Geographic Information for Vision–based Road Detection

Anonymous Submission

***Abstract*— Road detection is a vital task for the development of autonomous vehicles. The knowledge of the free road surface ahead of the target vehicle can be used for autonomous driving, road departure warning, as well as to support advanced driver assistance systems like vehicle or pedestrian detection. Using vision to detect the road has several advantages in front of other sensors: richness of features, easy integration, low cost or low power consumption. Common vision–based road detection approaches use low–level features (such as color or texture) as visual cues to group pixels exhibiting similar properties. However, it is difficult to foresee a perfect clustering algorithm since roads are in outdoor scenarios being imaged from a mobile platform. In this paper, we propose a novel high–level approach to vision–based road detection based on geographical information. The key idea of the algorithm is exploiting geographical information to provide a rough detection of the road. Then, this segmentation is refined at low–level using color information to provide the final result. The results presented show the validity of our approach.**

I. INTRODUCTION

Advanced driver assistance systems (ADAS) aim to understand the environment of the vehicle to the extent of contributing to traffic safety. The ability to detect the road ahead the target vehicle is an important task to develop these assistance systems. Road detection is a challenging task for machine perception since the system must be able to learn in a self–supervised manner and adapt to inter- and intra–class changes in the local environment. During the last decade, a large number of road following systems have been developed using either active sensors (i.e., laser and radar), or passive ones (i.e., monocular or stereo cameras). Using a single camera capturing scene information from the front windshield of the vehicle has several advantages: low cost, richness of features (color, texture), easy aesthetic integration and non–intrusive nature. However, detecting the road using a monocular vision–system is very challenging since the road is an outdoor scenario imaged from a mobile platform. Thus, the detection algorithm must be able to deal with continuously changing background, the presence of different objects (vehicles, pedestrian), different environments (urban, highways, off–road), different road types (shape, color), and different imaging conditions (varying illumination, different viewpoints and weather conditions).

Current road detection approaches using monocular vision systems use pixel features such as texture or color as visual cues for grouping pixels in two different groups: drivable road or background. However, algorithms based on these visual cues fail under wide lighting variations (penumbra, strong cast shadows and highlights or reflections among others) and these algorithms show dependency on highly structured roads, road homogeneity, simplified road shapes, and idealized lighting conditions. Thus, the performance of these systems is often improved by including temporal or road shape constraints. For instance, Michalke *et al.* [1] include temporal coherence averaging the results of consecutive frames in an image sequence. Sotelo *et al.* [2] include road shape restrictions by modelling the road using a second order curve. A similar example is found in [3] where color information is combined with edge information estimated based on synthetic road curvature models. Road shape is also included in [4] where road detection results using texture are classified as one of three different models (left turn, right turn and straight road). However, all these approaches use low–level properties to estimate the shape of the road ahead the target vehicle. That is, road features are extracted at a pixel level and grouped accordingly.

Our interest in this paper is estimating the geometry of the road in front of the target vehicle. Road geometry has been previously computed from images in [5]. Alvarez *et al.* use a scene classifier to provide the probability that a road image contains certain road geometry (left turn, straight, t–like junction). However, road geometries are learned off–line using training images. Thus, the algorithm is limited to a finite number of classes and the road shape is limited to the shapes in the training images.

Fig. 1. Geographic Information Systems store the information necessary to build a top–view map of specific areas of the Earth's surface. The image shows a satellite image of an urban area with superimposed road information.

