

Combining Priors, Appearance, and Context for Road Detection

Jose M. Álvarez, *Member, IEEE*, Antonio M. López, *Member, IEEE*,
Theo Gevers, *Member, IEEE*, and Felipe Lumberras

Abstract—Detecting the free road surface ahead of a moving vehicle is an important research topic in different areas of computer vision, such as autonomous driving or car collision warning. Current vision-based road detection methods are usually based solely on low-level features. Furthermore, they generally assume structured roads, road homogeneity, and uniform lighting conditions, constraining their applicability in real-world scenarios. In this paper, road priors and contextual information are introduced for road detection. First, we propose an algorithm to estimate road priors online using geographical information, providing relevant initial information about the road location. Then, contextual cues, including horizon lines, vanishing points, lane markings, 3-D scene layout, and road geometry, are used in addition to low-level cues derived from the appearance of roads. Finally, a generative model is used to combine these cues and priors, leading to a road detection method that is, to a large degree, robust to varying imaging conditions, road types, and scenarios.

Index Terms—Illuminant invariance, lane markings, road detection, road prior, road scene understanding, vanishing point, 3-D scene layout.

I. INTRODUCTION

ROAD detection is an important research topic in different applications of computer vision, such as autonomous driving [1], vehicle collision warning, pedestrian crossing detection, or road scene understanding. Detecting roads using a single color camera mounted on the windshield of a moving vehicle is a challenging task since algorithms must deal with



Fig. 1. Road detection algorithms must be able to deal with continuously changing imaging conditions, different environments and road types, and varying photometric conditions.

continuously changing imaging conditions (varying photometric and weather conditions), different environments (urban roads and highways), different road types (asphalt and off-road), and the presence of different objects (vehicles and pedestrians) with unknown movements (see Fig. 1).

Standard vision-based road detection methods rely on low-level features such as color [2], [3], texture [4], or their combination [1], [5]. The performance of these systems is usually improved by including constraints such as specific road shapes [2] and temporal coherence [6]. However, these constraints restrict the algorithms to highly structured roads and simplified road shapes, reducing their applicability in real-world scenarios. In addition to low-level features, contextual cues can be taken for road detection, such as horizon lines, vanishing points, lane markings, 3-D scene layout, and road geometry. In contrast to solely using low-level information, our aim is to exploit priors and contextual cues for road detection.

In this paper, we first propose an algorithm to estimate online road priors using geographical information systems (GISs), in terms of digital maps combined with a navigation system. These priors are obtained by building a road map using information, such as type of road and the number of lanes, retrieved from a database and then projecting the road map onto the driver's view. This projection includes a model to represent the inherent localization error, the knowledge about the ego position of the vehicle, and the uncertainty in the camera pose. The proposed algorithm does not directly rely on images, and hence, its result is independent of the image acquisition conditions.

Second, we use exploit scene analysis to obtain contextual visual cues containing information about the road. More precisely, we exploit five different contextual cues to detect the road: horizon lines (roads should be below them), vanishing points (where roads are aiming), lane markings (delimiting road areas), 3-D layouts (roads should be on ground areas), and road shapes (roads have predefined known shapes). Contextual information is used in addition to low-level visual cues. In particular, we consider photometric invariant information derived from the color appearance of roads.

Manuscript received June 24, 2013; revised October 13, 2013; accepted December 1, 2013. Date of publication January 22, 2014; date of current version May 30, 2014. The NICTA is funded by the Australian Government through the Department of Broadband, Communications, and the Digital Economy, as well as the Australian Research Council through the ICT Center of Excellence Program. This work was supported in part by the Spanish Government through the Research Program Consolider Ingenio 2010: Multimodal Interaction in Pattern Recognition and Computer Vision (MIPRCV) under Grant CSD200700018 and in part by the MINECO through Projects TRA2011-29454-C03-01, TIN2011-25606, and TIN2011-29494-C03-02. The Associate Editor for this paper was N. Zheng.

J. M. Álvarez is with the NICTA, Canberra, A.C.T. 2601, Australia, and also with the Australian National University, Canberra, A.C.T. 0200, Australia (e-mail: jose.alvarez@nicta.com.au).

A. M. López and F. Lumberras are with the Computer Vision Center, Department of Computer Science, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain.

T. Gevers is with the Computer Vision Center, 08193 Barcelona, Spain, and also with the Intelligent System Laboratory Amsterdam, Faculty of Science, University of Amsterdam, 1098 XH Amsterdam, The Netherlands.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2013.2295427

Finally, we present a generative model for combining priors and contextual and low-level information. The model considers individual cues as weak cues, which are sensitive to different imaging conditions, and combines them to improve the overall accuracy of individual cues. Therefore, the proposed method extracts information at the image and pixel level to obtain a diversified ensemble of cues, which is used as the input of a probabilistic framework.

II. RELATED WORK

Road Detection: Current monocular road detection approaches use low-level features, such as texture or color, as visual cues for grouping pixels in two classes: free road areas or background. Texture provides local-level information with some invariance to lighting variations at the expense of a higher computational cost [4], [7]. However, texture-based approaches are dependent on the strong textures parallel to the road direction, in the form of lane markings for paved roads or tracks created by other vehicles in rural (unpaved) roads. Color provides powerful information of the road, even in the absence of reliable shape patterns, and has been widely accepted as the main cue for road detection. Color also imposes less physical restrictions on the road shape and vehicle speed, leading to more versatile systems. The two most popular color spaces, which have proved to be robust to illuminant intensity changes, are *HSV* [2] and normalized *RGB* [6]. However, the color appearance of the road varies significantly under severe lighting variations (e.g., strong shadows and highlights). Furthermore, there is a lack of texture and color information under extreme situations (see Fig. 1). Therefore, low-level information is not completely feasible for real-world driving scenarios.

Other road detection approaches rely on scene analysis to deal with local lighting variations and localize the road in an image. For instance, a common approach consists of using lane markings [8]–[10] for structured roads or road boundaries [11] for general roads. Scene analysis is also exploited for road detection in [12], where typical road geometries (e.g., left turn, straight, and t-like junction) are learned offline, and then, a scene classifier is used to obtain the most probable road geometry in the current image. Based on this geometry, rough localization of the road is provided. The main disadvantage of these contextual-based approaches is the lack of accuracy to detect road areas, particularly in cluttered scenes (e.g., traffic or urban scenarios).

