Stop-line Detection and Localization Method for Intersection Scenarios

Tiberiu Marita, Mihai Negru, Radu Danescu, Sergiu Nedevschi
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
{Tiberiu.Marita, Mihai Negru, Radu.Danescu, Sergiu.Nedevschi}@cs.utcluj.ro

Abstract — **This paper presents a method for the detection and localization of stop-lines and other horizontal road markings as bicycle or pedestrian crossings encountered in road environments. The few existing stop-line detection approaches found in literature are based on monocular vision, and therefore cannot infer high accuracy information from the visual data. The proposed solution uses a hybrid approach that combines 2D detections in grayscale images with stereo-vision based 3D validations in order to increase the robustness and accuracy of the results. Model based reasoning is also used to eliminate false positive situations and to detect exactly the objects of interest (stop-lines, pedestrian or bicycle crossings). The detected horizontal road markings can be used as landmarks for a very accurate longitudinal localization of the ego vehicle within intersection scenarios compensating missing GPS data or its lack of accuracy.**

Keywords - environment perception; stereovision sensor; 3D localization; feature extraction; model-based classification

I. INTRODUCTION

Intersection scenarios provide the most complex traffic situations for driving assistance systems. Because they are both confusing and potentially dangerous, the research community has invested significant efforts to find solutions that can increase the safety of the traffic participants, to prevent and reduce the number of accidents and fatalities in intersections. Projects like INTERSAFE (part of the PReVENT project www.prevent-ip.org) and INTERSAFE-2 (www.intersafe-2.eu) had the goal to implement systems able to generate a complete and accurate description of the complex intersection environment based on state of the art passive and active on-board sensors, navigation maps, information from other traffic participants and from intelligent infrastructure.

An important task in these scenarios is the accurate ego-vehicle localization relative to the intersection. Although GPS based digital maps, with static or even dynamic content have been the solution world-wide adopted for navigation both in urban and non-urban environments, the conventional GPS signal is affected by various factors that introduce significant positioning errors (between 1 and 30 meters), and thus a more accurate localization of the ego-vehicle on a digital map is necessary. This can be obtained by sensorial localization of some road landmarks relative to the ego car.

In this paper a very accurate method for the longitudinal localization of the ego vehicle relative to an intersection is proposed. It is based on the detection and localization of some specific intersection landmarks: primarily stop-lines or wait-lines and secondarily other transversal road markings as pedestrian and bicycle crossings.

Vision sensors are extensively used in the detection of the painted road markings since they provide the highest spatial resolution compared to other sensors as laser scanners or LIDARS. Approaches based on monocular vision can provide accurate 2D detection of the road markings but they cannot provide alone localization information. In [1] a stop-line detection algorithm is presented that uses the Canny edge detector to search each image for properly oriented pairs of edges, one a transition from dark road to white stop line paint and its pair, a transition from white paint back to road. In [2] painted road signs are segmented in the IPM image and are recognized using a neural network classifier applied on the extracted features. In [3] a complete solution for detection, measurement and classification of painted road objects that exhibit mainly vertical/longitudinal features as arrows, lane delimiters and zebras is presented. The road markings are segmented in the 2D grayscale image while the 3D information is inferred from the dense-stereo map provided by a stereo-vision sensor [4]. The road markings are finally classified using a decision tree based classifier.

The proposed solution for detecting stop-lines, wait-lines, pedestrian and bicycle crossings combines the detection of horizontal features in the 2D image followed by 3D validation/classification. The detection of the 2D image features is based on the Hough transform which is used to detect only horizontal lines from the edge map of a grayscale image. These horizontal lines can belong to different classes of objects, as barriers [5], road features (transversal road markings, curbs etc), or other unclassifiable horizontal structures. Therefore, a series of validations steps based on primary dense 3D data provided by a dense stereo sensor [4], road surface parameters provided by a lane detection module [7] and model based reasoning applied on the markings' 2D shape and 3D size are performed for an accurate detection and localization. The identified transversal road markings are reported as flat 3D cuboids which are further tracked using the odometry data of the ego car.