In this paper, as a novelty, geographical information is used to infer the geometry of the road ahead the target vehicle. Geographical information refers to merging cartography and database technology. Geographical information systems (GIS) are often associated with a map that shows features and feature relationships on the Earth's surface (Fig. 1). Recently, the wide availability of high–quality location information

, mail

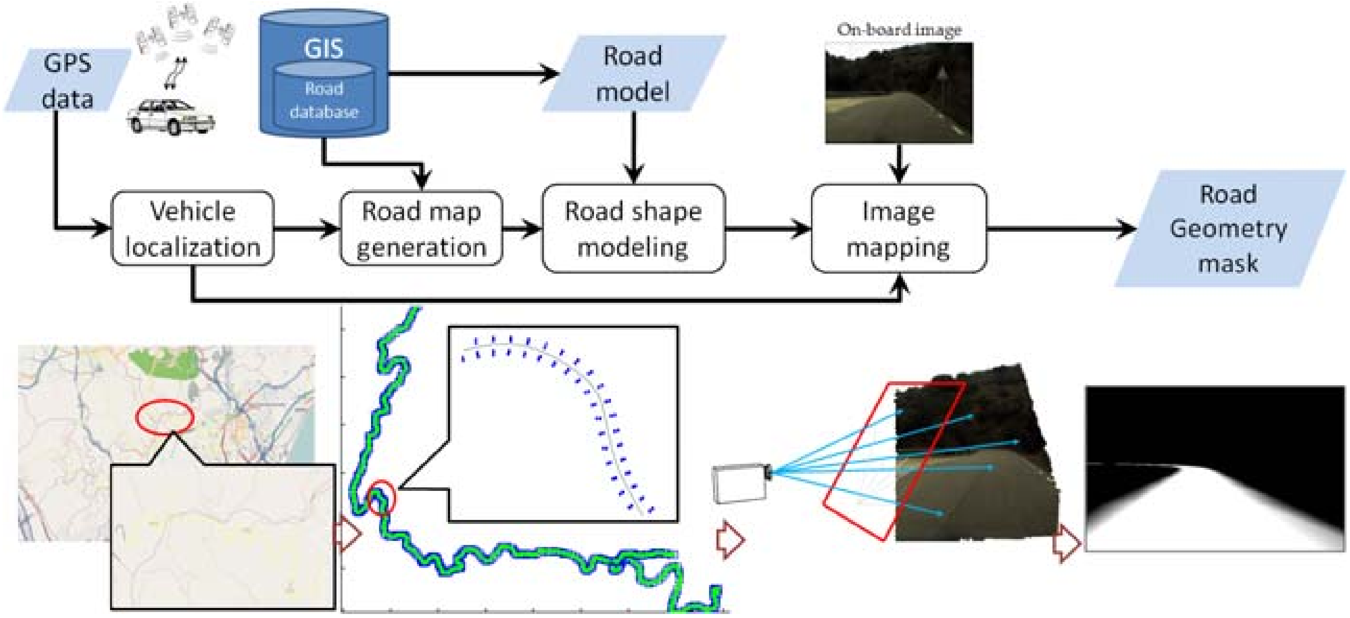

Fig. 2. Block–diagram of the algorithm for computing the road geometry using geographical information.

combined with affordable global positioning system (GPS) receivers has enabled mass–market mapping systems. These systems are widely used for vehicle navigation [6], [7], [8] or visual guidance using augmented reality [9]. Vehicle navigation refers to localizing the vehicle within a map. Visual guidance using augmented reality refers to superimposing traffic signal information into on–board images. Hence, none of these methods can be used for detecting the drivable area ahead the target vehicle. Therefore, in this paper, geographic information is exploited using a different approach. The key idea of the algorithm is detecting the road in images by projecting road map information provided by the GIS into the image plane of the on–board camera. The output of the algorithm is a rough segmentation of the road that is refined using pixel–level properties to provide the required accuracy.

The rest of this paper is organized as follows. First, in Sect. II the algorithm to estimate road shape from geographic information is outlined. Then, in Sect. III, experiments validating the algorithm are presented and results are discussed. Finally, conclusions are drawn in Sect. IV.

