Increased Accuracy Stereo Approach for 3D Lane Detection

Sergiu Nedevschi, Florin Oniga, Radu Danescu

Computer Science Department Technical
University of Cluj-Napoca
Cluj-Napoca, 400020, ROMANIA

Thorsten Graf, Rolf Schmidt

Electronic Research
Volkswagen AG
Wolfsburg, 38436, GERMANY

Abstract

A new approach for the stereovision problem is presented, aiming to increase the accuracy of stereo reconstruction. The proposed method is edge-based and consists of the correlation of left and right contours, detected with sub-pixel accuracy. The steps of the stereo matching process are: segmentation of each contour into basic contours (strongly- and weakly-curved), detection of correspondences between left-right image basic contours, and computation with sub-pixel accuracy of the pairs of corresponding left-right edge points. By consequence, one achieves the highest correlation and 3D reconstruction accuracy. 3D lane detection, based on a clothoidal model, is evaluated with the proposed stereo algorithm on synthetic and real world images.

1 Introduction

A 3D point can be reconstructed using stereovision if its projections onto the left-right image planes are determined and the optics and geometry of the two cameras are known [1], [2]. Finding pairs of left-right image points that represent the same 3D point is called stereo matching. The stereo matching problem is the most difficult part of reconstruction by stereovision.

Several aspects of stereo matching should be discussed: what image features are considered, what methods are used to find the correspondent features, what accuracy is achieved in finding the correspondents, the real time performance etc.

The features selected for reconstruction are application specific: low-level (discrete image points) or high-level (edge segments or uniform image patches). High-level features are used less often because they impose strong constraints to the shape of the objects from the scene (ex. apriori known geometric shapes [3]). Most approaches use discrete image points for reconstruction. Some perform full image reconstruction, called dense stereo, usually on scenes with rich textures [4], [5], [6]. In [7] only image points adjacent to vertical image structures are reconstructed. A real time stereo reconstruction system that reconstructs edge points placed on vertical and oblique edges is presented in [8].

The image correspondent is searched using area correlation, along the epipolar line, in the pair image [1]. The epipolar line is non-parallel to the image rows for the general geometry configuration. To make the search easier almost all algorithms use rectification first to obtain a canonical geometry: aligned image planes with horizontal epipolar line. The search is simple but accuracy is lost because of image re-sampling. In [8], [9] two possible methods were presented, one using image rectification and the second one without rectification. A better 3D accuracy for objects is obtained when the general geometry is used. A robust 3D lane detection algorithm was presented in [9].

If the calibration of the system is precise, the 3D reconstruction's accuracy depends on the accuracy of the left to right features correspondence. The standard method is fitting a parabola to a neighbourhood around the minimum position of the correlation function [5]. The method has an accuracy of about a quarter of a pixel. Each left-right

correspondent pair of points is mapped into the 3D world by computing the intersection of the two projection rays, a technique called triangulation [1], [2], [10].

2 Environment Model

The environment model is similar to the one in [9] and consists of a 3D lane, driving area delimiter, and 3D objects. We briefly describe in this section the 3D systems of coordinates and the 3D lane model. The stereo algorithm proposed in this paper improves directly only the 3D lane detection.. The driving area delimiter and the 3D objects models will not be detailed, because they are only indirectly improved, by better 3D lane detection.

The ego-car coordinates system has its origin on the ground plane and is the projection of the middle of the car front axis. Its axes (X_C,Y_C, Z_C) are parallel with the tree main axis of the car. The position of the two cameras relative to the car coordinate system is completely determined by the translation vectors $\mathbf{T}^i$ and the rotation matrices $\mathbf{R}^i$.

The world coordinate system has its origin in the middle of the current lane, the X_W axis is contained in the road plane and is perpendicular to the lane delimiters, the Y_W axis is perpendicular on the road surface, the Z_W axis is parallel with the tangents to the lane delimiters. The world coordinates system is moving along de lane mid axis together with the car and thus only a lateral and a vertical offsets between the origins of the two coordinates systems exists (vector $\mathbf{T}_C$ from Fig. 1). The relative orientation of the two coordinates systems ($\mathbf{R}_C$ rotation matrix) will be also changing due to static and dynamic factors: the loading of the car is a static factor; acceleration, deceleration and steering are dynamic factors, which also cause the car to change its height, pitch, yaw and roll angles with respect to the road surface.

