High Accuracy Stereovision Approach for Obstacle Detection on Non-Planar Roads

Sergiu Nedevschi, Radu Danescu, Dan Frentiu,
Tiberiu Marita, Florin Oniga, Ciprian Pocol
Computer Science Department
Technical University of Cluj Napoca
Daicoviciu Str. 15, Cluj Napoca
Romania
Sergiu.Nedevschi@cs.utcluj.ro, Team@vision.utcluj.ro

Thorsten Graf, Rolf Schmidt
Group Research, Electronics
Volkswagen AG
Brieffach 1776, D-38436 Wolfsburg
Germany
Thorsten.Graf@volkswagen.de, Rolf4.Schmidt@volkswagen.de

***Abstract*–This paper will present an obstacle detection system that that relies on the 3D information provided by stereo reconstruction. The 3D features must be separated in road features and obstacle features. Instead of relying on the flatness of the road, the vertical road profile is modeled as a clothoid, and is estimated from the lateral projection of the 3D points. The points above the road are selected for grouping into objects, based on vicinity criteria and the variation of the point density with the distance. The resulted objects are used as measurements for a model-based tracking algorithm. The resulted system is a high-accuracy, far distance obstacle detector, able to function in a large variety of real-world scenarios.**

I. INTRODUCTION

Detecting the surrounding environment of a moving vehicle is a complex and challenging task. One of the most important components of the environment is the set of obstacles, which can be other vehicles, stationary side objects, pedestrians, etc. Detection of the obstacles implies, directly or indirectly, the use of some kind of 3D information, and this is the reason why the active sensors, such as laser or radar, are the prime choice of industry. However, the use of a high resolution, high accuracy stereovision algorithm provides comparable results in 3D estimation, while delivering a larger amount of data, thus making the grouping and tracking tasks easier, and allowing a subsequent classification of the obstacle.

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 approach uses techniques such as object model fitting [1], color or texture segmentation [2,3], symmetry axes [4] etc. The estimation of 3D characteristics is done after the detection stage, and it 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, this feature being anything from a point to a complex structure. 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. The real time capability of the method is another important constraint. Such a method was proposed in [5]. The full 3D reconstruction of the visible scene is performed only on vertical or oblique edges. The list of obtained 3D points is grouped into objects based solely on density and vicinity criteria. The flat road assumption for the ground/obstacle points separation process was used. The system detects obstacles of all types, outputting them as a list of cuboids having 3D positions and sizes. The detected objects are then tracked using a multiple object-tracking algorithm, which refines the grouping and positioning, and detects the speed and orientation.

An important part in the obstacle detection process is the separation of the obstacle points from the road points. Most of the roadway obstacle detection methods are based on the flat road assumption [6,7]. This is a poor model since deviations from the flat road may be as large as or larger than the obstacles we wish to detect. In consequence the road objects separation and the 3D objects position estimation cannot be done. Therefore the non-flat road assumption is compulsory for a robust object detection method. In literature this assumption was introduced by non-flat road approximation by series of planar surface sections [8,9] or by modelling of the non-flat roads by higher order surfaces [10,11]. For instance the methods presented in [11,12,13] are fitting the parameters of a 3D clothoid model of the road lane using a monocular image and supplementary lane geometry constraints.

Our approach presented in this paper will model the vertical profile of the road surface with such a clothoid curve fitted directly on the detected 3D road surface points. These 3D road points are detected using a high accuracy stereovision method [5]. The obtained vertical profile will be used for the road-obstacle separation process in order to have a proper grouping of the 3D points in obstacles and precise estimation of their 3D position in the driving environment.

II. CALIBRATION OF THE STEREO SYSTEM

In order to reconstruct and measure the 3D environment using stereo cameras, 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 to a global world coordinate system).

The intrinsic parameters of each camera are calibrated individually. The estimated parameters are the focal length and the principal point coordinates and the lens distortions. The parameters are estimated by minimizing the projection error from multiple views of a set of control points placed on a coplanar calibration object with known geometry. For a stereo system of two cameras, the obtained intrinsic

parameters can be refined by inferring the stereo information available. This is done by introducing a new constraint in the estimation process which considers also the projection error of the control points image coordinates from one image to another [15].