II. ROAD GEOMETRY FROM GEOGRAPHICAL INFORMATION

Road geometry refers to the shape of the road ahead the target vehicle. Examples of geometries are left turn, right turn or t-like junction. This contextual information is crucial for reliable extraction of image regions such as the road. The key idea for extracting the road geometry from geographic information systems is projecting the road information in the database into the image plane of the on–board camera. The algorithm relies in the information provided by Geographic Information Systems (GIS). GIS are database systems that capture, store and manage geographically referenced information. That is, data which is linked to location. This information describes the world in geographic terms representing objects like rivers, lakes or roads) using simple geometries such as points, polylines or polygons. The proper combination of these geometries creates a bird's view map of an specific region of the Earth (Fig. 1). Road information is usually associated with additional attributes such as road name, construction level, direction information and number of lanes in each direction.

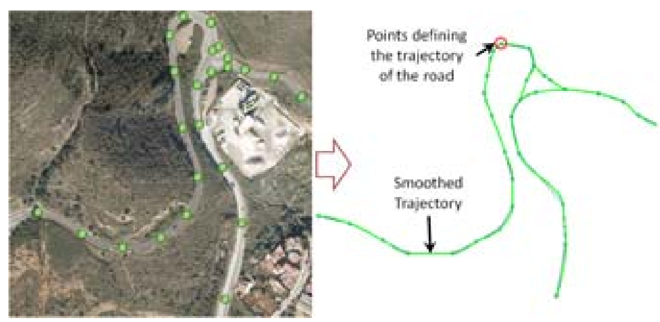

Fig. 3. Left: bird's view of the satellite road map with superimposed green points defining the road skeleton. Right: the road skeleton is obtained smoothing the road trajectory using cubic interpolation.

The algorithm devised for detecting the road using geographical information is shown in Fig. 2. The algorithm is divided in four main blocks: vehicle localization and navigation, bird's view map generation, road shape modelling and image mapping.

The first block is a GPS–based localization process and its goal is twofold. First, it provides the current position p of the vehicle relative to the map database. Second, consecutive positions of the vehicle are used to estimate the vehicle orientation in the current frame.

The second block consists in generating the road skeleton map of the area around the current position. The road map is modelled as junctions connected by piecewise continuous lines. Gaps between consecutive points are filled using cubic interpolation. As a result, a smoothed trajectory is obtained (Fig. 3).

The third block consists in modelling the road expading the road skeleton based on a road model. The road model consists of a drivable area per lane and two roadsides (Fig. 4). The width of these two parts (road lane W_{RL} and roadside R_{RS}) is estimated according the road attributes in the database and additional country's national road legislation.

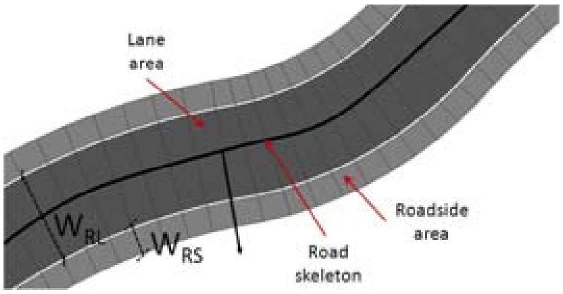

Fig. 4. Road shape model. The model consists in defining the width of the lane (W_{RL}) and the width of two roadsides (W_{RS}) perpendicular to the direction of each road segment defining the road skeleton.

Finally, the last block is the image mapping. In this process, the road geometry ahead the vehicle is obtained projecting the road map at the vehicle position onto the 2D driver's view.

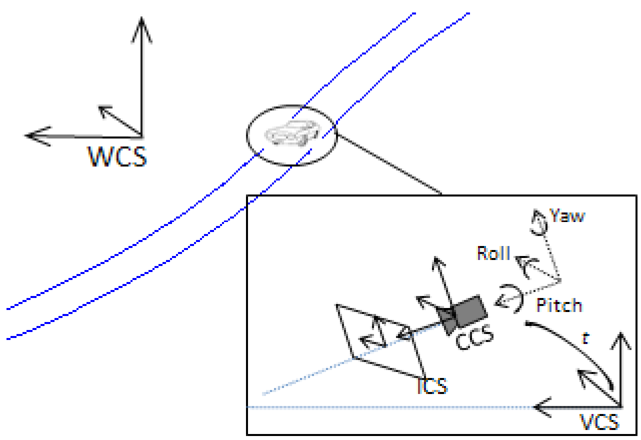

Fig. 5. Four coordinate systems are needed for mapping the road information onto the image plane of the camera: World (WCS), Vehicle (VCS), Camera (CCS) and Image (ICS) Coordinate Systems. The position of the camera is defined using a translation t and three rotations about coordinate axis (roll, pitch and yaw).

