Stereovision Sensor for Driving Assistance

Sergiu Nedevschi, Tiberiu Marita, Radu Danescu, Florin Oniga, Ciprian Pocol, Stefan Sobol, Corneliu Tomiuc, Cristian Vancea, Silviu Bota

Technical University of Cluj-Napoca, Department of Computer Science
E-mail: Sergiu.Nedevschi@cs.utcluj.ro

Abstract

This paper presents a high accuracy stereovision sensor for 3D lane and obstacle detection in traffic environments. Stereovision allows the elimination of the common assumptions used in most monocular systems: flat road, constant pitch angle or absence of roll angle. The accuracy of the 3D reconstruction is comparable with that of active sensors as radar, laser scanner or LADAR, while the quality of the detected information in terms of volume and meaning is much higher. However the stereovision sensor output can be used in a sensor-fusion system in conjunction with other sensors in order to obtain a more robust and complete description of the traffic environment. Possible applications of the developed stereovision sensor are the implementation of some driving assistance functions as lane keeping and lane changing assistance, frontal collision avoidance, pedestrian collision avoidance, stop and go, intersections assistance, ACC (Automatic Cruise Control) for highway and urban scenarios.

1. Introduction

From highway to urban traffic scenarios, the aim of a driving assistance system is to improve the safety of driving, and also to relieve the driver from repetitive and annoying tasks. Having a good description of the traffic environment is essential for any driving assistance system. There are two main classes of objects describing the driving environment: the lane (driving corridor) and obstacles.

The most common approach to obstacle detection is using active sensors such as lasers, ladars, or millimeter-wave radars. Their main advantage is that they can measure certain quantities (e.g., distance) directly requiring limited computing resources. But active sensors have also some drawbacks, such as low spatial resolution, and slow scanning speed. Moreover, when a large number of vehicles are moving simultaneously in the same direction, interference among sensors of the same type can occur [1].

Optical sensors, such as normal cameras, are usually referred to as passive sensors because they acquire data in a non-intrusive way. They are low cost sensors but do not perform direct measurements as the active ones. The measurement is done indirectly from the 2D image features and this process could be time consuming and its accuracy depends on the vision system setup. However, the use of a high resolution, high accuracy stereovision algorithm provides comparable results in 3D estimation, while delivering a larger amount of data [2].

Obstacle detection through image processing has followed two main trends: single-camera based detection and two (or more) camera based detection (stereovision based detection). The monocular approaches are using techniques such as object model fitting [3], color or texture segmentation [4], [5], symmetry axes [6] etc. The estimation of 3D characteristics is usually performed through a combination of knowledge about the objects (such as size), assumptions about the characteristics of the road (such as flat road assumption) and knowledge about the camera parameters available through calibration. The stereovision-based approaches have the advantage of directly measuring the 3D coordinates of an image feature [7], [8]. The main constraints concerning stereovision applications are to minimize the calibration and stereo-matching errors in order to increase the measurements accuracy and to reduce the complexity of stereo-correlation process for real time capabilities.

Lane detection has been for quite a long time the monopoly of the monocular image processing

techniques. Monochrome images tend to be preferred over color, due to better resolution and reduced data load. Lane detection methods become in this case a problem of fitting a certain 2D model to the image data. The 2D model can be a projection of a 3D road model (a clothoid curve is the most commonly used 3D model) in the image plane [9], usually the projection of a flat road of constant curvature, or it can be any mathematical model which can be matched under some robustness constraints, such as splines or snakes [10]. Some methods try to transform the 2D image into a surrogate 3D space by using inverse perspective mappings, under certain assumptions [11]. Sometimes the model itself is not explicitly stated, but it is implied by some assumptions that the algorithm takes (for instance, parallel lines). All the monocular lane detection methods suffer from their connection to a specific assumption (flat road, constant curvature, etc

In this paper we present a method for a full edge-based 3D reconstruction of the driving environment of a moving vehicle using stereovision. The stereovision algorithm allows the elimination of the assumptions of flat road, constant pitch angle or absence of roll angle. The lane is modeled as a 3D surface, defined by the vertical and horizontal clothoid curves, the lane width and the roll angle. The detection of the vertical profile is based on stereovision [2]. The horizontal profile detection is performed in the image space by a classical technique, but an ingenious method of switching between the 3D space and the 2D space is introduced [12], so that 3D lane parameters can be extracted from a single frame with good accuracy.