The extrinsic parameters of the cameras are estimated by minimizing against the extrinsic parameters the projection error for a set of 3D control points with measured coordinates in a world reference system [16, 17]. Due to the requirement that the reconstruction must work for far distances, a special calibration field is set up, with special targets, covering a distance of at least 40 m, as seen in figure 1. The 3D coordinates of each target are known in advance. The left and right images of the scene, taken from the on-board cameras, are processed and the image position of each calibration target is automatically detected based on the target's special shape.

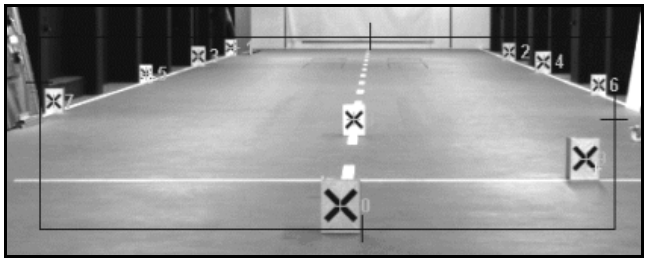

Fig. 1. The calibration scene and the detected targets

The obtained extrinsic parameters for each camera are a translation vector of the camera in the world coordinate system (**T**j) and a rotation vector (**R**j) relative to the same coordinate system.

The accuracy of the calibration is essential for the performance of the system, and this is the reason why this step must be performed with great care and patience.

III. STEREO RECONSTRUCTION

The stereo reconstruction algorithm that is used is mainly based on the classical stereovision principles available in the existing literature [14]: find pairs of left-right correspondent points and map them into the 3D world using the stereo system geometry determined by calibration.

Constraints, concerning real-time response of the system and high confidence of the reconstructed points, must be used. In order to reduce the search space and to emphasize the structure of the objects, only edge points of the left image are correlated to the right image points. Due to the cameras horizontal disparity, a gradient-based vertical edge detector was implemented. Non-maxima suppression and hysteresis edge linking are being used. By focusing to the image edges, not only the response time is improved, but also the correlation task is easier, since these points are placed in non-uniform image areas.

Area based correlation is used. For each left edge point, the right image correspondent is searched. The sum of absolute differences (SAD) function [7] is used as a measure of similarity, applied on a local neighborhood (5x5 or 7x7 pixels). Parallel processing features of the processor are used to implement this function. The search is performed along the epipolar line computed from the stereo geometry for general camera configuration.

To have a low rate of false pairs, only strong responses of the correlation function are considered as correspondents. If the global minimum of the function is not strong enough relative to other local minimums than the current left image point is not correlated. In figure 2 a successful correlation is shown along the first column, while the last two columns show ambiguous similarity functions with rejected correspondents. Repetitive patterns are rejected and only robust pairs are reconstructed.

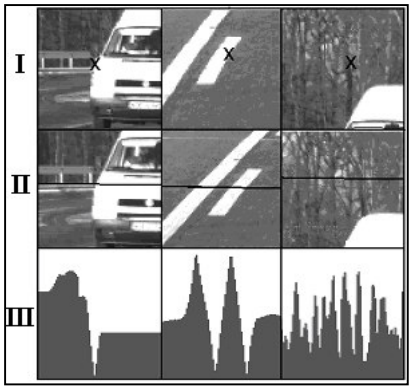

Fig. 2: Three correlation scenarios are shown on each column. Left image point marked by 'x' on row I, right image search area and the epipolar line on row II and the correlation function on row III.

A parabola is fitted to a local neighborhood (3 or 5 points) of the global correlation minimum in order to detect the stereo correspondence with sub-pixel accuracy. The obtained accuracy is about 1/4 to 1/6 and is dependent of the image quality (especially noise level and contrast). Our tests proved that the 3-neighbors parabola works better than the other one.

After this step of finding correspondences, each left-right pair of points is mapped into a unique 3D point [14]. Two 3D projection rays are traced, using the camera geometry, one for each point of the pair. By computing the intersection of the two projection rays, the coordinates of the 3D point are estimated.

