

Real Time Stereo Vision Based Pedestrian Detection Using Full Body Contours

Ion Giosan, Sergiu Nedevschi, Silviu Bota
Technical University of Cluj-Napoca
{*Ion.Giosan, Sergiu.Nedevschi, Silviu.Bota*}@cs.utcluj.ro

Abstract

There are many approaches to pedestrian detection in collision avoidance systems depending on the sensors (visible light, thermal infrared, RADAR, LASER scanner) used for acquiring the data and the features (depth, shape, motion) used for detection. In this paper we present a method for shape based pedestrian detection in traffic scenes using a stereo vision system for acquiring the image frames and a contour matching technique for classifying the scene objects as belonging or not to the pedestrian class. The 3D information is used for determining the foreground points of each 2D object and a contour extraction and merging algorithm is performed on these points. A 2D image filtering is performed and edges are extracted and used for objects contours refinement. A hierarchy of pedestrian full body contours and a matching technique are used for classifying the extracted objects contours from the scene.

1. Introduction

In the modern society the design and building of intelligent vehicles is a challenge for both to the researchers and vehicle constructors. Pedestrians are the most vulnerable participants involved in traffic. The objective is first to reliably detect pedestrians and after that to protect them. This protection is done by simply warning the driver or, in case of an imminent impact, by opening for example a special frontal part of the vehicle before hitting the pedestrian. The detection rate should be sufficiently high in order not to cause false alarms to the driver, falsely trigger the special protection parts of the vehicle.

Different technologies [1] such as ultrasound sensors, piezo-electrical sensors, laser scanners [2], microwave radars and video cameras [3] are frequently used for pedestrian detection. The detection based on processing video images acquired using video cameras is the most similar to the human vision system. Using

video cameras, the the field of view is wide and more scene information can be extracted as compared to other sensors which are limited in this sense. The detection in this manner is a passive and clean one because it does not imply any source of pollution and does not affect the people.

The pedestrian detection is a simple problem for humans but it is a complex one for artificial vision systems due to the large variations of human body poses, clothing, the accessories which they are carrying and of course due to the variations of background scene, the environment conditions, the distance from video cameras and their resolution, the unpredictable vehicle vibrations and scene cluttering. The most important aspect is that the pedestrian detection should be done in real time and this makes the detection a complex process which involves multiple efficient methods.

1.1. Related Work

The contour is a strong feature for pedestrian detection [4] motivated by the fact that it eliminates most of the problems previously presented which could cause weak detection (false positives and false negatives). The contour describes the shape of the pedestrian and it is invariant to other factors such as clothing and illumination. Scene objects are usually matched to a set of ideal pedestrian contour templates to determine if they belong to the pedestrian class. An approach for object detection using a hierarchy of contour templates and pattern matching with distance transform on edges is described in [5]. In [6] a system for pedestrian detection based on edges and shapes and using monocular vision is presented.

The main idea in pattern matching is to have a database with many pedestrian contours with different attitudes and poses. The database should contain a set of perfect pedestrian body contours obtained from ideal pedestrian images. A remark is that it is inefficient to perform an exhaustive search and match process through the whole template database as it would in-

volve performing a large number computationally complex matches between the pedestrian hypotheses and the templates. Such an approach would sacrifice the real time performance of the system. A hierarchy of pedestrian contours will decrease the matching time by reducing the search space. Some pedestrian detection systems using templates are described in [7], [8] and [9].

A pedestrian detection application based on local multi-scale oriented intensity differences, that uses the Haar wavelet transform and SVM for classification is presented in [10]. In [3] a complete and robust system that performs a stereo-based depth segmentation, a chamfer matching for shape, texture classification for verification using neural network, stereo-based verification, and tracking is presented.

Feature extraction and tracking based on structural changes of shape, especially the legs and symmetry detection method using morphological operators are used for pedestrian detection as presented in [11]. Active contour models can be used for pedestrian segmentation and stereo for guiding the active contour location since they are very sensitive to initial position [4].