Recent approaches to road scene understanding consist of combining cues at different levels to improve robustness against challenging situations. For instance, in [5], a set of appearance-based features is combined with motion cues to simultaneously segment and identify multiple objects in a road scene, such as buildings, cars, roads, or bikes. However, these approaches rely on specific training samples. Thus, they are restricted to perform on images highly correlated with the training samples.

The performance of any of these approaches is usually improved by including prior information, in terms of temporal coherence [6] or shape restrictions [2]. Temporal coherence refers to averaging the results of consecutive frames in a sequence. Shape restrictions refer to modeling the road shape in previous

results and using it to restrict the possible road area in the current image. These two priors rely on past detection results to constraint the analysis of the current image. Hence, these priors may propagate errors, particularly when other vehicles are present in the scene.

GPS Information: In recent years, the wide range of high-quality location information combined with affordable global positioning system (GPS) receivers has enabled mass-market mapping systems, such as vehicle navigation (e.g., localizing the vehicle within a map) [14] or visual guidance using augmented reality [15]. However, these approaches use information from the current image to perform their task, and thus, they fail with extreme image acquisition conditions.

In the next section, we propose a novel algorithm to obtain road priors by combining the localization information provided by an onboard standard GPS antenna and the information available in GISs. Furthermore, we present a road detection system that, in addition to low-level information, exploits priors and contextual cues.

III. ROAD PRIORS

Here, we propose a novel algorithm to estimate road priors online using information available in GISs. GISs are database systems that capture, store, and manage geographically referenced object information, such as rivers, lakes, or roads, and also include road attributes, such as road name, road type, and number of lanes in each direction [16]. Our algorithm consists of three main steps (see Fig. 2): data acquisition, road map modeling, and road shape projection.

Data Acquisition: The first step consists of acquiring georeferenced images using an onboard camera and a GPS antenna [see Fig. 2(a)]. GPS information is used to localize the vehicle in a digital map and to retrieve road information from the database within its surrounding area.

Road map Modeling: The second step consists of modeling the road map to generate a birds' view of the road [see Fig. 2(b)]. A road skeleton providing a rough description of the road is built using a set of points connected by piecewise straight lines. This skeleton is smoothed using cubic interpolation to obtain the road trajectory. Finally, this trajectory is expanded using a synthetic road model to obtain a realistic birds' view of the road. As shown in Fig. 2(b), the synthetic road model consists of a drivable area per lane and two roadsides. The number of lanes for the current road is obtained from the road information in the database. The width of the road lane and the roadside is estimated according to the information in the database and additional country's national road legislation.

Road Shape Projection: The last step of the algorithm consists of projecting road points in the birds' view map, i.e., $\mathbf{P}_W = [p_{W0}, p_{Wi}, \dots, p_{WN}]'$, into points in the image plane of the onboard camera (the 2-D driver's view), i.e., $\mathbf{P}_I = [p_{I0}, p_{Ii}, \dots, p_{IN}]'$ [see Fig. 2(c)]. This projection is decomposed in a coordinate system transformation and a perspective projection as follows:

$$\mathbf{P}_I = K M_{WC} \mathbf{P}_W$$

$$M_{WC} = \begin{bmatrix} R_x(\gamma) R_y(\beta) R_z(\alpha) & l \\ 0 & 1 \end{bmatrix}. \quad (1)$$

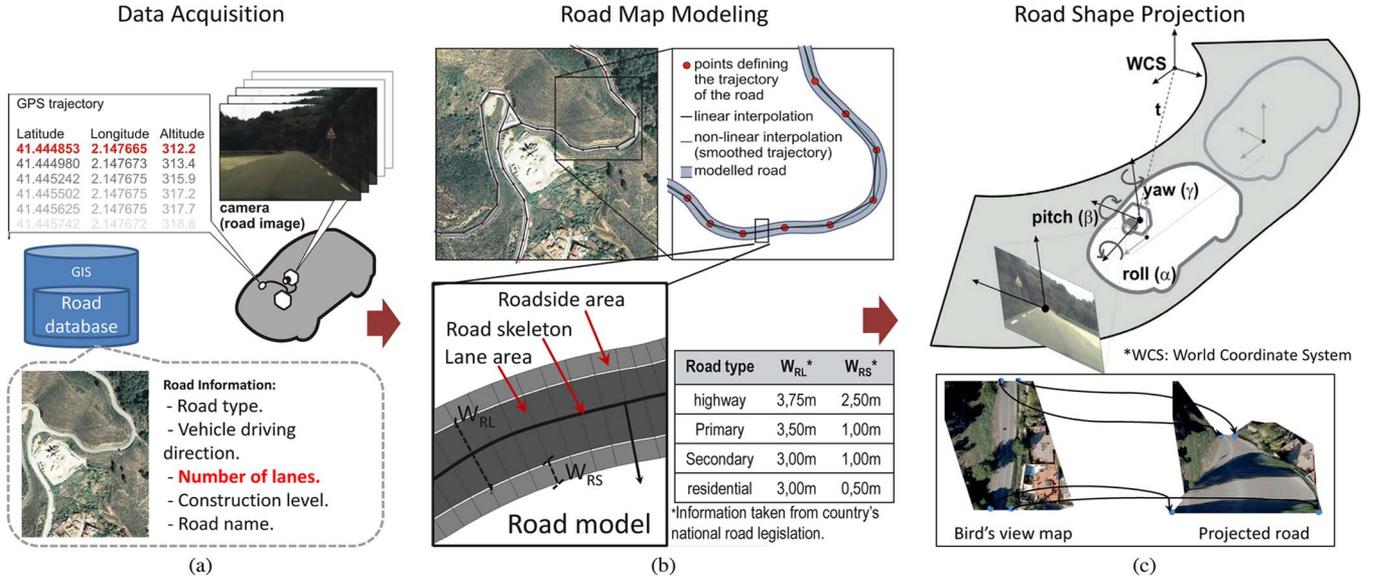


Fig. 2. Proposed algorithm to estimate road priors online using GISs. (a) Acquisition platform consisting of an onboard camera and a GPS antenna. The road database is obtained from OpenStreetMap [13] and provides geographical information and road attributes, such as road type (motorway, path, trunk, primary road, and secondary road), road name, maximum speed, or one or two ways. (b) Road map modeling based on a synthetic road model. (c) Road shape projection: the birds' view is projected into the camera image plane. WCS stands for World Coordinate System.