IV. VERTICAL ROAD PROFILE ESTIMATION

Many of the obstacle detection methods assume a flat road profile. Some take into account the car pitching – therefore admitting some degree of vertical profile change – but fail to account for a possible curved vertical profile. We'll try to extract the vertical profile of the road by approximating it with a first order clothoid curve (in the ego-car coordinate system):

$$Y_E = -Z_E\alpha + c_{0,v}\frac{Z_E^{\,2}}{2} + c_{1,v}\frac{Z_E^{\,3}}{6} \qquad (3)$$

where:

α, *is* the pitch angle of the ego-car
$c_{0,y}$ – vertical curvature
$c_{1,y}$ – variation of the vertical curvature

In order to extract the coefficients α, $c_{0,v}$ and $c_{1,v}$ which

will completely describe the vertical profile of the road, we'll make use of the 3D road points reconstructed by stereovision. The main advantage of using stereovision is the ability of directly extracting the vertical profile, independently of the lane detection process, sometimes even independently of the presence of any kind of delimiters. The key assumption, which makes this possible, is that there are none or very few 3D points under the road plane. Having a list of 3D points, it is easy to obtain a lateral projection in the YOZ plane, like in figure 3.

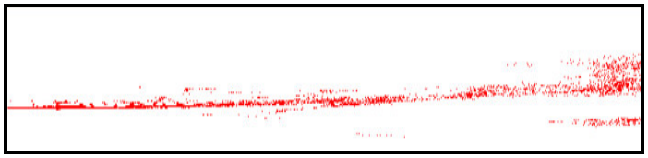

Fig. 3: Lateral view of the 3D points set (the projection of the 3D points on the YOZ plane)

As easily can be seen, there is a lot of noise in the set of points, and therefore a simple fitting of the curve to the lower points, or a least-square clothoid fitting is not enough. Our approach to detecting the vertical profile takes two simplifying assumptions:

- In the close vicinity of the ego-vehicle (30m), the points are on a straight line, and the effect of the curvature is sensed only after the 30m interval
- The effect of roll is negligible for the vertical profile detection, that is, the vertical displacement due to roll is negligible in comparison to the displacement due to pitch and vertical curvature.

These assumptions allow us to regard the problem as a 2D curve fitting to a set of 2D points corresponding to the lateral projection of the reconstructed 3D points (figure 3).

With these assumptions, first we want to estimate the pitch angle of the ego car coordinate system relative to the road surface (angle α from equation 3). 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 0-30 m (in which we consider the road flat). Therefore, an angle histogram is built for each possible pitch angle, using the near points. From the point of origin of the world coordinate system, uniformly spaced rays are cast. The distance between the rays must correspond to the acceptable error in pitch angle estimation, and the total set of rays must cover all the possible pitch values. The process is depicted loosely in figure 4. The points along each ray are counted in a polar histogram (a histogram having an entry for each ray). This histogram then undergoes a smoothing process, by convolution with a 1D Gaussian kernel. The result will look similar to figure 5.

Then the histogram is searched from under the road upwards. The first angle having a considerable amount of points aligned to it is taken as the pitch angle.

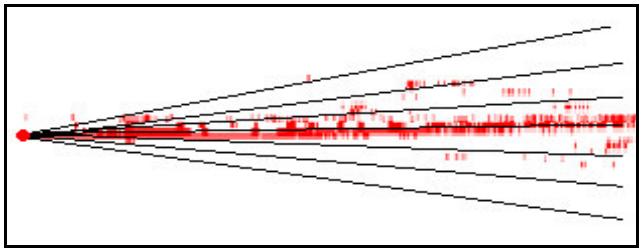

Fig. 4. The angle search space, depicted by a set of rays. The points along these rays will be counted in a polar histogram

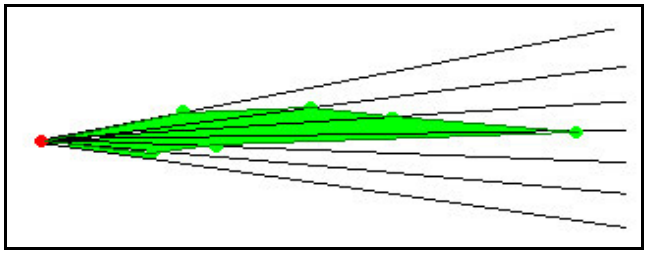

Fig. 5. The polar histogram

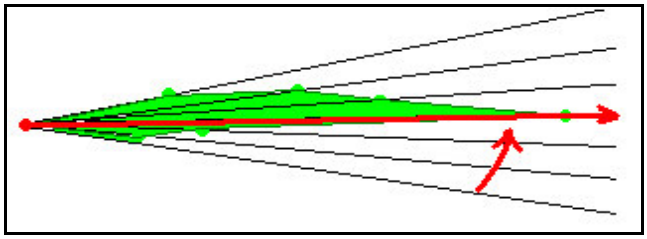

Fig. 6. Finding the ray that represents the pitch angle

After detecting the pitch angle, detection of the 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 (> 30 m) are taken into account, because the effect of a curvature is felt only in more distant points. The obtained vertical clothoid profile of the road is shown in figure 7. The variation of the vertical curvature will not be detected due to the large errors and to the fact that its value is very small and can be safely considered zero.