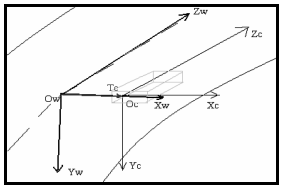

Fig. 1. The world, car and cameras coordinates systems

The lane is modelled as a 3D surface, defined by the vertical and horizontal clothoid curves. Lane detection is regarded as the continuous estimation of the following parameters [9]:

- 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 car reference system from the lane related world coordinate system (X component of $\mathbf{T}_C$)
- α, γ, ψ – the pitch, roll and yaw angles of the car (the rotation angles between the car reference system and the world reference system – the rotation angles corresponding to $\mathbf{R}_C$).

3 Increasing the accuracy of stereo reconstruction

The standard correlation method used in stereo vision has two main drawbacks: the left image features have pixel accuracy and are usually different than the interest image features (Fig. 2), and the right image feature is selected at pixel accuracy and then sub-pixel accuracy is obtained by interpolation (that has residual errors). For roads with high-curvature, distant lane-markings are close to horizontal and the matching cost has a non-sharp minimum, resulting in a lower accuracy.

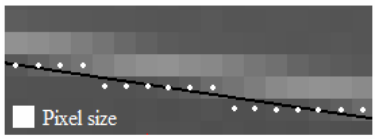

Fig. 2. Zoomed image patch of a lane-marking at 65 meters depth in a curve. White dots are the pixel centers used for correlation. The black line represents the true sub-pixel lower edge of the marking

These drawbacks are most visible for oblique edges, close to horizontal. For complex tasks, such as accurate lane detection in the 3D space, driving area detection and obstacle detection at high depths, an increased 3D accuracy is needed for reconstruction.

The 3D reconstruction's accuracy is directly linked to the discrete nature of the images. Sub-pixel

accuracy edges can be detected in both the left and right images and increased accuracy correspondences can be established between the sub-pixel edge points (Fig. 3).

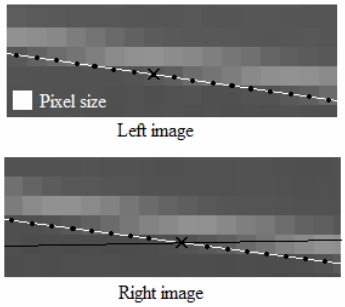

Fig. 3. Sub-pixel accuracy edges (black dots) are detected in both images and the correspondent of a left image point (the black x) is computed in the right image as the intersection between the epipolar line and the right image sub-pixel edges (linear piecewise approximation)

Sub-pixel accuracy epipolar lines can be computed from the system's geometry so image rectification is not required. This avoids accuracy loss by image re-sampling. The intersection of the epipolar line and the right contour must be computed using several adjacent sub-pixel edge points. Therefore, sub-pixel edges will be extracted as contours (chains of points) instead of individual points.

The stereo algorithm should have several steps: the sub-pixel contour extraction, contours segmentation, stereo matching along contours, and 3D reconstruction of pairs of sub-pixel accuracy points.

3.1 Sub-pixel accuracy contour extraction

An important part of the new algorithm is the detection of edge points with sub-pixel accuracy, in the form of contours. Although many theoretical aspects are presented about sub-pixel edges in the field of image processing, few methods exist that also extract contours with sub-pixel accuracy. Our choice was a robust method [12], based on the second numerical approximation of the Canny optimal operator [13]. It detects edges using the sub-pixel position of the zero-crossings of the second order derivative along the gradient's direction.

3.1.1 The sub-pixel contour extraction algorithm

The algorithm has the following steps [12]: the image filtering (by Gaussian smoothing), the computation of the first and the second order partial derivatives, the computation of the second order directional derivative in the gradient direction, contour tracking, extraction and closing. The second directional derivative in the gradient direction of image I can be computed as in [13]:

$$\frac{\partial^2}{\partial n^2}(I) = \frac{\frac{\partial^2 I}{\partial x^2}\left(\frac{\partial I}{\partial x}\right)^2 + 2\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\frac{\partial^2 I}{\partial x \partial y} + \frac{\partial^2 I}{\partial y^2}\left(\frac{\partial I}{\partial x}\right)^2}{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (1)$$

The sub-pixel position of a zero crossing is computed by interpolating the second order directional derivative inside a 2x2 pixels cell by a function: linear or cubic. The zero crossings detection, contour tracking and edge extraction take place simultaneously [12] on the second directional derivative image.

