestrian is calculated based on how frequently it appears as 1 in the training data. The template gives the probability of seeing a foreground at different pixel locations for pedestrians. For each pixel, the probability that the pixel has the correct classification given that the window contains a pedestrian is $p(x,y)$ if the pixel has a value 1 or $1-p(x,y)$ if the pixel has the value 0. The sum of these probabilities is calculated for all pixels and this gives the combined probability of the given window containing the person given the prior, see Fig. 6.

Figure 6. Probabilistic template and output frames (pedestrian heads are marked by blue contours).

V. Pedestrian Detection Using Periodicity of the Human Movement

This line of work has dealt specifically with scenes involving people walking laterally to the viewing direction. Periodicity of the human gait is a strong cue for the recognition of walking pedestrians, either derived from optical flow [9] or raw pixel data [2].

In [6], the recognition is based on the characteristic motion of the legs of a pedestrian walking parallel to the image plane. Each image is segmented into region-like image parts by clustering pixels in a combined color-position feature space. The clustering technique matches corresponding clusters in consecutive frames allowing clusters to be tracked over a sequence of frames. Based on the observation of clusters over time a classifier extracts those clusters which most likely represent the legs of pedestrians. A polynomial classifier performs a rough pre-selection of clusters by evaluating temporal changes of a shape-dependent cluster feature. The final classification is done by a time delay neural network with spatio-temporal receptive fields.

Figure 7. Image sequences showing pedestrians. The clusters recognized as pedestrians are marked by bounding boxes.

VI. Conclusion

Vision based pedestrian detection is still an open challenge. People blend with the background, walk, stand or change directions unpredictably. The background is complex and sudden changes are unavoidable in vision systems mounted on a moving vehicle. Different approaches have been developed to address these and other complexities. Some systems have already proved their effectiveness in on-road tests, and this encourages further research. Future work involves foreground/background segmentation of regions likely to contain pedestrians, their accurate detection and tracking in color video sequences taken from a camera mounted on a car.

References

[1] M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi, R. Chapuis and F. Chausse, "Pedestrian localization and tracking system with Kalman filtering", in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 584-589, June 2004.

[2] R. Cutler and L. Davis, "Real-time periodic motion detection, analysis and applications", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 326-331, vol. 2, June 1999.

[3] H. Elzein, S. Lakshmanan and P. Watta, "A motion and shape-based pedestrian detection algorithm", in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 500-504, June 2003.

[4] D. Gavrila, "Sensor-based pedestrian protection", in *IEEE Intelligent Systems*, vol. 16, no. 6, pp. 77-81, November 2001.

[5] D. Gavrila, J. Giebel and S. Munder, "Vision-based pedestrian detection: the PROTECTOR system", in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 13-18, June 2004.

[6] B. Heisele and C. Wohler, "Motion-based recognition of pedestrians", in *Proc. International Conference on Pattern Recognition*, pp. 1325-1330, vol. 2, August 1998.

[7] X. Liu and K. Fujimura, "Pedestrian detection using stereo night vision", in *IEEE Transactions on Vehicular Technology*, pp. 1657-1665, vol. 53, no. 6, November 2004.

[8] H. Nanda and L. Davis, "Probabilistic template based pedestrian detection in infrared videos", in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 15-20, vol. 1, June 2002.

[9] R. Polana and R. Nelson, "Low level recognition of human motion", in *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 77-82, November 1994.

[10] L. Zhao and C. Thorpe, "Stereo- and neural network-based pedestrian detection", in *IEEE Transactions on Intelligent Transportation Systems*, pp. 148-154, vol. 1, no. 3, September 2000.