where K is the intrinsic camera matrix and is estimated through calibration [17]. M_{WCS} is a transformation matrix defining the location ($l = [l_x, l_y, l_z]$) and orientation ($R_x(\gamma)R_y(\beta)R_z(\alpha)$) of the camera referred to the world coordinate system (WCS) [see Fig. 2(c)]. Rotation (γ, β, α) and translation (l_x, l_y, l_z) parameters can be set using calibration through registration [18] or empirical calibration [19]. The former consists of matching the projected road with road features extracted from the current image. However, registration algorithms rely on road information from the current image and are time consuming. The latter, i.e., empirical calibration, consists of learning the parameters from training images and assumes that these parameters do not vary over time [19]. Hence, this method is not usually feasible since the camera undergoes motions due to the vehicle dynamics and road imperfections. Moreover, there are localization errors in the GIS road information, and there is an inherent error in the estimation of l since the spatiotemporal resolution of GPS information is limited. Therefore, our proposal consists of modeling these errors within the projection step.

The key idea of our method to obtain online priors is exploiting localization errors and the variability of placement parameters. The algorithm, which is depicted in Fig. 3, first generates L sets of parameters representing different combinations of locations and positions [see Fig. 3(a)]. Then, for each set of parameters, we obtain a binary road mask P_I^j using (1). Each of these masks represents the road viewed from a different vehicle localization and camera pose in the vehicle [see Fig. 3(b)]. Finally, a confidence map representing the road prior \mathcal{L}_{gps} is obtained as follows [see Fig. 3(c)]:

$$\mathcal{L}_{gps}(x_i) = \frac{1}{L} \sum_{j=1}^L P_I^j(x_i) \quad (2)$$

where x_i is the i th pixel in the image. $\mathcal{L}_{gps}(x_i)$ ranges from 0 to 1. Higher values represent the higher potential of a pixel being a road pixel.

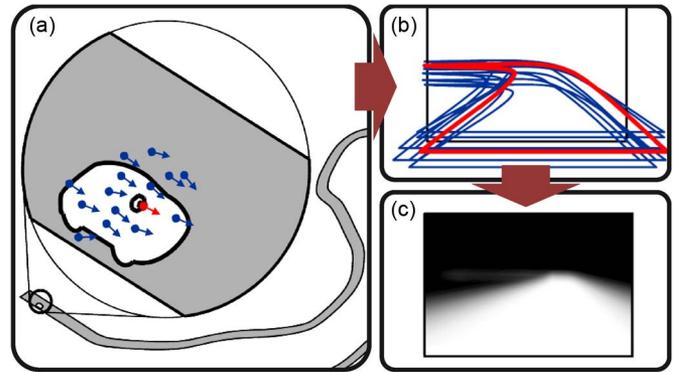


Fig. 3. Road prior (i.e., road confidence map) is obtained modeling the uncertainty (i.e., errors) in camera pose and vehicle position.

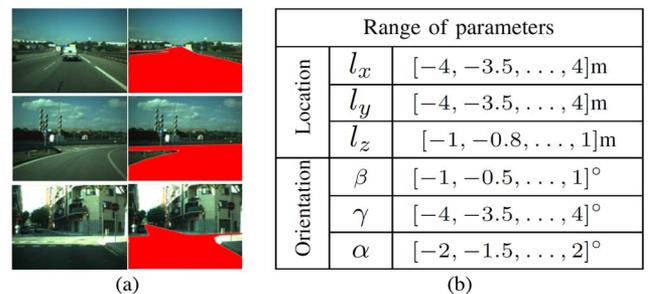


Fig. 4. (a) Ground truth used to calibrate the online road prior algorithm. Left column shows original images. Right column shows the manually generated ground-truth. This ground truth is generated by delimiting the road shape excluding objects on the scene. (b) Range of variability of localization and camera pose parameters used to generate the road prior.

In practice, we reduce the computational cost of the algorithm by using a Monte Carlo approach. Thus, each set of placement parameters is randomly generated from probabilistic (uniform) distributions to simulate the process of sampling their complete range. We adjust the range of values of each parameter using manually generated ground-truth [see Fig. 4(a)] as



Fig. 5. Example results of online road prior generation. Given (top row) a georeferenced image, a confidence map with the per-pixel potential of being a road pixel is generated based on (bottom row) digital road maps and GPS information. The algorithm does not depend on image features, and thus, the confidence map is recovered despite lighting variations or imaging conditions. Additional results can be found at <http://www.josemalvarez.net>.

follows. First, exhaustive variation of the range of values of each parameter is used to obtain different confidence maps. Then, the detection rate of each parameter set is evaluated by binarizing the confidence map. Finally, we select the set of parameters providing a 100% detection rate (covering all the road pixels) in at least 10% of the road masks (i.e., the confidence map \mathcal{L}_{gps} is binarized using 0.1). This criterion ensures that the prior does not limit the performance of the algorithm (in terms of detecting road pixels) since the entire road depicts a road probability larger than 0. As a result, the range of different parameters is listed in Fig. 4(b). The summary of the algorithm is listed in Algorithm 1.

Algorithm 1: Algorithm to estimate road priors using GPS information. A video demonstration can be found at <http://www.josemalvarez.net>

Input: Georeferenced onboard image and digital road map information.

Output: Pixelwise confidence map (\mathcal{L}_{gps}) showing potential road areas.

- ▶ Locate the image on the map using GPS information.
 - ▶ Retrieve road information of the vehicle (image) surroundings from the database.
 - ▶ Build the road skeleton using pairwise connections between consecutive points. Smooth this trajectory using cubic interpolation.
 - ▶ Retrieve road information (type of road and number of lanes) from the database to estimate W_{RL} and W_{RS} . Use these values to generate the road model.
 - ▶ Generate a road birds' view applying the road model to the road skeleton.
 - ▶ Generate L (i.e., $L = 100$) different sets of placement parameters (rotation angles γ , β , α and translation l_x , l_y , l_z) within the discrete range in Fig. 4(b).
 - ▶ Generate, for each set of parameters, a different driver's view image (P_I) using (1).
 - ▶ Obtain the road prior at each pixel using (2).
-

Qualitative online road prior results are shown in Fig. 5. As shown, confidence maps cover the entire road surface despite acquisition and illumination conditions. This is due to the fact that the proposed algorithm does not rely on image features but on digital maps and the GPS signal. The analysis of failures reveals errors in the positioning process due to the inherent