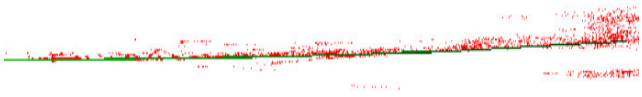

Fig. 7. The vertical profile fitted to the ground points

V. GROUPING THE POINTS INTO OBJECTS

We use only 3D points situated above the road surface. The road surface is modeled by the following clothoid equation in the world coordinates system:

$$Y = c_{0,v}\frac{Z^2}{2} + c_{1,v}\frac{Z^3}{6} \qquad (4)$$

The road/obstacle separation (figure 8) of the 3D points is done using thee following constraints:

- if $|Y_W - Y| < \tau$, the point is on the road surface, and classified as road point
- if $(Y_W - Y) < -\tau$, the point is below the road, and is rejected
- if $(Y_W - Y) > \tau$, the point is above the road

The threshold τ is a positive constant and its value is chosen depending on the on the error estimation of the disparity with the depth, and on the error estimation of the clothoid parameters and possible torsion of the road.

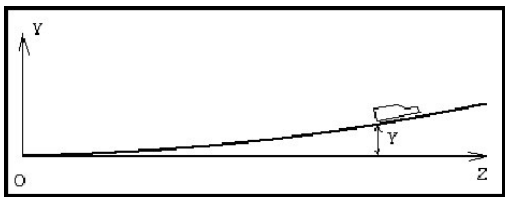

Fig. 8: Lateral view of the road surface in the world coordinate system

Some supplementary constraints are used to restrict the 3D points above the road: points higher then 4m above the road surface, points that are too lateral or too far are rejected. The remaining points belong to the so-called Space of Interest (SOI) in which is performed the grouping of the 3D points in objects. For the road geometry we have made the following assumptions: in highway and most of country road scenarios the horizontal curvature is slowly changing and the torsions can be neglected in our detection range (up to 100m). Therefore knowing the road vertical profile would be enough to characterize the driving surface in the SOI.

In our SOI, no object is placed above other. Thus, on a satellite view of the 3D points in SOI, we are able to distinguish regions with high points density, representing and locating objects. Regions with low density are assumed to contain noisy points and are neglected. The satellite view of 3D points is analyzed to identify objects. In figure 9 such a view is shown.

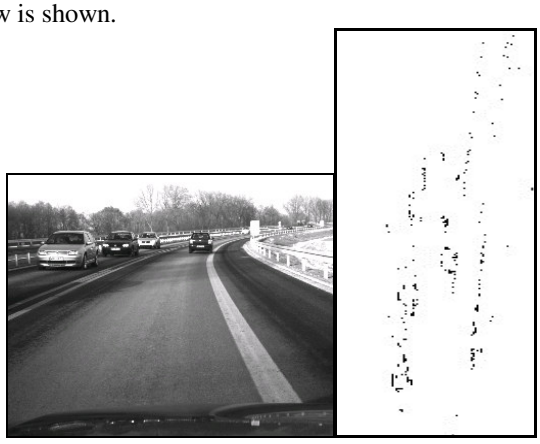

Fig. 9: Left image and the satellite view of the 3D points

An important observation is that the 3D points are more and more rare as the distance grows. To overcome this phenomenon, we compress the satellite view of the space, depending on distance, in such a way that local density of points, in the new space, is kept constant. Regardless the distance to an object, in the compressed space, the region where that object is located will have the same points density.

The compression factor depends on distance (Z):

$$Scale(Z) = f \cdot \frac{1}{Z} \cdot k \tag{5}$$

Where: Z – distance
F – focal length of the cameras
k – is a manually chosen factor, depending on the richness of 3D reconstructed points with the current reconstruction method. For X and Z axis the values for k can be different.