The availability of 3D information allows the separation between the road and the obstacle features. The list of obtained 3D points above the detected 3D road surface is grouped into objects based solely on density and vicinity criteria. To overcome the sparseness of the 3D points with the distance a compressed 3D space (with the depth) was introduced. In this way, the system detects obstacles of all types, outputting them as a list of cuboids having 3D positions and sizes, without having to make any assumption about their type [13]. Subsequent classification techniques can be employed for discrimination, if needed. The detected objects are then tracked using a multiple object tracking algorithm, which refines the grouping and positioning, and detects the speed.

2. Sensor and environment model

The stereovision sensor consists in two cameras mounted on a rigid rig. The position and orientation of the cameras' are completely determined by the

translation vectors $\mathbf{T_L}$ and $\mathbf{T_R}$, and the rotation matrices $\mathbf{R_L}$ and $\mathbf{R_R}$. The ego-car coordinate system has its origin on the ground in the front of the car, and its Z axis points in our direction of travel (Figure 1).
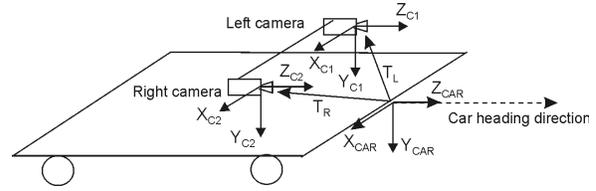

Figure 1. The stereovision sensor and the car coordinate systems

All 3D entities (points, objects) are expressed in the ego-car coordinate system, which is depicted in Figure 1. The detected objects are represented as cuboids, having position, size and velocity, as in Figure 2. The position (X, Y, Z) and velocity ($\mathbf{v_X}$ and $\mathbf{v_Z}$) are expressed for the central lower point C of the object's back face.

The lane is modeled as a 3D surface, defined by the vertical and horizontal clothoid curves (Figure 2). Lane detection is regarded as the continuous estimation of the following parameters [14]:

- W – the width of the lane
- $c_{h,0}$ – horizontal curvature of the lane
- $c_{h,1}$ – variation of the horizontal curvature of the lane
- $c_{v,0}$ – vertical curvature of the lane
- X_{cw} – the lateral displacement of the ego-car coordinate system from the lane reference system (lane center)
- α, γ, ψ are the pitch, roll and yaw angles of the car (the rotation angles between the car reference system and the world reference system).
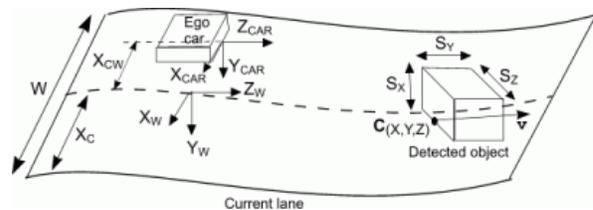
Figure 2. The environment model

These parameters describe the lane position and geometry through the following equations:

$$X_c = -X_{cw} - \psi Z + c_{0,h}\frac{Z^2}{2} + c_{1,h}\frac{Z^3}{6} \qquad (1)$$

$$X_L = X_c - \frac{W}{2} \qquad (2)$$

$$X_R = X_c + \frac{W}{2} \qquad (3)$$

$$Y = Z\alpha + c_{0,v}\frac{Z^2}{2} + \gamma X \qquad (4)$$

Equation (1) describes the horizontal profile - the variation of the lateral position (X) of the center of the lane with the distance Z. Equations (2) and (3) are expressing the lateral positions of the lane borders. Equation (4) describes the vertical position for any point on the road. The first two terms compose what we'll call the vertical profile, while the last term is due to the roll angle. All coordinates are given with respect to the car coordinate system, which is placed on the ground in the front center point of the ego vehicle.