TABLE I
PROPERTIES OF DIFFERENT VISUAL ROAD CUES FROM A THEORETICAL POINT OF VIEW. INVARIANCE IS INDICATED WITH “+” AND LACK OF INVARIANCE WITH “-”

	Color invariants	Horizon line	Vanishing point	Lane markings	3D scene layout	Road shape
Illum. changes	+	+	+	+	+	+
Strong shadows & highlights	+	+	-	+	-	+
Road type	+	+	-	-	+	+
Light traffic	+	+	+	+	+	+
Heavy traffic	+	+	-	-	+	-
Urban scenes	+	+	-	-	+	-
Ability to model the road	+	-	-	-	+	-

error in the accuracy of the GIS database and in urban scenarios where GPS signals are often blocked. These errors could be mitigated by using an inertial navigation system and integrating other positioning techniques, such as visual simultaneous localization and mapping (SLAM), to reduce the localization error. This improvement would also enable having precomputed maps to reduce the computational cost (e.g., reduce the prior to a set of parametric representations that can be retrieved in real time from the database). In addition, in areas where there is no GPS signal, the system could switch to using offline priors based on the quality flag provided by the GPS.

IV. ROAD CUES FOR ROAD DETECTION

Here, road cues are estimated from still images at two different levels. First, low-level (pixel-level) information is used to compute color cues using photometric invariants. Then, high-level scene analysis is used to estimate five contextual cues: horizon line, vanishing point, lane markings, 3-D scene layout, and road shape. These cues are mainly differentiated by their sensitivity to various imaging conditions and traffic situations, as summarized in Table I.

A. Color Cue

Photometric and color information is a powerful cue widely used for road detection. However, color appearance of the road varies depending on the acquisition conditions. Hence, photometric invariance is needed to provide robustness to photometric conditions. Therefore, in this paper, we focus on the photometric invariant approach proposed in [20]. Following this approach, an almost shadow-free image $\mathcal{I}(x_i)$ is obtained as

$$\mathcal{I}(x_i) = r(x_i) \cos \theta + b(x_i) \sin \theta \quad (3)$$

where $r(x_i) = \log(R(x_i)/G(x_i))$ and $b(x_i) = \log(B(x_i)/G(x_i))$, with $R(x_i)$, $G(x_i)$, $B(x_i)$ representing the red, green, and blue color values of the i th pixel in the input image. The angle θ is the invariant direction, which is device dependent, and it does not correlate with the lighting conditions. Hence, the calibration process for each camera need only to be done once using the approach in [21]. As a result, the invariant direction for the color camera used in our experiments is $\theta = 37.5^\circ$.

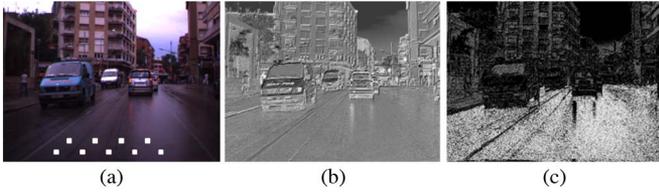


Fig. 6. Color cue from a single image. (a) Input image with seeds used to estimate the road model $p(\cdot)$. (b) Shadow-free image \mathcal{I} . (c) Road confidence map depicting the potential of a pixel being a road pixel. The whiter the color, the higher the potential.

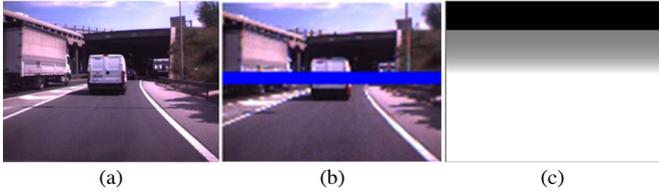


Fig. 7. Horizon line cue computed from a single image. (a) Input image. (b) Estimated horizon line. (c) Pixelwise road confidence map from the horizon line. The whiter the pixel, the higher the potential of being a road class pixel.

Given \mathcal{I} , we generate a confidence map \mathcal{L}_{inv} as follows [see Fig. 6(c)]:

$$\mathcal{L}_{\text{inv}}(x_i) = p(\mathcal{I}(x_i)) \quad (4)$$

where $p(\cdot)$ is the normalized histogram built using the areas of 7×7 pixels around nine seed pixels placed at the bottom part of \mathcal{I} [see Fig. 6(a)]. $\mathcal{L}_{\text{inv}}(x_i)$ ranges from 0 to 1, where the higher the value, the higher the pixel potential of being a road pixel. This color-based road detection algorithm is robust to varying illumination conditions [20]. However, the algorithm may fail when images are overexposed and in the presence of heavy traffic, due to a certain lack of discriminative power.

B. Contextual Cues

We consider five different contextual cues: horizon lines, vanishing points, lane markings, scene layout, and road shape.

1) *Horizon Lines*: Horizon lines contain information to infer where the road is located in each image (the road is usually below the horizon line). To estimate the vertical position of the horizon line in an image (h_t from now on), we follow the approach by Sivic *et al.* in [22], where nonlinear mixtures of linear regressors is applied to the GIST description of the input image [23], [24].

Given h_t , we generate a pixelwise confidence map \mathcal{L}_h as follows [see Fig. 7(c)]:

$$\begin{cases} \mathcal{L}_h(\cdot, i) = 1, & \text{if } i \leq h_t \\ \mathcal{L}_h(\cdot, i) = 0.98^{h_t-i}, & \text{otherwise} \end{cases} \quad (5)$$

where $\mathcal{L}_h(\cdot, i)$ refers to assigning the same value to all the pixels in the i th row. \mathcal{L}_h ranges from 0 to 1, where the higher the value, the higher the potential of being a road pixel.

For video sequences, a Kalman filter is used to smooth the position of the horizon line over time, avoiding radical single-frame errors. The filter takes the vertical position of the horizon line h_t as observation and outputs its filtered version \hat{h}_t . The

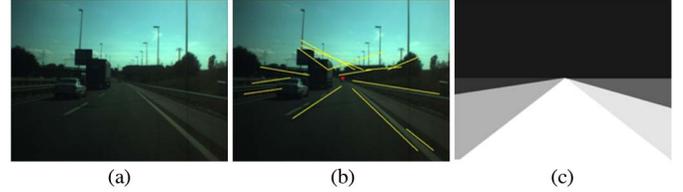


Fig. 8. Given the input image (a), vanishing point is estimated using line segments (b). Then, a pixelwise confidence map is computed (c). The main lane receives higher potential than side lanes. The whiter the color, the higher the pixel potential of being a road pixel.

filter considers the position of the horizon remains constant over time ($y_t = y_{t-1} + w$) and the state is directly the observation ($h_t = y_t + v$), both influenced by Gaussian noise biasing the model ($wN(0, n_w)$ and $vN(0, n_v)$, respectively).