There are different coordinate systems involved in the mapping process: World Coordinate System (WCS), Vehicle Coordinate System (VCS), Camera Coordinate System (CCS) and Image Coordinate System (ICS). The relationship between these coordinate systems is shown in Fig. 5. Accordingly, the mapping process consists in transferring the road model from VCS to the ICS, using WCS as reference. This process can be decomposed in a rigid body translation and rotation between VCS and CCS, and a perspective projection from CCS to ICS. Then, a set of N points $\mathbf{R} = [r_0, r_i, \ldots, r_N]$ defining a road segment is mapped from VCS to ICS as follows,

$$\kappa p_i = K \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_{ix} \\ r_{iy} \\ r_{iz} \\ 1 \end{bmatrix}, \tag{1}$$

where κ is an arbitrary scale factor, K represents the intrinsic parameters of the camera [10]. R and t are the extrinsic camera parameters which relate the camera and vehicle coordinate systems. The former, R, describes the orientation (pose) and the latter, $t = (t_x, t_y, t_z)$, describes the translation (position) of the camera system in a 3D world (Fig. 5).

The intrinsic camera parameters are fixed for the particular device being used and define pixel coordinates of image points with respect to coordinates in the camera reference frame [10],

$$K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{bmatrix}, \tag{2}$$

where f is the focal length. That is, the distance from the optical center to the image plane. Finally, the rotation camera matrix is decomposed as follows,

$$\begin{aligned} R &= R_{pitch} R_{yaw} R_{roll}, \\ R_{pitch} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix}, \\ R_{yaw} &= \begin{bmatrix} cos\phi & 0 & sin\phi \\ 0 & 1 & 0 \\ -sin\phi & 0 & cos\phi \end{bmatrix}, \\ R_{roll} &= \begin{bmatrix} cos\varphi & -sin\varphi & 0 \\ sin\varphi & cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned}$$

where ϕ (roll), θ (pitch), φ (yaw) refer to rotations about the respective axis.

Figure 6 shows different road shapes obtained using the proposed algorithm. As shown, the shape of the road is properly recovered at different daytime and in different situations such as soft and hard turns. The smoothing present in the road mask is due to a smoothing factor in the roadside area.

III. EXPERIMENTS

Our proposal for vision–based road detection consists in refining the road detection results via geographical information using pixel–level features. Thus, geographic information provides a general view of the road ahead the vehicle and the pixel–based features provide the accuracy required. The algorithm is depicted in Fig. 8. Given an image and the GPS signal, the road is detected using GIS as described in Sect. II. Then, the result is refined to provide the final road mask. The refinement is performed using color information of the road pixels in the road geometry mask. Pixels are characterized using the hue component from the HSI (Hue,