Additional information is used from standard sensors of the ego-car: ego-car speed, yaw rate or steering wheel angle.

3. Environment perception

3.1. Camera calibration and 3D reconstruction

In order to reconstruct and measure the 3D environment, the cameras must be calibrated. The calibration process estimates the camera's intrinsic parameters (which are related to its internal optical and geometrical characteristics) and extrinsic ones (which are related to the 3D position and orientation of the camera relative the car coordinate system – Fig. 1). The intrinsic parameters are the focal length and the principal point coordinates and are estimated using the Bouguet algorithm [15] by minimizing the projection errors from multiple views of a set of control points placed on a coplanar calibration object with known geometry.

The extrinsic parameters (translation vectors $\mathbf{T_L}$ and $\mathbf{T_R}$ and the rotation matrices $\mathbf{R_L}$ and $\mathbf{R_R}$ - figure 1) are determined during an off-line calibration procedure and are remaining unchanged during exploitation. The calibration process minimizing the projection errors of a set of control points placed in a calibration field with sizes comparable with the detection range [16]. The extrinsic parameters calibration method assures not only very accurate absolute extrinsic parameters of each camera individually but also very accurate relative parameters of the cameras' inside the stereo-rig, which

allows a precise estimation of the epipolar lines (near 0 pixel drift).

The stereo reconstruction algorithm used is mainly based on the classical stereovision principles available in the existing literature [17]. Constraints, concerning real-time response of the system and high confidence of the reconstructed points, were supplementary used [2],[13]:

- Exclusion from the correlation process of non-structured areas as road granule-like textures and off-road vegetation texture, which are hard to correlate and time consuming for disambiguation.
- Use of only edge features to reduce the search space in the correlation process.
- Use of area based correlation by searching along the epipolar lines computed from the general stereo-geometry using parallel processing features of the processor.
- Use of disambiguation techniques to obtain a low rate of false correlations.
- Sub-pixel disparity computation by fitting a parabola to the correlation function.

After finding correspondences, each left-right pair of edge points is mapped into a unique 3D point. Two 3D projection rays are traced, using the camera geometry, one for each point of the pair. The 3D point can be computed as the intersection of the two projection rays. Due to calibration parameters uncertainties and discrete image nature, the rays will not intersect most of the time. The 3D point is computed as the point of minimum distance from both rays [17] in least squares fashion.

3.2. Lane detection

Lane detection is integrated into a tracking process. The current lane parameters (described in chapter 2) are predicted using the past parameters and the vehicle dynamics, and this prediction provides search regions for the current detection (Figure 3).

Figure 3. Lane prediction in the image space

The detection starts with estimation of the vertical profile (pitch angle and vertical curvature), using the stereo-provided 3D information. The side view of the set of 3D points is taken into consideration (Figure 4). From all the 3D points only the ones that project inside the predicted search regions (Figure 3) are processed. The pitch angle is extracted using a method similar to the Hough transform applied on the lateral projection of the 3D points in the near range of 0-20 m (in which we approximate the road flat) [12]. After detecting the pitch angle, detection of the vertical curvature follows the same pattern. The pitch angle is considered known, and then a curvature histogram is built, for each possible curvature, but this time only the more distant 3D points are used.

Figure 4. Side view of the reconstructed 3D points inside the predicted search region

Afterwards the horizontal profile is detected using a model-matching technique in the image space, and using the knowledge of the already detected vertical profile. The 3D parameters of the horizontal profile are extracted by solving the perspective projection equation. The roll angle is detected last, by checking the difference in height coordinates of the 3D points neighboring the left and right lane border. The detection results are used to update the lane state through the equations of the Kalman filter [12].