Pedestrian detection and classification based on pattern matching are limited to 2D image intensity information [7]. Other techniques such as Adaboost feature classifiers are also implemented for improving the pedestrian detection [12]. Features based on 2D information are used for the detection of image regions where there are a significant number of vertical edges [13].

1.2. Vision System Characteristics

Pedestrian collision avoidance systems are classified by their field of view, angular resolution, detection range, range resolution, illumination type, hardware cost and algorithmic complexity [1].

Our stereo vision system has wide field of view, of 68 degrees, the angular resolution is medium, at about 8 minutes of arc and the detection range is medium, at 20m, as it becomes very difficult to detect pedestrians beyond this range, because of the wide field of view we use. The range resolution is high, thanks to the use of 3D information (the depth information is computed from stereo rather than inferred from mono) and a high precision calibration procedure [14]. We use no active illumination techniques which is an advantage. The hardware cost is medium as normal gray level cameras are used. A hardware stereo vision machine [15] is used for stereo correspondence and depth information extraction. The algorithmic complexity is medium to low, as using 3D information speeds up the matching process. The stereo system is mounted inside a vehicle.

1.3. Contributions

The novel aspects in our pedestrian detection system consist in using both 3D and 2D information from images acquired from the stereo vision camera system.

The pedestrian candidate objects are represented by their 3D surrounding box. We filter the pedestrian hypotheses based on their size and aspect ratio. An object is a pedestrian hypothesis only if it has a specific size (height between 1m and 2m and width between 0.25m and 1m) and an aspect ratio (height/width) between 4.00 and 1.00.

A novel aspect of our contour extraction algorithm is that it is performed on 2D foreground points obtained by projecting on the left image the 3D points located inside the pedestrian candidate's 3D box. Another contribution is that a single contour is obtained for each object by merging multiple incomplete contours. In order to improve the contour extraction we use intensity image edges filtered by length and continuity criteria. We use non-linear edge preserving smoothing pre-filters for obtaining better intensity image edges.

Another novel aspect is that we match templates contours with objects contours instead of matching them directly with edge points extracted from the intensity image. We compute the matching score as the average of a direct match (object contour to template contour) and an inverse match (template contour to object contour). In order to reduce false matchings we use full body contours.

1.4. Paper Structure

In section 2 we present the pedestrian detection using full body contours including the detection system architecture, the objects contour extraction and refinement, the hierarchy of full body pedestrian contours and the contour matching approach. In section 3 the experimental results regarding the pedestrian detection are shown. Finally in section 4 the conclusions are drawn and possible future improvements are proposed.

2. Pedestrian Detection Using Full Body Contours

This section describes our methods, algorithms and techniques used for pedestrian detection. The full body pedestrian contour is used as a strong shape feature in the contour matching process. Using this feature for the pedestrian classification provides a complete description of the pedestrian shape and leads to better results (smaller number of false detections and non-detections). The improved results will be presented in

the experimental results section.

2.1. Detection System Architecture

Figure 1 presents the processes involved and the data flow of the pedestrian detection system. Starting from gray level left and right intensity images (512x383 pixels) of the scene taken from the stereo vision system and using a hardware machine for stereo reconstruction a the 3D range image (depth image) is generated.

An object detection algorithm is applied on the 3D range image which finds the 3D boxes of all objects in the scene image (see figure 2). All the background points are eliminated by using depth information, the only remaining 3D points belong to the objects.

We compute the labeled image containing only the foreground points of the objects (see figure 3). A contour extraction algorithm is performed on these points resulting the objects contours. For each object we have a single continuous connected contour.

An optimized edge preserving non local means (NL-means) filter [16] is applied on the left intensity image. After filtering the image a 2D Canny edge detector is used for obtaining the filtered 2D edges. These edges are then involved in the refinement process of the previously extracted contour. The result is the refined 2D objects contour points.

The final step is to perform a full body contour matching to determine if that object is or not a pedestrian. The matching is realized between the 2D object contour and a hierarchy (multi-path tree) of full body pedestrian contour templates obtaining the detection result. These methods are detailed in the next subsections.