Detecting the road using the horizon line estimation is robust to lighting variations and scenarios, but it is not accurate since it is assumed that all the area below the horizon line belongs to the road surface.

2) *Vanishing Points*: Vanishing points provide powerful information to infer where the road is heading to [11]. Hence, given the vanishing point in an image, road detection can be reduced to the search for road boundaries pointing toward the vanishing point. In this paper, a contextual prior is included to weight the contribution of possible lanes in the scene. We follow the three-step algorithm proposed in [11]. First, edges in an image are detected using the maximum averaged response of a Gabor filter [25]. Then, soft voting, taking into account the distance between the candidate and the voter, is used to locate the vanishing point (see [11] for further details). Finally, we infer the location (boundaries) of different lanes of the road using the dominant and minor edges. In particular, we consider the angle between each edge pixel in the image and the vanishing point. The main lane is defined by lines starting at the vanishing point and using the two angles with highest support. Subsequently, secondary lanes are defined using angles with lower support.

Given the vanishing point and road-lane boundaries, we compute a pixelwise confidence map $\mathcal{L}_{\text{vp}}(X)$ as follows [see Fig. 8(c)]:

$$\begin{cases} \mathcal{L}_{\text{vp}}(x_i) = \max(1 - 0.2N, 0), & \text{if } x_i \text{ is within a lane} \\ \mathcal{L}_{\text{vp}}(x_i) = 0, & \text{otherwise} \end{cases} \quad (6)$$

where N identifies the lane ($N = 0$ for the main lane, $N = 1$ for secondary lanes, $N = 2$ for tertiary lanes, and so on). In this paper, lanes on the left to the main lane receive lower confidences than those located on the right side. \mathcal{L}_{vp} ranges from 0 to 1, where the higher the value, the higher the potential of being a road pixel.

For video sequences, a Kalman filter is used to smooth the position of the vanishing point over time. In this case, the measurement is the 2-D location estimation of the vanishing point in the image, and the state (output) is its filtered version. The dynamic of the filter is the 2-D location and the 2-D velocity of the point (almost zero with no acceleration) influenced by Gaussian noise. Furthermore, the state is directly the measurement, and it is also influenced by Gaussian noise.

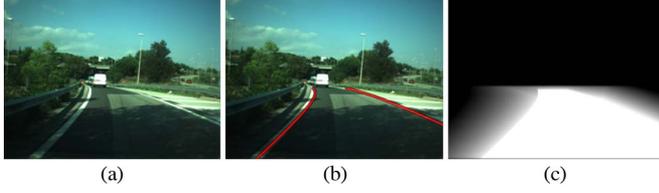


Fig. 9. (a) Original image. (b) Detected lane markings overlapped on the input image. (c) Road confidence map estimated from lane marking results. The whiter the pixels, the higher the pixel potential of being a road pixel.

Detecting the road using vanishing points is robust to global lighting variations, different road types, and damaged roads. However, it is not robust against curved roads, heavy traffic, or strong shadows.

3) *Lane Markings*: Lane markings also provide powerful information to infer where the road is heading to. Given the lane markings of the main lane, road detection is reduced to areas within those markings. To detect the lane markings, we follow the single-lane algorithm based on ridges and random sample consensus (RANSAC) proposed in [10]. This algorithm consists of two main steps. First, a binary image is obtained based on how much a pixel resembles a ridge (the medial axis of a thick, bright, and elongated structure). Then, lane markings are modeled using a pair of hyperbolas with a common asymptote [see Fig. 9(b)]. The parameters of these curves are estimated simultaneously using RANSAC.

Given the lane markings, we compute a road confidence map as follows [see Fig. 9(c)]:

$$\begin{cases} \mathcal{L}_{lm}(x_i) = 1, & \text{if } x_i \text{ is within lane markings} \\ \mathcal{L}_{lm}(x_i) = e^{-d_{x_i}/25}, & \text{otherwise} \end{cases} \quad (7)$$

where d_{x_i} is the distance between x_i and the closest road-lane boundary.

For video sequences, a Kalman filter is used to smooth changes in the lane model from one frame to the next one; see [26] for details.

Detecting the road using lane markings is, to a certain extent, robust to lighting variations and shadows. However, it is not suitable for urban scenarios since the algorithm is limited to the presence of continuous lane markings or a minimum vehicle speed (around 80 km/h) for dashed lane markings. In addition, the resulting road confidence map is not accurate for multiple lanes or when other vehicles are present in the scene.

4) *Three-Dimensional Scene Layout*: Another contextual cue providing relevant information about the road is the 3-D layout of the scene (sky, vertical surface, and ground pixels) by limiting the road to be in ground image regions. We consider the approach in [27] to compute the scene layout. This method provides, for each pixel, a label and three values P_g , P_v , and P_s ranging from 0 to 1, representing the confidence of the classifier for assigning a ground, vertical, or sky label, respectively (see Fig. 10).

Given P_g , P_v , and P_s , we obtain a road confidence map as follows [see Fig. 10(d)]:

$$\begin{cases} \mathcal{L}_{3D}(x_i) = P_g(x_i), & \text{if } P_g(x_i) \geq \max(P_v(x_i), P_s(x_i)) \\ \mathcal{L}_{3D}(x_i) = 0, & \text{otherwise} \end{cases} \quad (8)$$

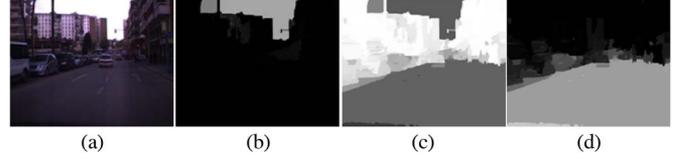


Fig. 10. (a) Input image is partitioned in three 3-D geometry classes: sky, vertical surface and ground. Using the approach of Hoiem *et al.* [27], a pixelwise confidence is estimated for each class [(b) sky areas P_s , (c) vertical surfaces P_v , and (d) ground areas P_g]. The whiter the pixels, the higher the potential of a class.