3.1.2 Sub-pixel accuracy of the contour detector

The analysis of the contour elements detected by this method revealed the presence of systematic sub-pixel level noise (Fig. 4). This noise limits the accuracy of the 3D reconstruction. The error mainly depends on the direction and the strength of the edge [11].

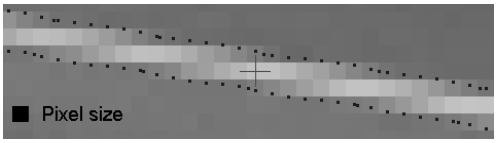

Fig. 4. A periodic error in the sub-pixel location due to the interpolation function

Synthetic images, with known sub-pixel edge position, are used for estimating the errors of gradient-based sub-pixel edge detectors [11]. We have used the same model (Fig. 5) to estimate the errors of the contour detector.

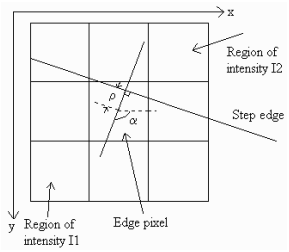

Fig. 5. Step edge model used in evaluations. The intensity of the edge pixel is the area-weighted average of the two intensities around the edge

3.1.3 Error evaluation results

Each interpolation method of the contour detector was evaluated using various values for two edge properties: edge strength (the difference between the two intensities around the edge) and edge orientation.

Ten distinct values were used for the slope of the edge, ranging from 0.05 to 0.95 (0, π/4, the rest is symmetrical). Three strengths were used for the edge: strong edge –145, normal edge – 80, and weak edge – 30.

Fig. 6. Samples of strong, normal and weak edges are shown together with the true edge line (black) with the slope 0.2.

The maximum measured errors (in pixels) are shown in Table 1. Systematic errors are too high for weak edges, more than one tenth of a pixel.

TABLE 1

Interpolation / Strength	Linear	Cubic
Weak	0.14	0.14
Normal	0.09	0.06
Strong	0.08	0.05

3.1.4 Propagation of sub-pixel errors to the 3D reconstruction

A simple canonical stereo model can be used (close to the system we use for testing), baseline B=0.35 m and focal length F=1200 pixels, to evaluate how the sub-pixel errors propagate to the reconstructed 3D coordinates.

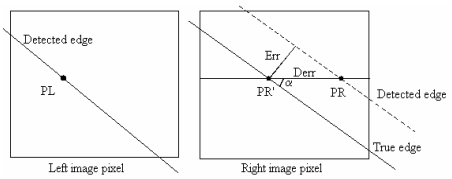

Fig. 7. The disparity error is *Derr*, assuming a true detected edge in the left image, and a sub-pixel error *Err* of estimating the edge position in the right image. Each square represents one pixel

The true correspondent of the left image (sub-pixel accuracy, Fig. 7) point *PL* is *PR'*. The depth error *Zerr* of a point reconstructed at depth *Z*, is computed using the canonical geometry:

$$Derr = \frac{Err}{\sin(\alpha)}, \quad Zerr = \left| \frac{Z^2 * Derr}{B * F - Z * Derr} \right| \qquad (2)$$

Considering a systematic sub-pixel error of 0.1 pixels, the values of the depth error are shown (in meters) in Table 2, for a depth range (first column) between 10 and 100 meters and the angle (first row) of the edge between 0.1 (close to horizontal) and 1.5 radians (close to vertical).

TABLE 2

Z \α	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5
10	0.2	0.1	0.05	0.04	0.03	0.03	0.03	0.02
30	2.1	0.7	0.46	0.35	0.29	0.26	0.25	0.24
50	6.0	2.0	1.32	1.01	0.85	0.77	0.73	0.73
70	11.8	4.1	2.65	2.05	1.75	1.60	1.55	1.56
100	24.4	8.6	5.64	4.44	3.87	3.62	3.58	3.71

The effect of sub-pixel errors increases roughly as the edge gets closer to the horizontal level (perspective projection of lane markings of the side lanes and curves).

3.1.5 Sub-pixel error correction through least square line fitting

One important feature of the error is periodicity and symmetrical distribution around the true edge position. The local image curvature of affected features (lane markings and road borders) is usually small.

A robust way to reduce the errors is to fit 2D lines to sub-pixel contour patches and correct the sub-pixel

points to fit the lines. The length of the patch is computed proportional to the error's period (function of the edge's slope), to insure a strong filtering. There are several places where filtering is not used because it can destroy the local shape of the contour: corners and places with high 2D image curvature.