II. PROBLEM STATEMENT

Most of the horizontal road markings (stop-lines, wait-lines, and pedestrian or bicycle crossings – Fig. 1.a) exhibit few 3D reconstructed points when using a dense stereo vision engine with the two cameras displaced horizontally. This can be explained by the impossibility to correlate areas with uniform intensity/color (i.e. no texture) along the horizontal direction (Fig. 1.b). 3D points' density increases only at the side ends of the horizontal road markings where edge like intensity variations are better handled by the most of the stereo reconstruction engines [7]. Accordingly, the accuracy of the 3D points along horizontal features is questionable due to the correlation uncertainties (Fig. 1.c)
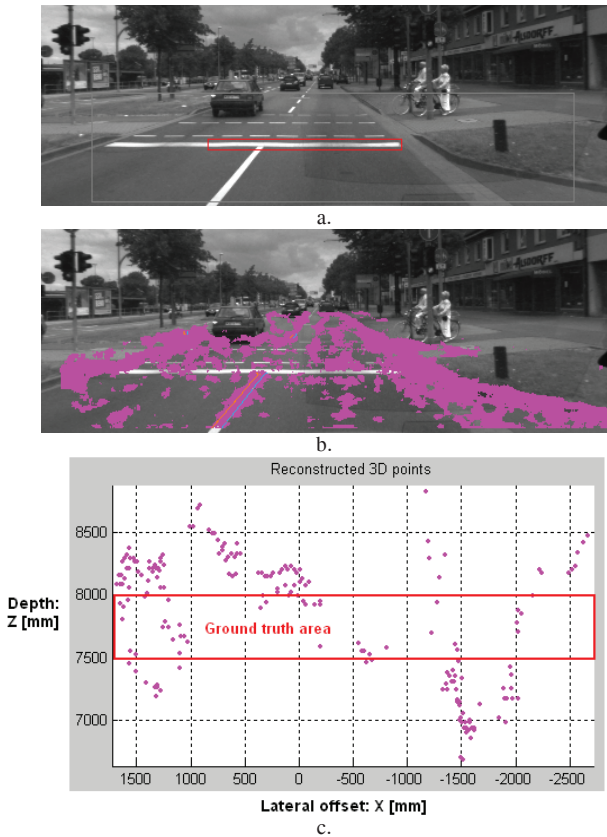

Figure 1. a. Typical urban scenario with stop line and crossings; b.Reconstructed road points; c. Top view of the 3D road points reconstructed on the stop-line highlighted in Fig 1.a (red box). Ground truth is about 7750 mm for the depth and 500 mm for the thickness of the stop-line.

The proposed solution uses a hybrid approach by combining the 2D detection based on grayscale image features and 3D validations based on dense stereo data in order to increase the robustness of the detection. Model based reasoning uses the shape, proportions and size features of the objects of interest in order to eliminate false positive situations (e.g. the scenario from Fig. 2 with a curb of a transversal road) and to detect only stop/wait lines, pedestrian and bicycle crossings. False negative situations which can occur due to obstructions or missing 3D information are compensated by a tracking algorithm based on the ego-vehicle odometry.
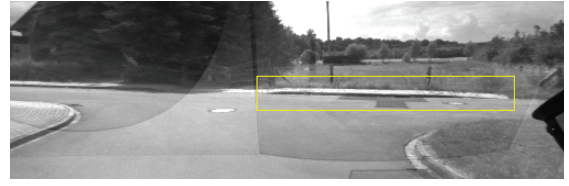

Figure 2. False positives can be generated by other horizontal structures.