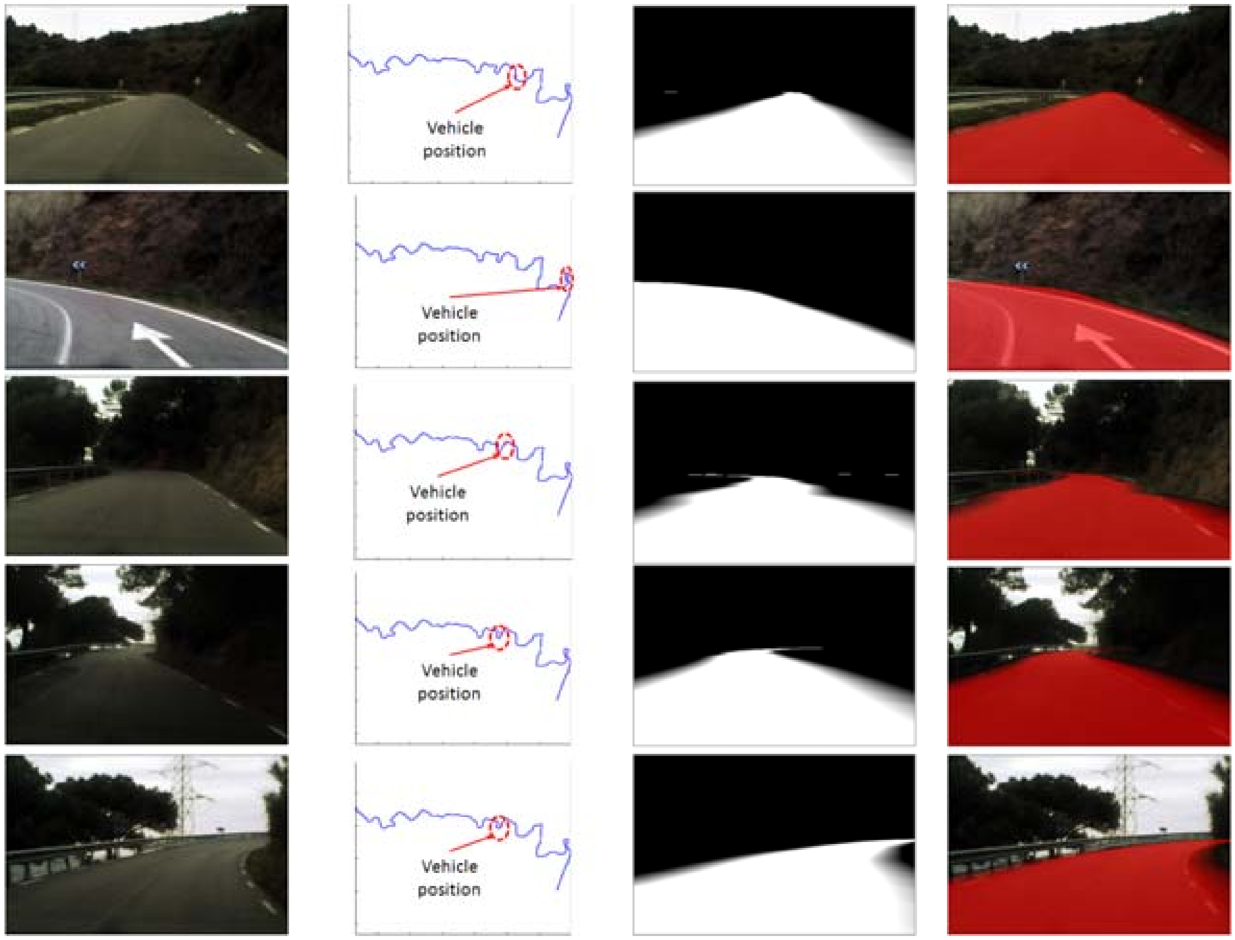

Fig. 6. Example results of road detection via geographical information. a) Input image. b) Bird's view of the road map and vehicle localization. c) Road geometry mask. d) Road geometry mask superimposed in the original image.

Saturation and Intensity) color space [10] and the chromaticity coordinates (u and v) of the perceptually uniform color space, CIELuv [11]. Hue describes color characteristics in terms of visual sensation according to perceived colors. That is, hue values are completely different between a red and a blue car. The CIELuv color space has been widely used in color image segmentation methods using clustering techniques [12]. Finally, the refinement consists in discarding those road pixels for which the Mahalanobis distance to a chromatic road model is larger than a fixed threshold. The chromatic road model is built using road pixels of the geographic mask in the bottom part of the image. Hence, the algorithm assumes the bottom part of the image belongs to the road surface. This is a common assumption [2], [13] since this region of the image corresponds to a distance of 2 meters in front of the target vehicle. Thus, the assumption is not restrictive if the vehicle keeps a gap with preceding ones that is safe for driving.