This filtering provided accurate results: the error is reduced 3 up to 5 times in magnitude on synthetic images. On real images a visual evaluation was performed. The filtering proved efficient even for errors introduced by image noise.

3.2 Contours segmentation

Contours are usually detected along more than one 3D object due to perspective projection (Fig. 8.a). Since our matching (correlation) scheme is based mainly on contour following and continuity (two successive contour points are neighbours in the 3D world), this spread effect might introduce false correlations. In addition, when the edge direction varies roughly (corners), the sub-pixel accuracy is low (signal frequency much higher than the image sampling rate) and the correlation works better using the traditional area-based method. To make the task of correlation easier, these contours are split into smaller basic contours. Two classes of basic contours are defined: strongly-curved and weakly-curved. The local angle (Fig. 8.b) of the contour is computed by triangulation on a contour patch. This angle is used to discriminate between strongly-curved and weakly-curved points.

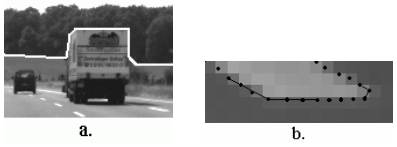

Fig. 8. A complex contour (white) spreads across more than one 3D object (a). Two apertures can be used to determine angles (b)

After this classification, basic contours are extracted as the sub-contours with the same class of the points. Each basic contour receives the class of its points: strongly-curved for corners or weakly-curved otherwise.

3.3 Stereo matching

For each left image pixel, the search space for a correspondent is a segment placed on the epipolar line in the right image. The extremities of the segment are computed from the geometry of the system in such a way that the 3D coordinates of the reconstructed point are inside of a 3D (a large cube) space of interest.

3.3.1 The correlation function

To determine the similarity between left-right image points a complementary comparison condition was introduced along with the classical area-based correlation function. For oblique edges, with different 2d image orientations, area-based correlation is appropriate (not invariant to rotation). The intensities of the surfaces placed around the left-right edge points are also compared [14] to reject false correlations.

3.3.2 The matching scheme for basic contours

The right correspondent contour is searched for each left image basic contour *CL* that is weakly-curved and is not horizontal (slope lower than 0.05). A simple correlation scheme (example in Fig. 9) is used determine pairs of left-right contours:

1) The middle point *PL* of *CL* is selected for correlation.
2) The correspondent *PR* of *PL* is searched in the right image along the discrete epipolar line in a search interval determined by the stereo geometry, considering the followings: only right image locations where a contour exists are considered (CR1 and CR2); the left-right contour orientations must be close enough; only strong minimums (compared to the other local matches) are accepted.

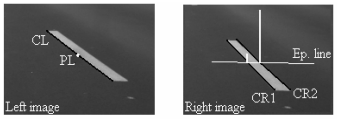

Fig. 9. A simple correlation process is presented for left contour *CL* and two possible right image candidates *CR1* and *CR2*

For each pair of correspondent contours, sub-pixel accuracy pairs of points are generated. For a left contour point *PL*, correspondent PR is at the intersection of the epipolar line with the piecewise linear approximation of the right contour (Fig. 10).

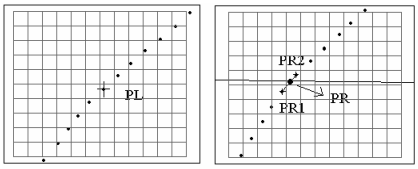

Fig. 10. A point *PL* on the left contour (a), and (b) the intersection *PR* of the right contour and the epipolar line

Strongly-curved basic contours usually present irregular sub-pixel shapes, and the accuracy obtained by the above-described method can be less than the one of standard correlation. The classical method is used for these basic contours (area-based correlation plus parabola fitting).

Each left-right correspondent pairs of points is mapped into the 3D world by computing the intersection of the two projection rays associated with the points [1].

4 Lane Detection

Lane detection, detailed in [9], is integrated into a tracking process. The current lane parameters are predicted using the past parameters and the vehicle dynamics, and this prediction provides search regions for the current detection.

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 (Fig. 11). From all the 3D points only the ones that project inside the predicted search regions are processed.

Fig. 11. Side (lateral) view of the 3D points inside the search region

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). 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.

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 detection results are used to update the lane state through the equations of the Kalman filter.