The k factors are chosen to satisfy two conditions of the found objects:
- to not divide a real object into many smaller objects;
- to not unify many real objects into one bigger object.

The equations used to find the position (row, col) in the compressed space, of a point (X, Z) in the uncompressed space, are:

$$row = \log_{1+\frac{k}{f}} \frac{z}{z_{\min}} \tag{6}$$

Z_{min} = low distance limit of SOI

$$col = X \cdot Scale(Z) \tag{7}$$

The compressed space of the scene depicted in figure 6 is presented in the figure 10.

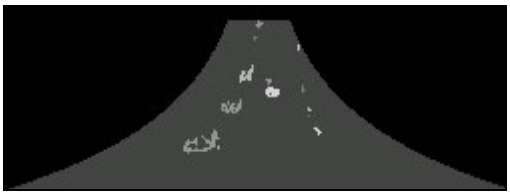

Fig. 10. The compressed space and the identified objects

Also, in figure 10, objects were identified as dense regions.

For the resulted objects their limits along the Y-axis are found. In figure 11 the cuboids circumscribing objects are shown.

Fig. 11: Perspective view of object cuboids painted over the image

VI. RESULTS

The detection system has been deployed on a standard 1 GHz Pentium® III personal computer, and the whole processing cycle takes less than 100 ms processing time, therefore securing a 10 fps detection rate. This makes the system suitable for real-time applications. The system has been tested in various traffic scenarios, both offline (using stored sequences) and online (on-board processing), and acted well in both conditions. 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 working range of about 90 m, with maximum of reliability in the range 10-60 m. The depth measurement error is, naturally, higher than one can obtain from a radar system, but it is very low for a vision system: less than 10 cm of error at 10 m, about 30 cm of error at 45 m and about 2 m of error at 95 m.

In figure 12 the detection results on a non-flat road are outlined. The scene from figure 6.a is at the end of a concave slope. The detected road surface has a concave vertical curvature $c_{0,v} \approx 2e^{-4}$. The far objects (the two cars at 71m, respectively 81m and the traffic sign from 91m have vertical offsets in the world coordinates system (having the XOZ plane coincident with the road surface below the current car position) of 0.51m, 0.66m and 0.96m respectively, due to the non flat road. But using the non-flat road modeled by a vertical clothoid (figure 8) the objects are detected correctly on the road surface.

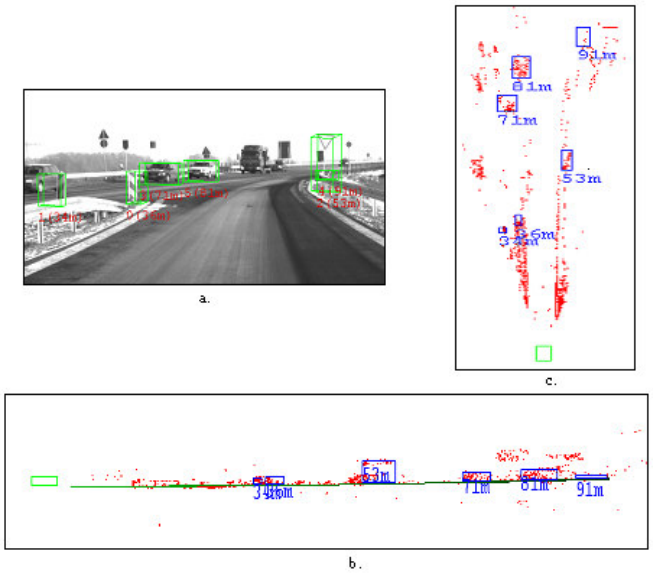

Fig. 12: a. Image of the scene with the detected object (cuboids with ID and distance); b. Side view of the detected objects and the detected road surface; c. Top view of the detected objects.

The result of obstacle detection in multiple scenarios is shown in figures 13 through 15. The system was tested and performed well in urban traffic (figure 13), busy highway traffic (figure 14) and country roads (figure 15), where the main difficulty were the high relative speeds of the incoming objects.