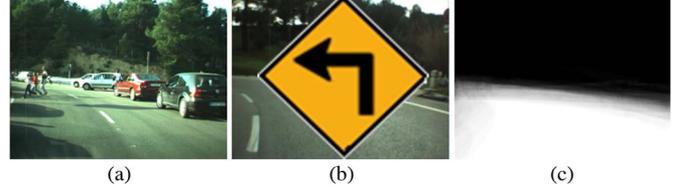


Fig. 11. Geometric information for road detection. (a) Input image. (b) Assigned road shape via scene classification. (c) Pixelwise confidence map learned offline using manually segmented images.

\mathcal{L}_{3D} ranges from 0 to 1, where the higher the value, the higher the potential of being a road pixel.

Road detection using a scene layout is robust to lane markings, pedestrian crossings, and the presence of other vehicles in the scene. However, it is sensitive to shadows and acquisition conditions since the algorithm uses color superpixel segmentation.

5) *Road Shapes*: The shape of the road is another valuable cue for road detection since it reduces the areas in the image where the road is located. In this paper, we consider the approach in [28] to detect the shape of the road ahead the target vehicle. The algorithm first learns road shape models from training images containing specific road shapes, such as left turn, right turn, straight road, or junctions. Then, a support vector machine (SVM) classifier is used to infer the road shape in the test image.

Given the class representing the road shape of the test image, we obtain a confidence map as follows [see Fig. 11(c)]:

$$\mathcal{L}_{sh} = \frac{1}{M} \sum_{j=1}^M GT_j^c \quad (9)$$

where c is the class assigned to the test image by the classifier, and GT_j^c is the j th annotated mask in the set of M training images for the c th class. \mathcal{L}_{sh} ranges from 0 to 1, where the higher the value, the higher the potential of being a road pixel.

Detecting the road using road shape models is robust to local illumination effects such as shadows and highlights. Furthermore, the result cannot degenerate since it is forced to be in one of the road shape classes. However, it is limited to the number of classes in the training set.

V. COMBINING CUES AND PRIORS FOR ROAD DETECTION

Here, we propose an algorithm for road detection based on visual cues and priors. First, a naive Bayesian framework is used to combine cues and online priors. Then, for video sequences, an adaptive model [29] is used to exploit interframe

correlations due to the sequential nature of the data. Finally, a binary road mask is obtained using a threshold. The algorithm is summarized in Algorithm 2.

Naive Bayesian Framework: The Bayesian framework performs in a frame-by-frame basis to estimate the probability of each pixel being a road pixel. We consider each visual cue as an observation (i.e., color, horizon line, vanishing point, lane markings, 3-D scene layout, and road shape) providing six different confidence maps (in the range $\in [0, 1]$) and interpreted as road likelihood functions \mathcal{L}_{inv} , \mathcal{L}_h , \mathcal{L}_{vp} , \mathcal{L}_{lm} , \mathcal{L}_{3D} , and \mathcal{L}_{sh} , respectively. Then, given an image, the probability of each pixel x_i being a road surface pixel given is computed as follows:

$$P_r(x_i) \propto \mathcal{L}_{gps} \prod_j \mathcal{L}_j(x_i = R), \quad j \in [inv, h, vp, lm, 3D, sh] \quad (10)$$

where the online prior acts as a probabilistic description of the road before any visual observation is processed.

Algorithm 2: Road detection algorithm by combining cues and priors.

Input: Georeferenced onboard image/sequence, digital road map information, and parameter set (τ , λ , and α).

Output: Binary mask B depicting road areas.

- ▶ Retrieve a georeferenced image I_k , its geolocation, and related information contained in the road database.
 - ▶ Compute \mathcal{L}_{gps} using (2).
 - ▶ Compute \mathcal{L}_{inv} , \mathcal{L}_h , \mathcal{L}_{vp} , \mathcal{L}_{lm} , \mathcal{L}_{3D} , and \mathcal{L}_{sh} using (4)–(9), as detailed in Section IV.
 - ▶ Compute $P_r(x_i)$ by combining cues and priors at the frame level using (10).
 - ▶ For single frames:
 - ▶ $P_r^t(x_i) = P_r(x_i)$.
 - ▶ For video sequences:
 - ▶ Save P_r in a buffer as the most recent frame.
 - ▶ Obtain $\hat{P}_r(x_i)$ using (12).
 - ▶ Obtain $P_r^t(x_i)$ using (11).
 - ▶ Shift the buffer to discard the oldest frame.
 - ▶ Compute the road binary mask O as follows:
 - if** $P_r^t(x_i) > \tau$ **then**
 $B(x_i) = 1$.
 - else**
 $B(x_i) = 0$.
 - end if**
-

Temporal Adaptation: For video sequences, the algorithm uses an adaptive model [29] to consider the inherent temporal road consistency between consecutive frames. The adaptation consists of a pixel-by-pixel combination between $P_r(x_i)$ and $\hat{P}_r(x_i)$. The former is the output of the Bayesian framework for the current image. The latter is a road confidence map predicted using previous frames (see Section V-A). The combination is performed as follows:

$$P_r^t(x_i) = \alpha P_r(x_i) + (1 - \alpha) \hat{P}_r(x_i) \quad (11)$$



Fig. 12. Example images from the database for three different scenarios: (top row) highways, (middle row) urban scenes, and (bottom row) secondary structured roads.

where $P_r^t(x_i)$ is the road confidence map considering temporal information, and α is an adaptation parameter. The lower the α , the more persistent the model is.

Pixel Classification: Finally, the last step consists of assigning a road or background label to each pixel based on a fixed threshold τ . The road class label is assigned to a pixel if $P_r^t(x_i) \geq \tau$. Otherwise, the pixel belongs to the background class.