2.2. Objects Contour Extraction and Refinement

In the labeled objects foreground points image each object can contain several parts (connected components). These objects parts are the input of the contour extraction module. The output of the contour extraction module is a continuous chain of contour pixels which best approximate the object (see figure 4) [17].

The scene objects contour refinements were necessary to improve the pedestrian contour matching correctness. The problems that lead to an imperfect contour extraction (larger than the real object area or flowing into the real object area) are the faulty 3D reconstructed points and the lack of 3D reconstructed points. Because of these problems the contour isn't perfectly fitted on the object boundary. The solution is to try to use 2D intensity image information in addition 3D information. Our approach is to find the object's exterior

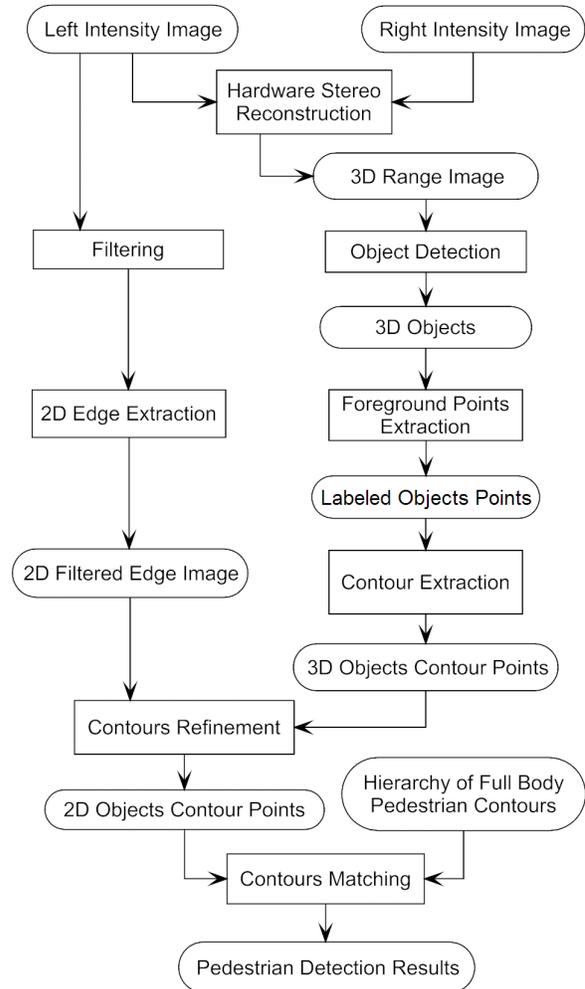


Figure 1. Processes and data flow in the pedestrian detection system

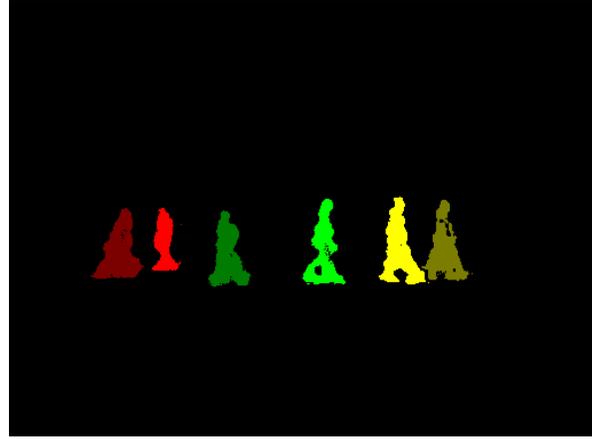
edges and to refine the contour using these edges. A problem with this approach is the lack of edges when it becomes impossible to decide where the the real boundary of the object is. This method removes the inner-object edges and keeps only the exterior edges.

A preprocessing step was necessary for filtering the intensity image to improve the extracted contours. The method for filtering is NL Means. The algorithm computes the similarity between the neighborhood of each pixel and all the pixels of the objects areas from the image, performs noise reduction and preserves the original image. This filtering will remove the noisy edges. Important edges will be enhanced and smoothed.

A comparison between the edges on an object from the original image and those from the filtered image is shown in figure 5.