III. STOP-LINE DETECTION AND 3D LOCALIZATION ALHORITHM

The main steps of the horizontal road markings detection and localization algorithm are described bellow:

A. Search space reduction

One of the main constraints of the algorithm is the real time capabilities in the context that it must be integrated into a more complex system that performs a complete description of the traffic environment with a 20 fps speed. Therefore the limitation of the search space of both 2D and 3D features was compulsory by introducing *Regions of Interest* (ROIs):

- a 3D ROI used for the 3D validation is specified (depth and lateral offsets). If a current lane exists, the lateral offsets are adjusted according to the lane limits.

- a 2D ROI used for the 2D detection is computed automatically based on the camera setup (pitch angle, nearest blind spot, maximum detection range.

The two ROIs are permanently merged (intersected) in the detection/validation process.

B. Detection of horizontal line structures in 2D

Based on quasi-horizontal edges, the Hough transform is applied to detect horizontal lines. The Hough lines are limited to near horizontal angles to reduce the processing time. Closely detected Hough lines (belonging to the same horizontal structure/marking) are grouped based on vicinity and orientation criteria [5].

C. Hypotheses generation

This step generates stop/wait line or crossing hypotheses. A 2D Hough line is validated if it neighbors pixels that are image projections of 3D road points [5]. From the neighboring 3D points a corresponding 3D line is generated. Close 3D lines are grouped (clustered) based on their vicinity in the YOZ space (Fig. 3) using the MBSAS algorithm [8]. Each cluster will be a stop-line or crossing hypothesis.
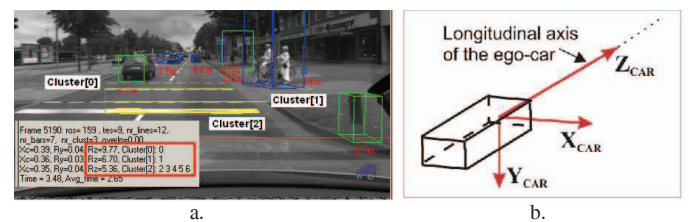

Figure 3. a. Clustering of the 3D horizontal lines in the YOZ space. b. Ego car coordinate system

D. Hypothesis validation

The 3D stop-line and crossing hypotheses obtained in the previous step are further validated in 2D.

1) Local 2D ROI generation

First a local 2D ROI is generated for each 3D stop-line or crossing hypothesis by projecting it on the image. This is obtained by projecting each cluster characterized by a height Y a depth Z and lateral offsets X_L and X_R on to the 2D image. Lateral offsets of the 3D clusters are restricted inside the current lane (if exists) or to about ± 2.5m relative to ego-car longitudinal axis. The height of each 2D ROI is set to a value equal with two times the expected stop-line/crossing image projection thickness plus a tolerance ε in order to cope with slightly deviations of the 2D image of the stop-line or crossing from the horizontal. The expected stop-line or crossing image projection thickness SLT_{2D} is adjusted dynamically as a function of the depth (Z) of the 3D hypothesis, camera focal length (F_Y), camera height (T_{CY}), camera pitch angle (R_{CX}) and the real 3D stop line thickness (SLT):

$$SLT_{2D} = F_Y\left[tg\left(tg^{-1}\frac{T_{CY}}{Z - SLT/2} - R_{CX}\right) - tg\left(tg^{-1}\frac{T_{CY}}{Z + SLT/2} - R_{CX}\right)\right] \quad (1)$$

2) Local 2D ROI analysis

Valid stop-line cells *(SL-cells)* are searched inside each local 2D-ROI from left to right. A valid SL-cell is a column vector of grayscale levels that obeys the following pattern: black-to-white transition followed downside by a white-to-black transition (Fig. 4). The distance between the 2 transitions should be around the expected image projection thickness (SLT_{2D}) corresponding to its depth Z for a stop-line or wait-line, around $SLT_{2D}/2$ for a bicycle crossing and around $SLT_{2D}/4$ for a pedestrian crossing (see models from Table 1). For each SL-cell its 3D position is computed based on neighboring 3D points given by the dense stereo-engine.
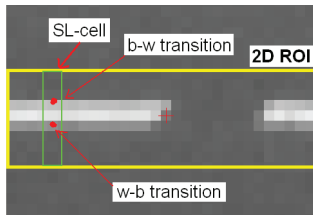