A. Temporal Road Prior

In this paper, temporal priors are estimated using an exponentially weighted moving average (EWMA) algorithm [30]. The algorithm assumes that the road detected in the current frame is correlated (similar) to the road detected in previous frames. However, instead of averaging over previous results as in [2], the EWMA uses a decay factor to weight the influence of each past result. Therefore, more recent results receive higher weights than older ones. The result is a process able to cope with sudden changes in the data. This is particularly relevant when other objects are present in the scene (e.g., urban scenarios or junctions). Using the EWMA, temporal priors $\hat{P}_r(x_i)$ are computed as follows:

$$\hat{P}_r(x_i) = \frac{1}{\sum_{j=1}^T \lambda^{j-1}} \sum_{j=1}^T \lambda^{j-1} P_r^{(t-j)}(x_i) \quad (12)$$

where $P_r^{(t-j)}(x_i)$ is the output of the Bayesian framework for the i th pixel j frames before the frame being analyzed, and λ is the decay factor. The lower the decay, the higher weight to recent values. Finally, T is the number of past observations taken into account. We set T to a large number since the weighting procedure will eventually place a zero weight on results far in the past (e.g., $0 < \lambda < 1$; hence, $\lambda^n \rightarrow 0$ when $n \rightarrow \infty$).

VI. EXPERIMENTS

Here, we apply the proposed algorithm to a data set consisting of different georeferenced image sequences acquired using a moving vehicle at different days, different daytime (daybreak, morning, noon, and afternoon), and different weather conditions (sunny, cloudy, and rainy) and for three different scenarios (urban like, highways, and secondary structured roads) exhibiting different challenges, as shown in (see Fig. 12). Ground truth is generated by manually segmenting a subset of 2000 images.

The proposed algorithm is used without specific training to ensure maximum adaptability to unobserved road sequences.

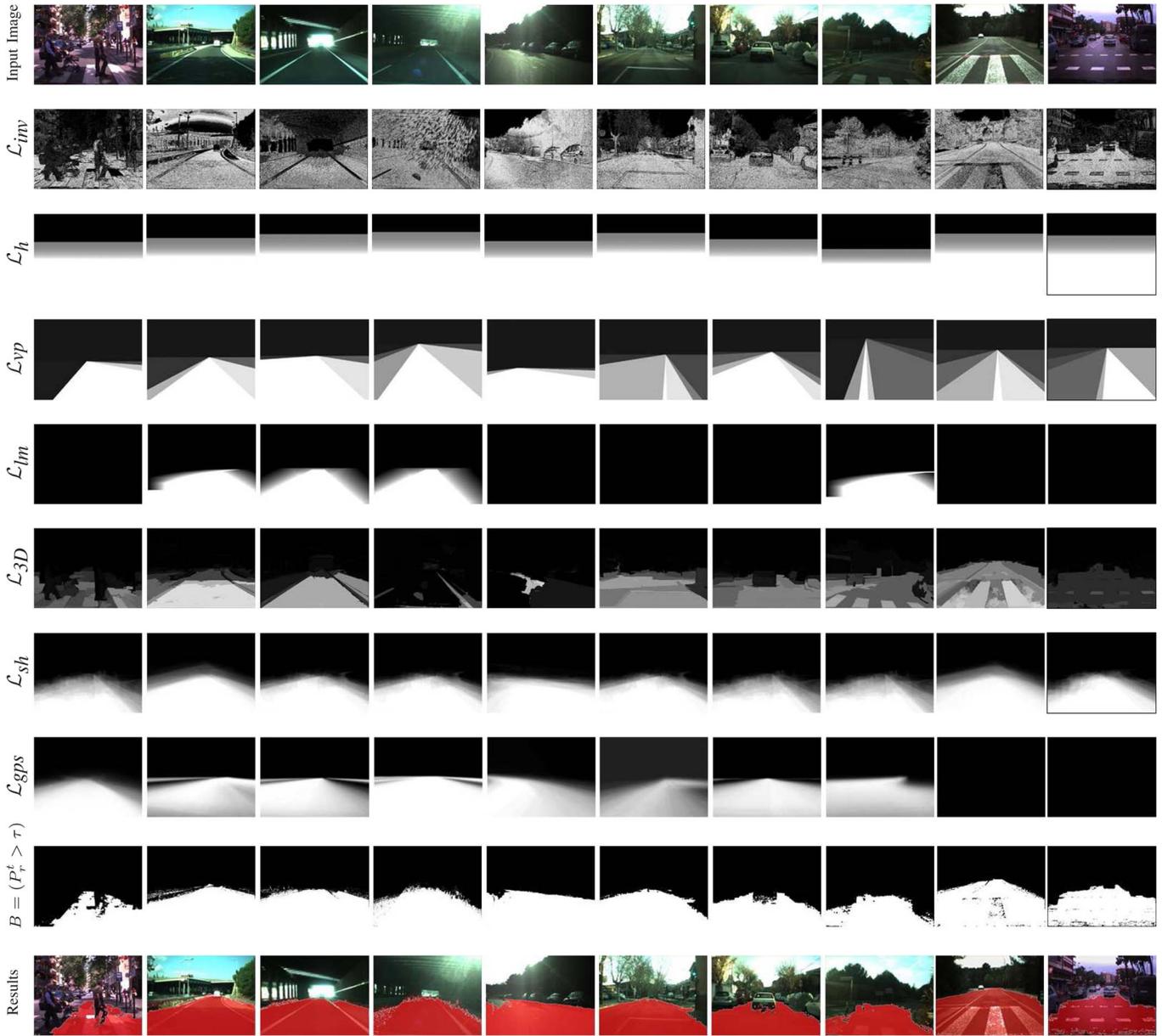


Fig. 13. Example road detection results using cues and priors. From top to bottom: input image; color-based cue \mathcal{L}_{inv} ; horizon line \mathcal{L}_h ; vanishing point \mathcal{L}_{vp} ; lane markings \mathcal{L}_{lm} ; 3-D scene layout \mathcal{L}_{3D} ; road shape \mathcal{L}_{sh} ; online prior \mathcal{L}_{gps} ; final road mask B ; and the road mask overlapped on the input image.

Nevertheless, some adjustments are done offline as follows. First, online priors are adjusted using the algorithm proposed in Section III. The characteristic direction of the camera θ is estimated once using the algorithm described in Section IV-A. The 3-D scene layout and horizon line approaches are used without specific training. Finally, uniform weights are assigned to each cue and prior in the Bayesian framework.

Quantitative evaluations are provided using average receiver operating characteristic (ROC) curves on the pixelwise comparison between the ground truth and the results obtained varying τ (see Section V). ROC curves represent the tradeoff between the true positive rate $TPR = (TP/TP + FN)$ and the false positive rate $FPR = (FP/FP + TN)$. Furthermore, for performance comparison, we consider the area under the curve (AUC) and the equal error rate (EER) defined as the intersection of the curve and the line, where error rates are equal, i.e., $(1 - TPR) = FPR$.