Experiments validating the algorithm are conducted on images acquired using an onboard camera based on the Micron MT9V023 sensor. This is a high dynamic range CMOS sensor of 752×480 pixels and 10 bits per pixels. The camera is equipped with a 6mm focal length microlens. The sensor uses Bayer pattern for capturing color information. Standard Bayer pattern decoding (bilinear interpolation) is used to obtain a 3–channels color image (RGB) of 752×480 pixels per channel and 10 bits per pixel.

Each acquired image is geo–referenced (synchronized) with GPS information (latitude, longitude and altitude) provided by a standard GPS antenna (Woxter Slim II) using NMEA protocol [14]. This protocol defines the interface between different electronic equipment. In this way, NMEA protocol sends a line of data called sentence that is totally self contained and independent from other sentences. Standard sentences for GPS devices have a two letter prefix (GP) which is followed by a three letter sequence that defines the sentence contents. For instance, GGA sentences provide 3D location and accuracy data, see Fig. 7. Each location (latitude and longitude) is converted to Cartesian coordinates (X, Y, Z) using the datum 84 (WGS–84). This geodesic datum is a reference used to describe the localization of points on the Earth's surface. The WGS–84 defines the Earth surface as a pole-flattened (oblate) spheroid, with major (transverse) radius $a = 6.378137$m at the equator, and minor (conjugate)

GPGGA,123519,4807.038,N,01131.000,E,1,08,0.9,545.4,M,46.9,M,,*47

Where:

GGA	Global Positioning System Fix Data
123519	Fix taken at 12:35:19 UTC
4807.038,N	Latitude 48° 07.038' N
01131.000,E	Longitude 11° 31.000' E
1	Fix quality: 0 = invalid, 1 = GPS fix (SPS) , 2 = DGPS fix , 3 = PPS fix, 4 = Real Time Kinematic, 5 = Float RTK, 6 =estimated, 7 = Manual input mode, 8 = Simulation mode
08	Number of satellites being tracked
0.9	Horizontal dilution of position
545.4,M	Altitude, Meters, above mean sea level
46.9,M	Height of geoid (mean sea level) above WGS84 ellipsoid
(empty field)	time in seconds since last DGPS update
(empty field)	DGPS station ID number
*47	Checksum data, always begins with *

Fig. 7. GGA NEMEA sentences provide 3D location (latitude, longitude and altitude).

radius $b = 6.356752,314$m at the poles.

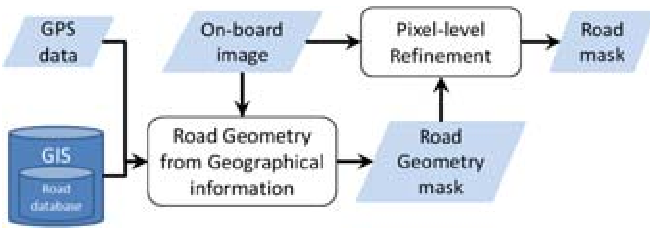

Fig. 8. Algorithm used to validate the vision–based road detection proposal.

The road database is obtained from OpenStreetMap [15]. OpenStreetMap is an opensource database containing geographical information and roads attributes in XML format [16]. These attributes comprise features such as type (motorway, path, trunk, primary road, secondary road), name, maximum speed, one/two ways [17]. Points describing road segments are interpolated using cubic interpolation to improve the resolution of the system. Finally, the road model is parameterized using equivalences in Table I. The rest of parameters of the algorithm are fixed empirically.

TABLE I

ROAD LANE AND ROADSIDE WIDTH EQUIVALENCES USED TO MODEL THE ROAD.

Road type	highway	primary	secondary	residential
W_{RL}	3.75m	3.5m	3.00m	3.00m
W_{RS}	2.5m	1m	1m	0.50m

Example results for different images are shown in Fig. 9. Images contain different road geometries and the presence of other vehicles in the scene. As shown, the road layout is properly recovered and objects are discarded using chromatic information. These results suggest that a reliable road segmentation algorithm is obtained by combining geographic information and a pixel–based refinement. Road surface is properly recovered most of the time despite lighting conditions and complex road shapes.