(a) Scene image



(b) Labeled objects foreground points image

Figure 3. Image points corresponding to the objects from the scene

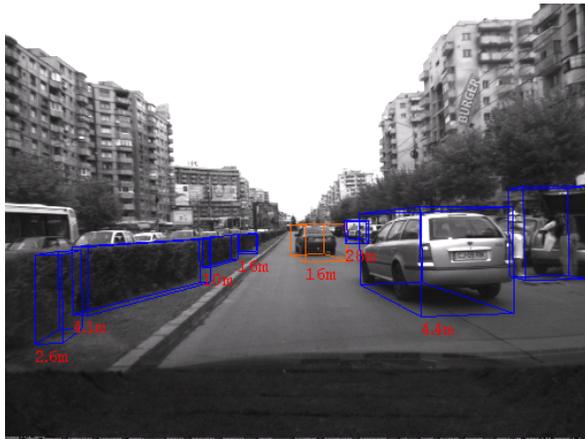


Figure 2. Example of 3D oriented boxes of the objects



Figure 4. Example of the extracted contours of the scene objects

2.3. Hierarchy of Full Body Pedestrian Contours

The pedestrian templates hierarchy is a multi-path tree with five levels height. The number of templates from each group is not constant. There can be groups with different number of templates, the minimum being two models in a group. Only the root of the tree is a group with one template. The contour matching algorithm is started from root down to leafs, storing the best score and choosing everytime the branch having the minimum score of matching.

Pedestrian templates were obtained by using the background subtraction method for determining the pedestrian area, manually adjusting the small imperfections of the obtained area, and finally applying the con-

tour extraction algorithm on that refined area. The hierarchy contains 1872 models (936 original models and another 936 obtained by flipping the original ones over the vertical axis – opposite direction of walking) of full body pedestrian contours having various attitudes. .

In the hierarchy building process these models were scaled to a fixed size having an aspect ratio considered suitable for pedestrians. The height and width of the scaled templates were set to 150 respectively 71 pixels resulting an aspect ratio of about 2.1. These values were obtained by initially averaging all the heights and widths of the original pedestrians template contours. The figure 6 shows some models of the scaled pedestrian contours contained in the hierarchy. The pedestrian contours hierarchy is used in real-time contour matching algorithm for obtaining the matching score



Figure 5. An original and filtered object (left) and corresponding extracted edges (right)

which determines if a pedestrian hypothesis is truly a pedestrian.

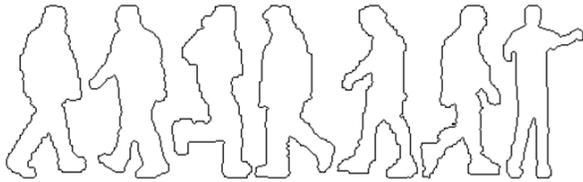


Figure 6. Examples of pedestrian templates used in the hierarchy

2.4. The Matching Contours Approach

The contour matching is performed between the contour of the pedestrian hypothesis from the scene image and the hierarchy of the pedestrian contour templates. There are two phases in the matching process.

First phase consists in determining the contour-to-template matching score (see figure 7). In this phase the contour of the hypothesis is superimposed on the template contour of the pedestrian in the hierarchy. Then a Distance Transform (DT) is applied on the template contour. The matching score consists in the average sums of the pixels intensities (distances) in the template

contour DT image that lie below the contour hypothesis points.

The second phase is the reverse of the first phase (see figure 7). It consists in determining the template-to-contour matching score. In this phase the template is superimposed on the hypothesis contour of the pedestrian. A DT is applied on the hypothesis contour. The matching score consists in the average sums of the pixels intensities (distances) in the pedestrian hypothesis contour DT image that lie below the template contour points.