Figure 4. The SL-cells search process

For a local 2D-ROI a feature vector containing the average 2D thickness *W* of the containing SL-cells and the horizontal binary pattern vector of the SL-cells ('1' for a valid SL-cell and '0' for an invalid one) is built and its lateral limits are refined (Fig. 5.a). The horizontal binary pattern vector is filtered out by an adaptive median filter in order to eliminate the speckle type noise that can occur due to SL-cells detection errors. The filter size is adapted accordingly to the size and proportions of the image projection of the candidate stop-line or crossing.

An example of the horizontal binary pattern vector for a bicycle crossing at 10.37 m depth is presented in Fig. 5.b. Character '*' means a valid SL-cell while character '.' means a

gap cell. The length of continuous SL-cell segments and gaps are expressed in pixels and are given also numerically in the form of a binary *Run-Length-Encoding* (RLE). Also the average lengths of the filled segments and gaps are given as they are expected at the given depth and measured, along with the total fill factor of the horizontal binary pattern vector.
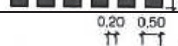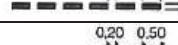

a.
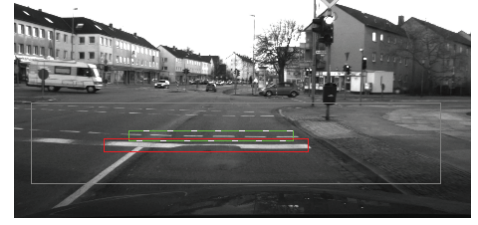
The horizontal binary pattern vector:
***************.....***************.....***************.....**************......**************......

Corresponding RLE encoding: [16 5 16 5 16 5 15 6 15 6]

Average expected values [pixels]:
  Fill_segment_length: EF=15.3, Gap_segment_length: EG=6.1, Fill_factor: EFF = 0.71

Average measured values [pixels]:
  Fill_segment_length: MF=15.8, Gap_segment_length: MG =5.3, Fill_factor: MFF= 0.75

Total fill factor: TFF = 0.74

b.

Figure 5. a. Refined local 2D ROIs after the local analysis process; b. Example of features computed from the horizontal binary pattern of a bicycle crossing at 10.37 m depth

TABLE I. 3D MODELS OF THE STOP-LINE AND CROSSING MARKINGS (I.E. SPECIFIED BY GERMAN LEGISLATION [9])

Line Type	3D shape and measures [m]	Marking Type
Continuous line	0,50	Stop line
Dashed line (fill/gap ratio 2:1)	0,25 0,50 0,50	Wait line
Dashed line (fill/gap ratio 2.5:1)	0,20 0,50 0,12	Pedestrian crossing
Dashed line (fill/gap ratio 2.5:1)	0,20 0,50 0,25	Bicycle crossing

3) Model based 2D classification

By analyzing the size, shape and proportions of the detected feature vector against country specific models (i.e. models from Table I) the hypotheses are classified as: stop/wait lines, crossings, transversal lane delimiters (others then stop lines) and unknown structures. The classification process uses simple decision-tree based rules. An example of some rules used in the decision tree classifier is given bellow, where the meaning of the used variables is explained in step *D.2* of the algorithm. All used thresholds were set experimentally based on the models from Table I and are customizable.

If (TFF > 0.85) and (W≈SLT_{2D})
  then class=stop-line;
If (TFF > 0.85) and (W≤SLT_{2D}/2)
  then class=continuous-line;
If (0.5 <TFF< 0.85) and (W≈SLT_{2D}) and (EF≈MF) and (EG≈MG)
  then class=wait-line;
If (0.5 <TFF< 0.85) and (W≤SLT_{2D}/2) and (EF≈MF) and (EG≈MG)
  then class=crossing;
Otherwise: class=unknown;