3.3. Object detection

The object detection process consists primarily in a labeling/grouping of the reconstructed 3D points by density and vicinity criteria. The Lane Detection module provides a pre-classification of the reconstructed 3D points. The vertical and frontal lane profiles can be used to obtain a separation of the 3D points into three important classes: road points, points above the road and points below the road.

For that purpose, an expected Y_E coordinate of each 3D point is computed using the vertical profile (4) and taking into consideration the roll of the road:

$$Y_E = Y_c + (X - X_c)\gamma \qquad (5)$$

where:

Y_C – is the vertical coordinate of the lane center, computed from (4);
X_C – is lateral coordinate of the lane center, computed from (1).

The points in a limit of 20 cm around this expected Y are considered as road points. The ones above are taken for object grouping, and the ones below are rejected. Are also rejected the points that are outside a predefined Space Of Interest (SOI) which limits also the object labeling in depth and lateral displacement.

The remaining 3D points labeled for object grouping are first projected on a top view plane (XZ). The top view space is compressed with the distance in order to obtain a constant point density with any distance [13]. Afterwards, adjacent points of the top-view histogram of the compressed coordinates (Figure 5) are connected and finally the objects segmentation is refined on vertical direction (Y).

Object tracking is used in order to obtain more stable results, and also to estimate the velocity of an object along the X, Y and Z axes [2], [13].
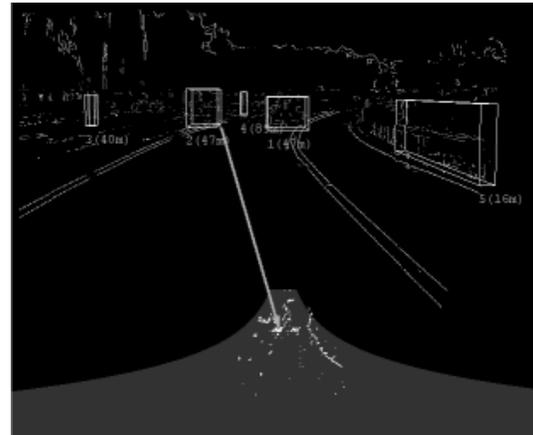
Figure 5. Top-view histogram of the compressed coordinates (lower part) and the segmented objects (upper part)

4. Applications for driving assistance

4.1. Lane detection

Lane detection can be performed in various scenarios and conditions. On highways the model can be easily extended to detect also the side lanes (Figure 6). On country roads it performs well even in the absence of painted lane delimiters (Figure 7).

In Figure 8 are presented the results of lane detection on a road with high vertical curvature while in Figure 9 are presented the results of the lane

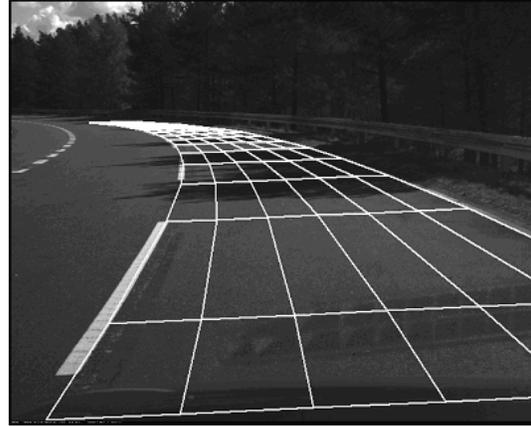detection on a road with high horizontal curvature and shaded lane markers.

Figure 6. Current lane, side-lanes and very far obstacles detected on highways

Figure 7. Current lane, opposite lane and far obstacles detected on partially marked country roads

Figure 8. Results of lane detection on non-flat roads with vertical curvature

Figure 9. Results of lane detection on roads with high horizontal curvature and shaded lane markers

4.2. Driving area detection

Having the 3D lane surface detected, and taking the advantages of the stereovision to detect the elevated continuous structures as fences and poles (Figure 10) or construction area delimiters (Figure 11) the so called *driving area* [2] can be detected.