The analysis of failures reveals errors in the positioning process due to two main causes. The former is the inherent error in the accuracy of the GPS information. The latter occurs in urban areas where GPS signals are often blocked. These errors could be mitigated including navigation algorithms within the vehicle localization step and using Real Time Kinematic (RTK) satellite navigation instead of 'normal' GPS navigation.

IV. CONCLUSIONS

In this paper, as a novel approach, geographic information is introduced in the context of road detection. The key idea of the method is inferring the road geometry ahead the target vehicle using available geographic road databases. Further, the result obtained is refined using chromatic information at pixel level. Hence, road geometry from geographic information provides a rough detection of the road while color provides the accuracy required to discard other objects present in the scene (such as vehicles or pedestrians).

In the future, we aim to incorporate navigation algorithms to minimize the localization error.

V. ACKNOWLEDGMENTS

..

REFERENCES

[1] T. Michalke, R. Kastner, M. Herbert, J. Fritsch, and C. Goerick, "Adaptive multi-cue fusion for robust detection of unmarked inner-city streets," June 2009, pp. 1 –8.

[2] M. Sotelo, F. Rodriguez, L. Magdalena, L. Bergasa, and L. Boquete, "A color vision-based lane tracking system for autonomous driving in unmarked roads," *Auton. Robots*, vol. 16, no. 1, 2004.

[3] Y. He, H. Wang, and B. Zhang, "Color–based road detection in urban traffic scenes," *IEEE Trans. on Intel. Transp. Systems*, vol. 5, no. 24, 2004.

[4] P. Lombardi, M. Zanin, and S. Messelodi, "Switching models for vision-based on–board road detection," in *Procs. IEEE Intl. Conf. on Intel. Transp. Systems*, 2005.

[5] J.M.Alvarez, T. Gevers, and A.M.Lopez, "Vision based road detection using road models," in *ICIP '09: Procs. of the 2009 IEEE Inter. Conf. on Image Processing*. Cairo, Egypt: IEEE Computer Society, 2009, pp. 2073–2076.

[6] J. Levinson, M. Montemerlo, and S. Thrun, "Map-based precision vehicle localization in urban environments," in *Proceedings of Robotics: Science and Systems*, Atlanta, GA, USA, June 2007.

[7] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, "Self-supervised monocular road detection in desert terrain," in *Robotics: Science and Systems*, 2006.

[8] C. Wang, Z. Hu, N. Hamada, and K. Uchimura, "An improved road network modeling and map matching for precise vehicle localization." in *ICIRA: Intelligent Robotics and Applications, Second International Conference*, vol. 5928. Singapore: Springer, December 2009, pp. 174–184.

[9] Z. Hu and K. Uchimura, "Real-time data fusion on tracking camera pose for direct visual guidance," in *Intelligent Vehicles Sysmposium*, 2004, pp. 842–847.

[10] G. Stockman and L. G. Shapiro, *Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[11] R. Gonzalez and R. Woods, *Digital Image Processing (2nd Edition)*. Prentice Hall, 2002, ch. Section 10.4.

[12] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, May 2002.

[13] J. M. Álvarez, A. M. López, and R. Baldrich, "Illuminant-invarariant model-based road segmentation," in *Procs. of the 2008 IEEE Intel. Vehicles Symposium (IV'08)*, Eindhoven, The Netherlands.

[14] National Marine Electronics Association, "NMEA 0183: Standard for interfacing marine electronic devices. version 3.01," *NMEA*, Januray 2002.