In the 3D objects generation step (next step *E*) only the stop/wait lines and crossings are taken into account (Fig. 6). Stop-lines and wait-lines are reported as a merged unique class labeled *"STOP-LINE"* while pedestrian and bicycle crossings are reported as a unique *"CROSSING"* class.
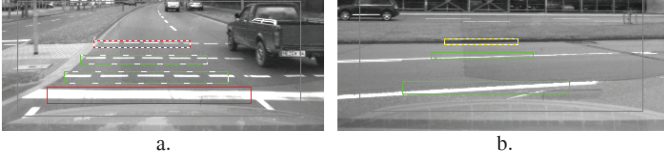

Figure 6.   a. Only stop/wait lines and crossings are reported (true positives); b. transversal lane delimiters and unknown structures (i.e. curbs) are rejected.

E. 3D objects generation

Based on the 3D data associated to the SL-cells, the 3D position, size and shape of each detected horizontal road structure is refined. For the validated hypotheses classified as *STOP-LINEs* and *CROSSINGs* corresponding 3D flat-box shape objects are generated. Only the nearest stop-line and/or bike-crossing objects are reported. If the application requires it, crossing objects can be ignored and only stop-lines are maintained in the generated 3D objects list (Fig. 7).
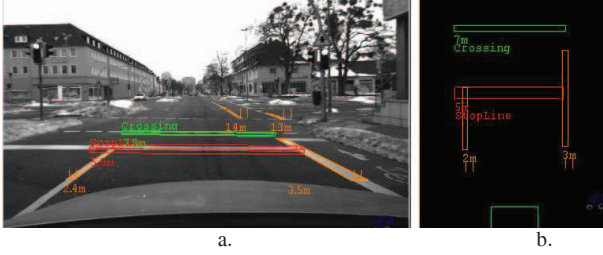

Figure 7.   Detected stop-line (flat red box), crossing (flat green box) and longitudinal lane markings (orange rectangles) [3] represented as 3D objects: a. perspective view; b. top view.

F. 3D objects tracking.

The position of the detected 3D objects is filtered over time using the odometry data of the ego car.

The basic odometry information available through the CAN bus of a modern car is the speed v and the yaw rate $\dot{\psi}$. Together with the time interval Δt elapsed between measurements *(t-1)* and *(t)*, these parameters can be used to compensate for the ego-motion, and separate it from the independent motion of the objects in the scene. Between measurements, the ego-vehicle rotates with an angle ψ, and travels a distance d.

$$\psi = \dot{\psi}\Delta t \tag{2}$$

$$d = \frac{2v\Delta t \sin\dfrac{\psi}{2}}{\psi} \tag{3}$$

Consequently the origin of the ego-vehicle's coordinate system suffers a displacement between points *o(t-1)* and *o(t)*

which is expressed by displacements along the two coordinate axes by d_X and d_Z:

$$d_X = d\sin\frac{\psi}{2} \tag{4}$$

$$d_Z = d\cos\frac{\psi}{2} \tag{5}$$

A point on the road, at position given by z and x (see the ego car coordinate system from Fig. 3.b), is displaced between moments of time *(t-1)* and *(t)* by the following equation:

$$\begin{bmatrix} x(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} x(t-1) \\ z(t-1) \end{bmatrix} - \begin{bmatrix} d_X \\ d_Z \end{bmatrix} \tag{6}$$

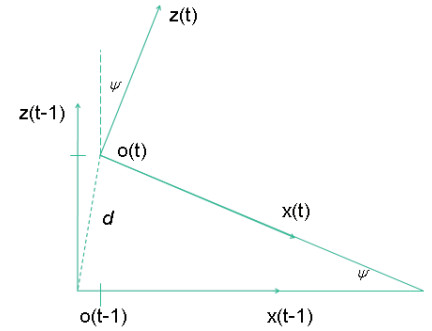Figure 8.   Displacement of the ego-car's origin between two consecutive measurements: *(t-1)* and *(t)*

IV. RESULTS

For the assessment of the stop–line detection and localization method both quantitative and qualitative evaluations were performed. The quantitative ones were focused to prove the precision of the measurements against some ground truth positions while the qualitative ones aimed to prove the accuracy in terms of true positive detections against false positives and false negatives.