Figure 10. Highway scenario with lane and driving area delimiters (poles and fences) detected

Figure 11. Urban scenario with construction area: lane and construction area delimiters are detected

4.3. Obstacle detection

The system can be used to detect any kind of static and/or moving obstacles identifiable trough 3D stereo-reconstruction. An obstacle can be considered any object which can interfere with the ego vehicle's trajectory. This can be a proceeding vehicle moving on the current or side lanes (Figures 6, 7 and 13), stationary vehicles parked on the road side (Figure 12), driving area delimiters as poles (Figure 10) or other traffic participants as pedestrians or bikers (Figures 12 and 13). Also traffic signs which are inside the SOI can be detected as 3D objects (Figure 13).

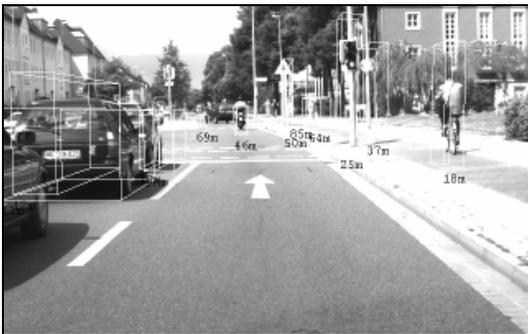**Figure 12**. Obstacles detection in urban scenarios (I)

Figure 13. Obstacles detection in urban scenarios (II)

5. Results

The developed applications were deployed on a standard 2.8 GHz Pentium® IV personal computer, and the average processing cycle takes up to 100 ms on 688 x 515 resolution grayscale images, securing a 10 fps detection rate. This makes the system suitable for real-time applications. Tests covered as much traffic conditions as possible, from highways to country roads (Figures 6-13).

The lane detection algorithm works with almost any kind of lane delimiters, provided that they obey the clothoid constraints and there are not too many noisy road features (a constraint usually fulfilled by most of the roads). The lane was detected even in the presence of high vertical or horizontal curvatures, or in the presence of obstacles on the current lane.

In all situations the obstacles were reliably detected and tracked, and their position, size and velocity measured. The detection has proven to have a maximum reliable working range of about 100 m, with maximum measurement errors of 5% in depth. The edge performance of the tracking algorithm was tested for ego-car speeds up to 150 km/h and relative incoming traffic speeds up to 250km/h.

6. Conclusions

In the current paper a stereovision sensor able to reconstruct the driving environment was presented. The developed algorithms are implementing two fundamental applications for any driving assistance system: lane detection and obstacle detection.

The presented lane detection method combines several detection and tracking techniques and enhances them for a more accurate and general lane model matching. The lane is represented as a 3D surface, and the assumptions of flat road and zero pitch and roll angles are eliminated due to the availability of the 3D information. This technique can be easily scaled up to another lane model, due to the generality of the model generation, matching and of the 3D parameters reconstruction method.

The developed obstacle detection method works on the 3D points corresponding to the object edges, in a large variety of traffic scenarios, and under real-time constraints. Because the vertical profile of the road is detected by the lane detection module a correct road-obstacle separation is possible for non-flat road environments. This way the grouping of the 3D points in relevant objects was greatly improved, and the objects 3D positioning accuracy was increased. The functions of the object detection module can be greatly extended in the future. Because any type of object is detected this algorithm can form the basis for any type of specific object detection system, such as vehicle detection, pedestrian detection, or even traffic sign detection. The classification routines can be performed directly on our detected objects, with the advantage of reduced search space and additional helpful information such as distance, size and speed,

The stereovision sensor output consists in the lane parameters and a list of the detected objects/obstacles in the format specified in chapter 2. This output can be used in a sensor-fusion system, which is a world description estimator integrating the information from

multiple heterogeneous sensors (vision, radar, range laser, ladar) into a unique model that is tracked over time. This way the week points of one sensor's can be compensated with the strong points off others obtaining a more robust, accurate and complete description of the driving environment.

Further researches will be focused on the implementation of driving assistance functions based on the presented applications, as: lane keeping and lane changing assistance, frontal collision avoidance, pedestrian collision avoidance, intersections assistance, ACC (Automatic Cruise Control) for highway and urban scenarios etc.