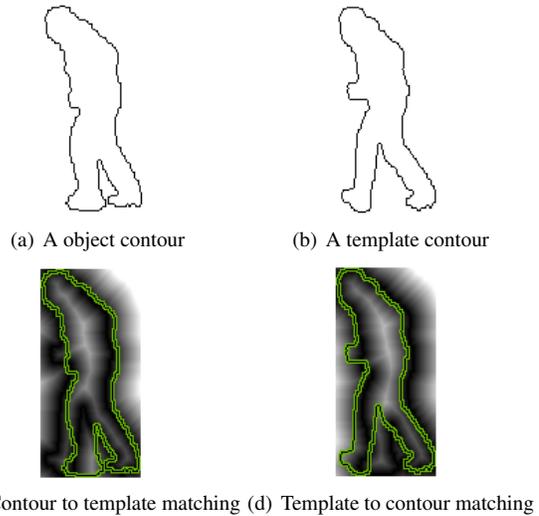


Figure 7. DT based matching

In the matching process the contour hypothesis of the pedestrian and the template contour from hierarchy are scaled to the same size for calculating the matching score. An object is considered a pedestrian hypothesis only if it has a specific size (height between 1m and 2m and width between 0.25m and 1m) and an aspect ratio (height/width) between 4.00 and 1.00. The matching score is computed as being the average of contour-to-template and template-to-contour matching scores. A hypothesis is considered to be a pedestrian if the matching score is below a threshold.

3. Experimental Results

In this section we present the pedestrian detection results obtained using the full body contour extraction and matching with a full body pedestrian contour templates hierarchy.

The proposed methods and algorithms for pedestrian detection were tested on sequences with thousands of images from different environments. The ROC curve of our detector is presented in figure 8. In table 1 we

present the pedestrian detection results obtained on a sequence taken from a medium-high complexity environment. Of course there is a compromise between the rate of non-detected pedestrians (false negatives) versus wrong detected pedestrians (false positives) that can be adjusted depending on the desired application by using asymmetrical cost functions.

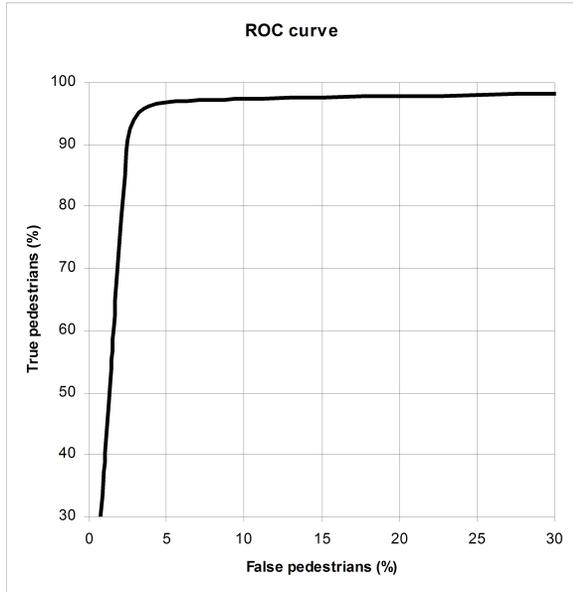


Figure 8. Pedestrian detector ROC curve

| Results | | |
|------------------------------|------|-------|
| Total objects | 3195 | 100% |
| Pedestrians | 2703 | 84.6% |
| Other objects | 492 | 15.4% |
| Correct pedestrian detection | 2963 | 92.7% |
| Non-detected pedestrians | 130 | 4.1% |
| Wrong detected pedestrians | 102 | 3.2% |

Table 1. Pedestrian detection results

Some frames from a test sequence and the pedestrian detection results are presented in the figure 9. The extracted objects contours are drawn in green color. Pedestrian full body contour templates from hierarchy that matched successfully over the objects contours are drawn in yellow color. Those objects are detected as being pedestrians.

The pedestrian detection using this technique is performing in real-time mode (it achieves 25 fps when running on a computer system having an Intel Core 2 Duo processor at 2.66 GHz).

4. Conclusions and Future Work

We have presented a method for shape based pedestrian detection in traffic scenes using a stereo vision based approach and a contour matching technique for detecting the pedestrians from the scene. The 3D information is used for depth masking points of each object and a contour extraction algorithm is applied on the 2D projections of these masked points. A image filtering is performed and the intensity image edge information is used in objects contour refinement. A hierarchy of pedestrian full body contours and a matching technique are used for classifying the extracted objects contours.