For the quantitative evaluation a stereo vision system with 8.5 mm focal length lenses, 190 mm baseline and 42 degrees horizontal field of view was used. For this setup a reliable dense stereo reconstruction is possible up to 35 m in depth. Hence features based on quasi-vertical edges and their neighboring textures can be detected in the above mentioned range [10], detection of the stop lines is limited up to 18 … 20 m in the current camera configuration. Above this range the stop-line's image projection thickness is below 1 pixel and sub-pixel estimations of the black-to-white and white-to-black transitions are not possible. Pedestrian and bicycle crossings detection range is reduced up to 10 … 15 m due to their reduced thickness (compared to a stop-line).

The evaluation experiment was performed in a controlled road scenario (Fig, 9). The aim was to evaluate the accuracy of the longitudinal distance (depth) between the ego car and the detected stop-line. The car was driven perpendicular to the

stop-line in a depth range from 18 m down to 4 m with stops at every 2 m (see the plot of the estimated position of the stop line – Fig. 10). At each stop the estimated distance was averaged over 200 … 300 frames and the results are show in Table II (2-nd column). As it can be noticed, the relative depth errors are below 2% (absolute values errors are given for the relative errors).

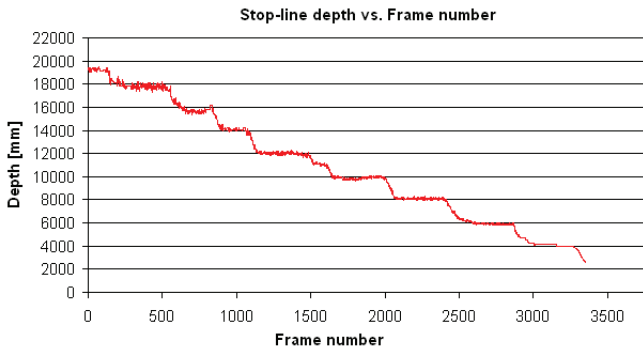Figure 9.   The experimental scenario used for the quantitative evaluation.

Figure 10. Plot of the estimated depth for a set of over 3000 measurements in a controled scenario.

TABLE II.   AVERAGE ERRORS AT THE REFERENCE DEPTHS

Ground truth [mm]	Estimated depth [mm]	Absolute error [mm]	Relative error [%]
4000	4065	65	1.63
6000	5913	-87	1.45
8000	8099	99	1.24
10000	9880	-120	1.20
12000	11983	-17	0.14
14000	14087	87	0.62
16000	15711	-289	1.81
18000	17831	-169	0.94