Fig. 13: Detection results in urban traffic

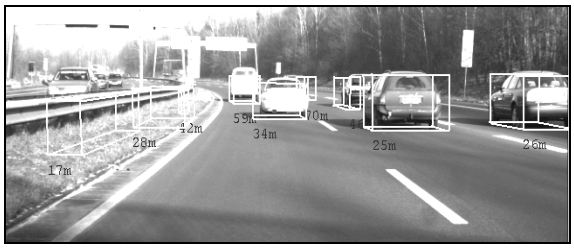

Fig. 14: Detection results in highway traffic

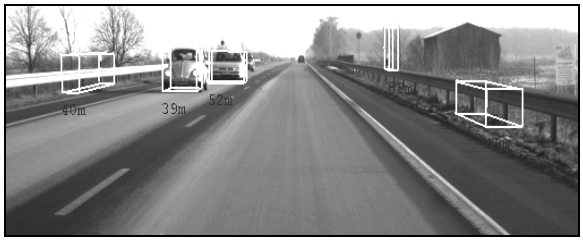

Fig. 15: Tracking of incoming traffic on country road

VII. CONCLUSIONS

We have presented a stereovision-based obstacle detection system that reconstructs and works on 3D points corresponding to the object edges, in a large variety of traffic scenarios, and under real-time constraints. Because the stereovision module reconstructs any feature in sight (that means also the road features) the vertical profile of the road was detected. This way a correct road-obstacle separation was possible. The grouping of the 3D points in relevant objects was greatly improved, and the objects 3D positioning accuracy was increased.

The functions of this system can be greatly extended in the future. An intelligent correlation function should be developed, one that can disambiguate, not reject, repetitive patterns and reconstruct points from horizontal edges. Moreover, 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, which can also reduce the class hypotheses. The vertical road profile detection from stereovision can be the base for a 3D lane detection algorithm, which will give a complete 3D description of the driving environment in a lane related coordinates system.

VIII. REFERENCES

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

[2] 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.*

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

[4] 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.

[5] 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", *IEEE Intelligent Vehicles Symposium,* Parma, Italy, June 14-17, 2004.

[6] J.Weber, D. Koller, Q.-T. Luong and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance", *Fifth International Conference on Computer Vision,* Boston 1995*, Collision Avoidance and Automated Traffic Management Sensors*, Proc. SPIE 2592, 1995.

[7] T. A. Williamson, "A high-performance stereo vision system for obstacle detection", *PhD Thesis CMU-RI-TR-98-24, Robotics Institute Carnegie Mellon University*, Pittsburg, September, 1998.

[8] J. Hancock, "High-Speed Obstacle Detection for Automated Highway Applications", *Tech. Report CMU-RI-TR-97-17, Robotics Institute, Carnegie Mellon University*, Pittsburg, May, 1997

[9] R. Labayrade, D. Aubert, J.P. Tarel, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through V-disparity Representation", Proceedings of IEEE Intelligent Vehicle Symposium, June (IV'2002), 18-20 June 2002, Versailles, France.

[10] J. Goldbeck, B. Huertgen. Lane Detection and Tracking by Video Sensors, IEEE International Conference on Intelligent Transportation Systems (ITSC 99). Tokyo, Japan, 5-8. Oktober 1999.

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

[12] R. Aufrere, R. Chapuis, F. Chausse, "A fast and robust vision-based road following algorithm", *IEEE-Intelligent Vehicles Symposium 2000,* Dearborn, Michigan (USA), October 2000, pp.192-197.

[13] Takahashi, A. & Ninomiya, Model-based lane recognition. *Proceedings of the IEEE Intelligent Vehicles Symposium 1996*, pp. 201–206.

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

[15] Jean-Yves Bouguet, "Camera Calibration Toolbox for Matlab", MRL - Intel Corp., http://www.vision.caltech.edu/bouguetj/calib_doc/, 2003.

[16] S. Nedevschi, T. Marita, M. Vaida, R. Danescu, D. Frentiu, F. Oniga, C. Pocol, D. Moga, "Camera Calibration Method for Stereo Measurements", Journal of Control Engineering and Applied Informatics (CEAI), Vol.4, No. 2, pp.21-28, Bucuresti, Romania, 2002

[17] S. Nedevschi, T. Marita, R. Danescu, F. Oniga, D. Frentiu, C. Pocol, "Camera Calibration Error Analysis in Stereo Measurements", *microCAD International Scientific Conference,* pp. 51-56, Miskolc, Hungary, March 2003.