Pedestrian detection by full body contour matching can be improved by extracting better objects contours. Using a higher resolution intensity image is important in the process of contour refinement. The idea is to search in the neighborhood of the extracted contour and to consider the edges in that area. The length and orientation of the edges can be analyzed and finally the contour traced close to these edges (within the complexity constraints of real time processing performance).

The lack of edges (see figure 10) on the true object boundary makes the extraction of perfect objects contours difficult. An incomplete extracted contour gives suboptimal matching results. The solution is to use at least one color camera and a higher resolution for enhancing the edges that do not appear in gray images.



Figure 10. Lack of edges in gray level images



Figure 9. Objects classified as pedestrians (full body contour based matching)

References

- [1] T. Gandhi and M. M. Trivedi, "Pedestrian collision avoidance systems: A survey of computer vision based recent studies," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, 2006, pp. 976–981.
- [2] B. Fardi, U. Schuenert, and G. Wanielik, "Shape and motion-based pedestrian detection in infrared images: a multi sensor approach," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, June 6-8 2005, pp. 18–23.
- [3] D. M. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: the protector system," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, June 14-17 2004, pp. 13–18.
- [4] C. Hilario, J. M. Collado, J. M. Armingol, and A. de la Escalera, "Pedestrian detection for intelligent vehicles based on active contour models and stereo vision," in *Proceedings of the IEEE Intelligent Vehicular Sytems*, October 19 2005, pp. 537–542.
- [5] D. M. Gavrila and V. Philomin, "Real-time object detection for "smart" vehicles," in *Proceedings of the IEEE International Conference on Computer Vision*, 1999, pp. 87–93.
- [6] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, October 3-5 2000, pp. 215–220.
- [7] D. M. Gavrila, "Pedestrian detection from a moving vehicle," in *Proceedings of the European Conference on Computer Vision*, January 01 2000, pp. 37–49.
- [8] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," in *International Journal of Computer Vision*, 2007, pp. 41–59.
- [9] S. Nedeveschi, C. Tomiuc, and S. Bota, "Stereo based pedestrian detection for collision avoidance applications," in *Proceedings of ICRA Workshop: Planning, Perception and Navigation for Intelligent Vehicles*, 2007, pp. 39–44.
- [10] C. Papageorgiou and T. Poggio, "A trainable system for object detection," in *International Journal of Computer Vision* 38(1), 2000, pp. 15–33.
- [11] L. Havasi, Z. Szlavik, and T. Sziranyi, "Pedestrian dection

- tection using derived third-order symmetry of legs,” in *Proceedings of the IEEE International Conference on Computer Vision and Graphics*, 2004, pp. 11–17.
- [12] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laugreau, “Vehicle detection combining gradient analysis and adaboost classification,” in *Proceedings of the IEEE Intelligent Transportation Systems*, 2005, pp. 66–71.
- [13] M. Bertozzi, E. Binelli, A. Broggi, and M. D. Rose, “Stereo vision based approaches for pedestrian detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, p. 16.
- [14] S. Nedeveschi, T. Marita, R. Danescu, F. Oniga, and C. Pocol, “Camera calibration method for high-accuracy stereovision,” in *Proceedings of the Joint-Hungarian-Austrian Conference on Image Processing and Pattern Recognition, Vespem, Hungary*, 2005, pp. 303–310.
- [15] J. I. Woodfill, G. Gordon, and R. Buck, “Tyzx deepsea high speed stereo vision system,” in *Proceedings of the IEEE Computer Society Workshop on Real Time 3D Sensors and Their Use, Conference on Computer Vision and Pattern Recognition*, 2004, p. 41.
- [16] A. Buades, B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one,” in *Multiscale Model Simulation*, Vol.4, No. 2, 2005, pp. 490–530.
- [17] I. Giosan and S. Nedeveschi, “Building pedestrian contour hierarchies for improving detection in traffic scenes,” *Computer Vision and Graphics (Lecture Notes in Computer Science)*, Springer Berlin / Heidelberg, pp. 154–163, 2009.