The qualitative evaluation of the method was performed by manual assessment of the number of true positives, false negatives and false positives occurrences of the detected sensorial landmarks (stop-lines and crossings). Real urban traffic scenarios were used for this test summing a total number of 12275 image frames, taken in Romania and Germany (with specific geometry constraints). The test images were acquired with several stereo system configurations (in terms of focal length, baseline length, HFOV). Some examples among the several test scenarios are shown in Fig.11.
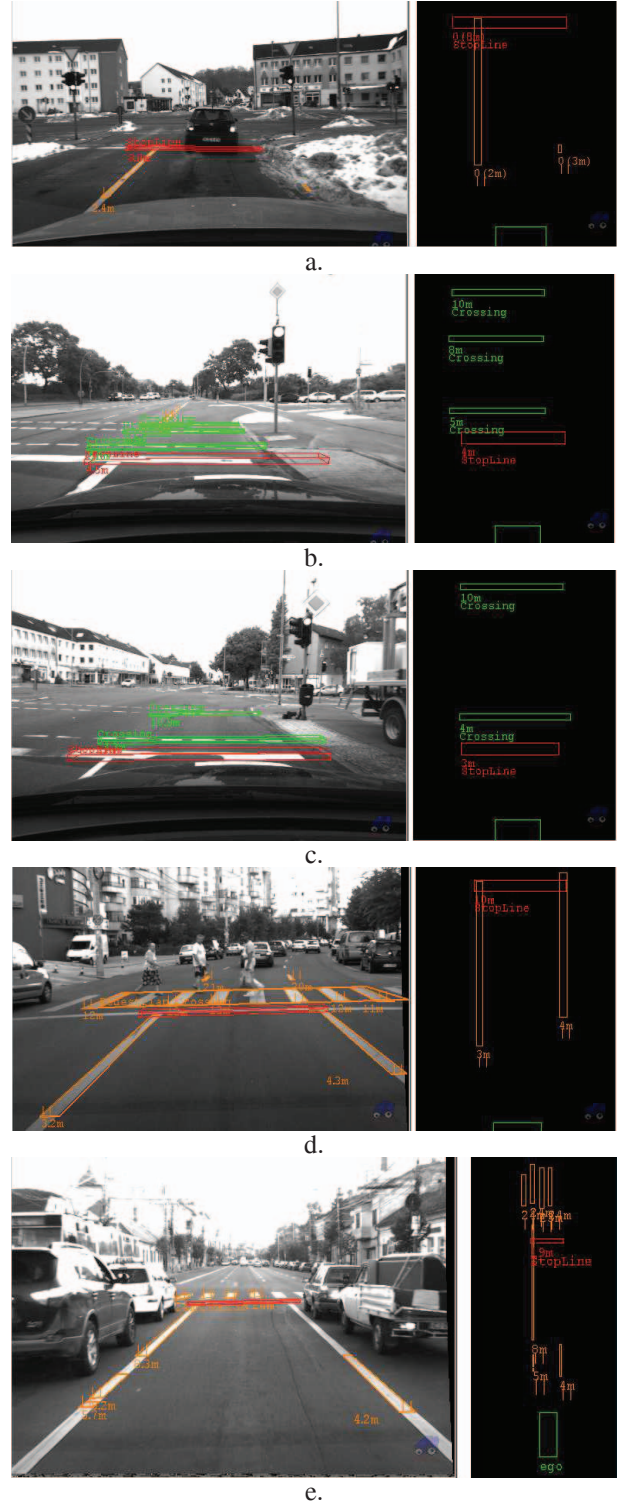

a.

b.

c.

d.

e.

Figure 11. Stop-lines and crossings detection examples in several test scenarios.

In Table III, the true positive rate (TPR) and the false positive rate (FPR) are shown for the detected landmarks. The false negative rates (FNR) can be deduced from the true positive rates by: FNR=100% - TPR, and is not shown.

Landmarks with more complex models (i.e. crossings) exhibit a very low FPR (due to model complexity which cannot match on artifacts), but on the other hand the TPR is lower due to the real traffic environment conditions. These marking can suffer due degradation and consequently validation conditions based on the horizontal pattern feature vector are not fulfilled. Moreover their image projection thickness estimation at high distances has a lower accuracy due to the impossibility to estimate it with sub-pixel precision which can cause additional classification or rejection errors. Landmarks with simple models (stop-lines) have high TPR but they are prone to be confused with other transversal road artifacts (FPR ≈ 1%). However, the sensorial stop-line detection can be over imposed by GPS-based map information, thus compensating the FPR.

TABLE III.   DETECTION RATES OF THE PAINTED ROAD LANDMARKS

Landmark type	TPR [%]	FPR [%]
Stop line	96.50	1.05
Crossing (pedestrian or bicycle)	84.89	0.00

The implementation of the algorithms was done by taking into account real time requirements. Repetitive and intensive computations as the trigonometric values used in the Hough transform or the dynamic estimation of the stop line image projection thickness were replaced by the use of look-up tables values computed off-line. The processing time depends on the size of the 2D-ROI which is dynamically adjusted at every frame. Consequently the number of the horizontal features detected in the 2D-ROI will influence the computation costs of the algorithms implied. Processing time was evaluated on an Intel E6750 processor using 512x383 resolution input images. In a normal scenario with true positives (a stop-line and up a 2 crossing lines) the average processing time is about 3 ms. As the number of horizontal Hough features increases inside the 2D-ROI the processing time can take up to 10 ms.