7. References

[1] Z. Sun, G. Bebis, R. Miller, "On-Road Vehicle Detection Using Optical Sensors: A Review", *in Proc. of IEEE Intelligent Vehicles Symposium,* Parma, Italy, June 14-17, 2004, pp.585-590.

[2] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, C. Pocol, S. Sobol, T. Graf, R. Schmidt, "Driving Environment Perception Using Stereovision", *Procedeeings of IEEE Intelligent Vehicles Symposium,* (IV2005), June 2005, Las Vegas, USA, pp.331-336.

[3] D. M. Gavrila, "Pedestrian Detection from a Moving Vehicle", *Proc. of European Conference on Computer Vision,* Dublin, Ireland, 2000, pp. 37-49.

[4] I.Ulrich and I. Nourbakhsh, "Appearance-Based Obstacle Detection with Monocular Color Vision", *Proc. of the AAAI National Conference on Artificial Intelligence, Austin, TX,* July/August 2000, 866-871.

[5] T. Kalinke, C. Tzomakas, and W. von Seelen, "A Texture based Object Detection and an Adaptive Model-based Classification", in *Proc. of IEEE Intelligent Vehicles Symposium'98*, (Stuttgart, Germany), Oct. 1998, pp. 341–346.

[6] A. Kuehnle, "Symmetry-based vehicle location for AHS", in *Procs. SPIE Transportation Sensors and Controls: Collision Avoidance, Traffic Management, and ITS*, vol. 2902, (Orlando, FL), Nov. 1998, pp. 19–27.

[7] U. Franke, I.Kutzbach, "Fast Stereo-based Object Detection for Stop&Go Traffic", *IEEE International Conference on Intelligent Transportation Systems (ITSC 96).* Tokyo, Japan, 1996, pp.339-344.

[8] M. Bertozzi, A. Broggi., GOLD: a Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection, *IEEE Transactions on Image Processing*, 7(1):62-81, January 1998,

[9] R. Aufrere, R. Chapuis, F. Chausse, "A model-driven approach for real-time road recognition", *Machine Vision and Applications,* Springer-Verlag, pp. 95-107, 2001

[10] Y. Wang et al., "Lane detection using spline model", *Pattern Recognition Letters* vol. 21 (2000), no. 8, pp. 677-689.

[11] M. Bertozzi, A. Broggi, A. Fascioli, and A. Tibaldi, "An Evolutionary Approach to Lane Markings Detection in Road Environments", In *Atti del 6 Convegno dell'Associazione Italiana per l'Intelligenza Artificiale*, Siena, Italy, September 2002. pp. 627-636.

[12] S. Nedevschi, R..Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, "3D Lane Detection System Based on Stereovision", *in Proc. of IEEE Intelligent Transportation Systems Conference (ITSC),* Washington, USA, October 4-6, 2004, pp. 292-297.

[13] S. Nedevschi, R. Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, "High Accuracy Stereo Vision System for Far Distance Obstacle Detection", *in Proc. of IEEE Intelligent Vehicles Symposium,* Parma, Italy, June 2004, pp.161-166.

[14] J. Goldbeck, B. Huertgen, "Lane Detection and Tracking by Video Sensors", *In Proc.of IEEE International Conference on Intelligent Transportation Systems,* October 5-8, 1999, Tokyo Japan, pp. 74–79.

[15] Jean-Yves Bouguet, Camera Calibration Toolbox, http://www.vision.caltech.edu/bouguetj/calib_doc/.

[16] T. Marita, F. Oniga, S. Nedevschi, T. Graf, R. Schmidt, Camera Calibration Method for Far Range Stereovision Sensors Used in Vehicles, *Proceedings of IEEE Intelligent Vehicles Symposium*, (IV2006) , June 13-15, 2006, Tokyo, Japan, pp. 356-363.

[17] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision", *Prentice-Hall*, 1998.