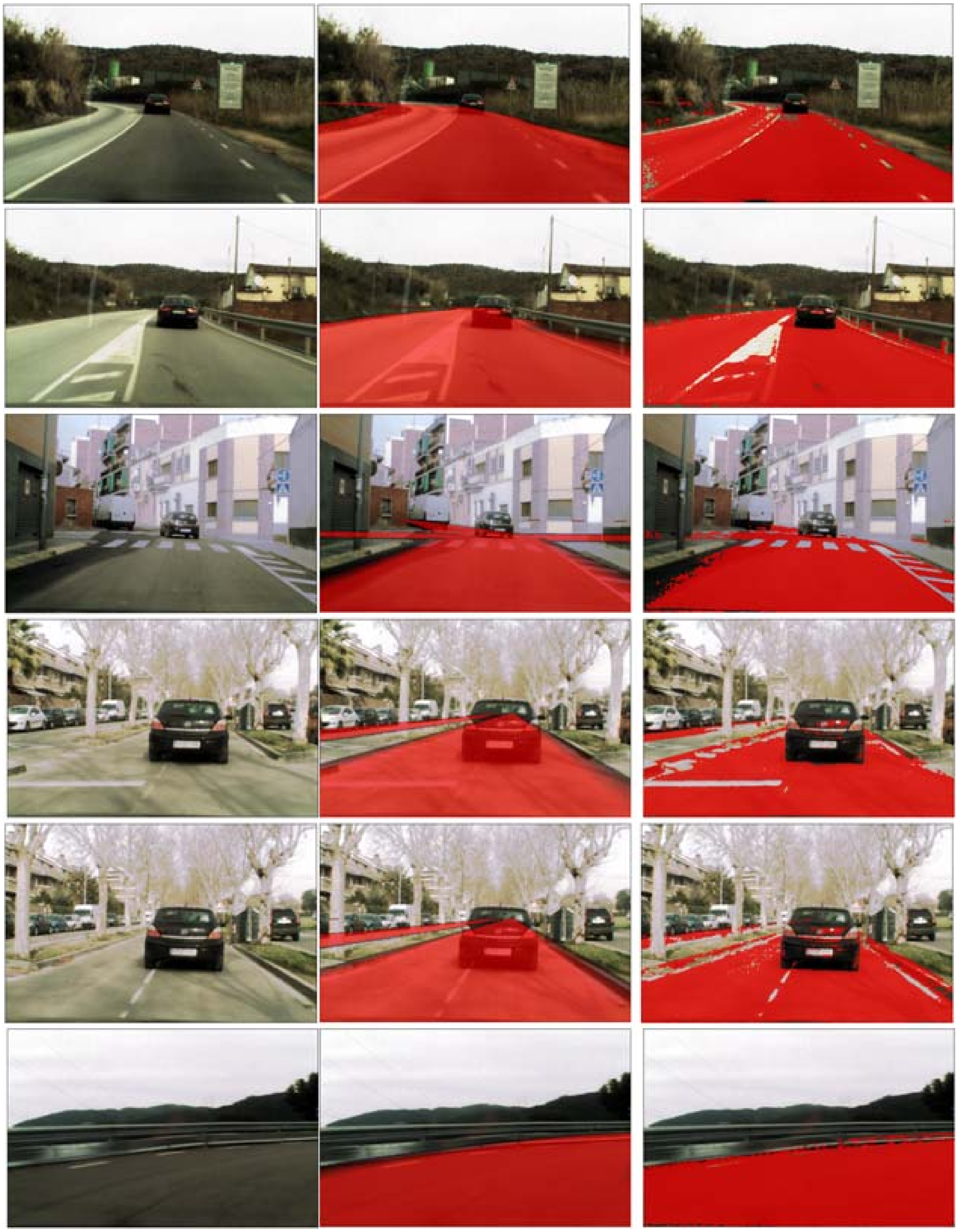

Fig. 9. Example results. Left column shows the original image. Middle column shows the road geometry results. Refined results using chromatic information are shown in the right column.

[15] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 12–18, Oct.-Dec. 2008.

[16] OSM. Openstreetmap: Data Primitives, "Available online: http://wiki.openstreetmap.org/wiki/data_primitives," December 2009.

[17] OSM. Openstreetmap: Map Features, "Available online: http://wiki.openstreetmap.org/wiki/map_features," December 2009.