V.   CONCLUSIONS

In this paper a method for the detection and localization of stop-lines and other horizontal/transversal road markings as bicycle or pedestrian crossings encountered in road environments was presented. The method is based on 2D horizontal features detection using the Hough transform followed by a series of 2D and 3D validations in order to increase the detection robustness and the localization accuracy. The 2D local analysis of the horizontal features provides the proper selection of the interest objects as stop-lines, wait-lines, bicycle and pedestrian crossings while localization and size information is inferred from the 3D data provided by a dense stereo-engine.

The detected horizontal road objects are reported as flat, un-oriented, 3D objects and can be used for very accurate ego-vehicle localization within intersections. If used in conjunction with digital GPS maps, and other measurements performed with on-board or infrastructure sensors, this can provide an augmented static and dynamic description of the intersection scenarios, improving specific driving assistance applications.

The implementation of the method has real time capabilities providing an average processing time witch is less then 10 ms even in complex situations. The accuracy of the method was assessed by tests in both controlled and real traffic environments for quantitative and qualitative evaluations.

As future work an optimization of the method will be performed in order to handle oriented stop-line and crossing objects which can occur in some intersection scenarios where these markings are not any more transversal or perpendicular on the lane delimiters.

REFERENCES

[1] Miller, I., Campbell, M., Huttenlocher, D., and all: "Team Cornell's Skynet: Robust Perception and Planning in an Urban Environment", Journal of Field Robotics 25(8), 2008, pp 493–527.

[2] A. Kheyrollahi, T.P. Breckon, "Automatic Real-Time Road Marking Recognition Using A Feature Driven Approach", Machine Vision and Applications, DOI 10.1007/s00138-010-0289-5

[3] R. Danescu, S. Nedevschi, "Detection and Classification of Painted Road Objects for Intersection Assistance Applications", Proceedings of the IEEE Intelligent Transportation Systems Conference 2010 (ITSC 2010), September 2010, Madeira, Portugal, pp. 433-438.

[4] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, C. Pocol, S. Sobol, C. Tomiuc, C. Vancea, M.M. Meinecke, T. Graf, T. B. To, M.A. Obojski, "A Sensor for Urban Driving Assistance Systems Based on Dense Stereovision", Proceedings of 2007 IEEE Intelligent Vehicles Symposium, (IV2007), Istanbul, Turkey, June 13-15, 2006, pp 276-283,

[5] T.Marita, "Barriers Detection Method for Stereovision-Based ACC Systems", Proceedings of the 5-th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP2009), Cluj-Napoca, Romania, August 27-29, 2009, pp. 95-102.

[6] R. Danescu, S. Nedevschi, M.M. Meinecke, T.B. To, "Lane Geometry Estimation in Urban Environments Using a Stereovision System", Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC 2007), Seattle, USA, 2007, pp. 271-276.

[7] TYZX Inc, DeepSea V2 Stereo Processor, http://tyzx.com/products/DeepSeaV2.html, (cited April 2011).

[8] Bundesministerium für Verkehr, HAV 11. Auflage, 7. Fahrbahnmarkierungen-Landstrassen, BMV Erlass von 23.08.1993, pp. 365.

[9] Theodoridis, K. Koutroumbas, Pattern Recognition, 2-nd Edition, Academic Press, 2003.

[10] S. Nedevschi, T. Marita, R. Danescu, F. Oniga, S. Bota, "On-board Stereo Sensor for Intersection Driving Assistance. Architecture and Specification", Proceedings of the 5-th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP2009), Cluj-Napoca, Romania, August 27-29, 2009, pp. 409-416.