5 Implementation

The platform used was a Pentium 4 processor at 2.6 GHz, 512 MB memory, running the Windows 2000 operating system. High quality digital grayscale cameras and a grabber were used for synchronous stereo image acquisition. The image resolution was 644x512 pixels. The stereo algorithm was implemented in C++, with MMX/SSE processor optimizations used for the correlation function and for the sub-pixel contour extraction. The stereo algorithm runs in real time, varying from 13 to 18 frames per second.

6 Evaluation methodology

To provide a quantitative evaluation of the improvements brought by the proposed algorithm, synthetic stereo images were generated with various 3D lane configurations. Real world textures were used for the road surface and anti-aliasing was used for realistic projection of edges. Stereo geometry and optics of the virtual cameras were the same as for the real system. Four lane configurations were used (Fig. 12): vertical curvatures of 1500 meters and 3000 meters (V1500 and V3000), both with the pitch angle of 1 degree, and horizontal curvatures of 400 and 600 meters (H400 and H600).

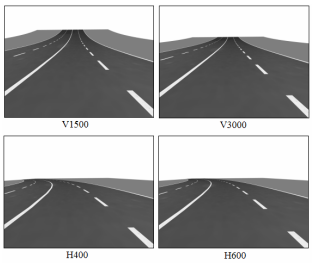

Fig. 12. Synthetic configurations used for evaluation

A standard canonical stereo reconstruction engine was used for comparison. It reconstructs edge points optimally detected with a gradient-based Canny edge detector. Rectification is used to get a canonical configuration. Parabola is fitted to the (area-based SAD) matching cost for sub-pixel accuracy.

The Canny edge detector and our contour detector were tuned so they extracted the same set of edge points at pixel level. This is necessary to avoid differences in the lane modelling due to different sets of reconstructed points. The only difference is the sub-pixel accuracy of the contour detector.

The contour-based stereo performed much better on horizontally curved roads (Fig.13). It was able to reconstruct more accurate 3D points in the depth interval 60 to 90 meters.

The following parameters of the 3D lane were considered for evaluation: width W, horizontal curvature radius c_h, vertical curvature radius c_v, and the pitch, yaw, and roll angles α, γ, ψ.

TABLE 3

W(m)	V1500	V3000	H400	H600
Canonical	3.06	3.07	3.05	3.11
Contour	2.91	2.90	2.92	2.91
True value	2.90			

TABLE 4

c_h (m)	V1500	V3000	H400	H600
Canonical	4500	5000	310	460
Contour	11000	8500	435	615
True value	+∞	+∞	400	600

TABLE 5

c_v (m)	V1500	V3000	H400	H600
Canonical	1750	3290	9300	8800
Contour	1650	3320	9700	8450
True value	1500	3000	+∞	+∞

TABLE 6

α (deg)	V1500	V3000	H400	H600
Canonical	0.96	1.12	0.02	0.03
Contour	0.99	1.05	0.01	0.03
True value	1		0	

TABLE 7

γ (deg)	V1500	V3000	H400	H600
Canonical	0.66	0.4	0.74	0.55
Contour	0.14	0.17	0.11	0.10
True value	0			

TABLE 8

ψ (deg)	V1500	V3000	H400	H600
Canonical	0.08	0.06	0.07	0.08
Contour	0.09	0.05	0.04	0.06
True value	0			

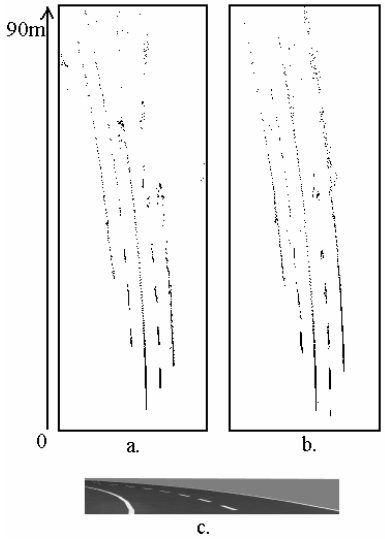

Fig. 13. A top-view (bird eye) of the set of reconstructed points with canonical stereo (a) and the contour-based stereo (b). The scenario H400 is used for reconstruction. Canonical stereo has difficulties for features between 60 and 90 meters (image patch shown in c).

Minor improvements are noticed for the roll and pitch angles (Tables 6 and 8). Medium improvements are present for the estimation of the lane's width and the vertical curvature radius (Tables 3 and 5).

The contour-based stereo greatly improves the estimation of the horizontal curvature radius and the yaw angle (Tables 4 and 7). For canonical stereo the 3D model of the lane is fitted to inaccurate and/or

incomplete 3D data in the far-distance range (60 to 90 meters).

Same improvements were seen with real test images (Fig. 14) especially for curved lanes.

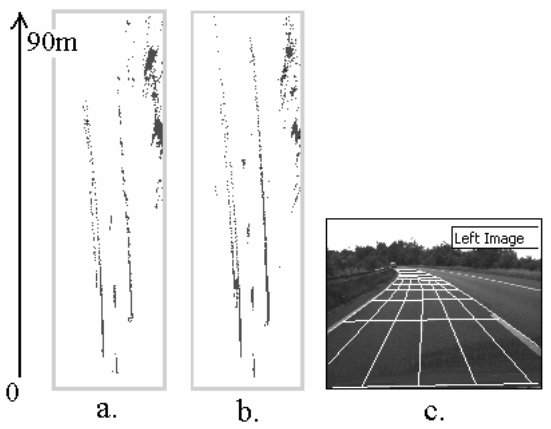

Fig. 14. Top-view of the 3D points on real images: from 60 to 90 meters, the contour based stereo (b) performs much better than the canonical stereo (a). The left image and the projection of the 3D lane is shown in c

7 Conclusions

A novel idea for 3D stereo reconstruction has been presented. Edge points are detected with sub-pixel accuracy as contours in both left and right images, and correspondences are computed between the sub-pixel accuracy edges, with increased precision. The new algorithm was implemented and integrated in a stereo acquisition and processing framework. It runs in real time and provides increased 3D reconstruction accuracy.

The increased accuracy can prove valuable for applications that require an accurate 3D reconstruction and description of the highway-driving scenario: collision avoidance, lane following and departure. An accurate environment description is required especially for high speeds of the ego vehicle. A future extension can consist of speed improvements by implementing time-consuming parts on dedicated hardware (ex. FPGA).

References

[1] E. Trucco, A. Verri , "Introductory techniques for 3D Computer Vision", Prentice Hall, 1998.

[2] J. Ramesh, R. Kasturi, B. G. Schunk, *Machine Vision*, McGraw-Hill, Inc., 1996.

[3] S. T. Barnard, M. A. Fischler, "Computational stereo", *Computing Surveys* 14(4), 553-572, 1982.

[4] R. Szeliski, D. Scharstein, "Symmetric Sub-Pixel Stereo Matching", *7th European Conference on Computer Vision,* Copenhagen, Denmark, volume 2, pp. 525–540, May 2002.

[5] T. A. Williamson, "A High-Performance Stereo Vision System for Obstacle Detection", technical report CMU-RI-TR-98-24, Robotics Institute, Carnegie Mellon University, September, 1998.

[6] K. Konolige, L. Iocchi, "A Multiresolution Stereo Vision System for Mobile Robots", *AIIA (Italian AI Association) Workshop*, Padova, Italy, 1998.

[7] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler, "Autonomous driving approaches downtown", *IEEE Intelligent Systems*, vol.13, no. 6, pp. 40-48, 1998.

[8] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, R. Schmidt, T. Graf, "High Accuracy Stereo Vision System for Far Distance Obstacle Detection", *IEEE Intelligent Vehicles Symposium* (IV2004), Parma, Italy, 292-297, 2004.

[9] S. Nedevschi, R..Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, "3D Lane Detection System Based on Stereovision", *IEEE Intelligent Transportation Systems Conference (ITSC)*, Washington, USA, pp.161-166, 2004.

[10] H. Pan, "General Stereo Image Matching Using Symmetric Complex Wavelets", *Technology Park*, Adelaide, Australia, 1996.

[11] P. Rockett, "The Accuracy of Sub-Pixel Localisation in the Canny Edge Detector", *The British Machine Vision Conference,* Nottingham, UK, pp. 392-401, 1999.

[12] S. Nedevschi, C. Goina, "Edge extraction and contour closing with subpixel accuracy". A.C.A.M. Sci. Journal, 1(2):75-86, 1992.

[13] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 679-698, Nov. 1983.

[14] C. Baillard, C. Schmid, A. Zisserman, A. Fitzgibbon, "Automatic line matching and 3D reconstruction of buildings from multiple views", *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, IAPRS Vol.32, Part 3-2W5, 1999.