A. Single Cues

Representative qualitative road detection results are shown in Fig. 13, and average ROC curve (including AUC and EER) values per cue for each scenario and the entire data set are shown in Fig. 14. As shown, contextual cues tend to recover a big portion of the road [even in failure cases (see Fig. 13)] producing curves with strong slopes (i.e., fast increment of true positive rate while the false positive rate does not). Moreover, the number of false positives in contextual cues is limited to surrounding road areas and object surroundings, and it is negligible compared with the number of true positives leading to higher AUC values compared to pixel-based cues Fig. 14. This effect is particularly noticeable using road shapes where the output is a fuzzy road map. Hence, for most images, there exists a certain threshold that provides an almost perfect road binary map.

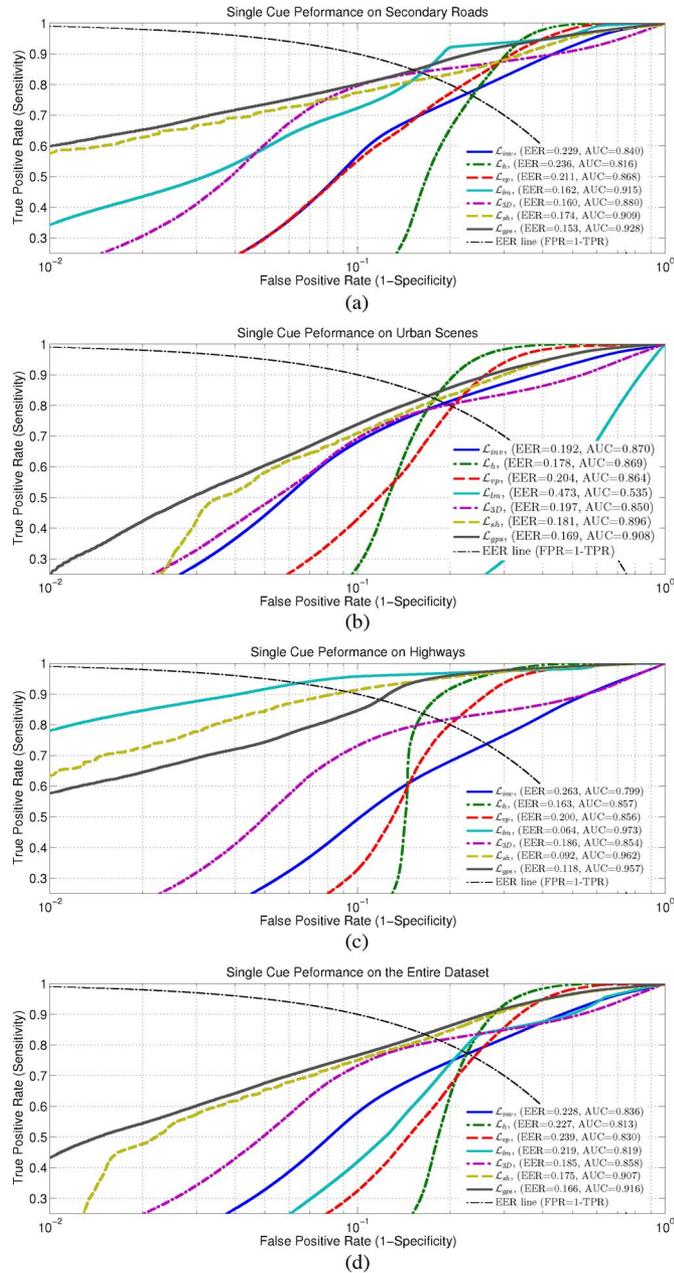


Fig. 14. Average ROC curves per cue and scenario: color-based cue \mathcal{L}_{inv} ; horizon line \mathcal{L}_h ; vanishing point \mathcal{L}_{vp} ; lane markings \mathcal{L}_{lm} ; 3-D scene layout \mathcal{L}_{3D} ; road shape \mathcal{L}_{sh} . (a) Curves for secondary structured roads. (b) Curves for urban scenes. (c) Curves for highways. (d) Curves considering the complete data set. Best viewed in color.

The horizon line and online priors are the algorithms exhibiting the highest true positive rate at the expense of higher false positive rates. The horizon line algorithm increases the number of false positives by systematically assuming that all the pixels below the horizon belong to the road surface. On the other hand, online priors provide false positives by recovering road areas, even if they are occluded by buildings or objects in the scene (e.g., these priors are calibrated to oversegment the road); see Fig. 5. Vanishing points and lane markings (fourth and fifth rows in Fig. 13, respectively) provide the lowest global performance, due to the lack of lane markings or crowded scenes (i.e., other vehicles or pedestrians). However, they are particularly

relevant [exhibiting the highest performance; see Fig. 14(b)] in highways where lane markings are present (see Fig. 13). From these results, we can conclude that the performance of these contextual cues drops off for crowded scenes, due to the presence of objects.

More accurate road masks are provided by color and 3-D scene layout (second and sixth rows in Fig. 13, respectively) since they perform at the pixel level. However, their performance is affected in extreme situations, such as penumbra and highlights, due to the lack of color information. In these cases, both algorithms may fail to detect any road pixel or retrieve all pixels as road pixels. This is an important difference with other color-based road detection algorithms applying constraints, such as region growing [2], [20], to reduce the presence of false positives. In our approach, the color algorithm is not constrained in any sense, whereas the 3-D layout only retrieves pixels exhibiting higher probability of being ground than vertical surface. Furthermore, the performance of the color cue is also affected, to a certain extent, by the lack of discriminative power due to the dimensionality reduction (i.e., from color to a single channel). The performance of the 3-D scene layout is also affected by ground surfaces that do not belong to the road surface (see Fig. 13).

B. Combining Cues and Priors

Here, we analyze the performance of different ensemble configurations. First, we consider ensembles of cues without priors and then ensembles combining different cues and priors.

The performance of leave-one-cue-out ensembles per scenario and for the entire data set is shown in Fig. 15. As shown, the performance varies depending on the scenario being analyzed and, more prominently, on the cue being excluded. The ensemble exhibiting the highest performance is the one excluding the vanishing point cue. Furthermore, the worst performance corresponds to the ensemble excluding the 3-D scene layout. Nevertheless, ensembles combining different cues provide smoother curves with lower variations interscenario. From these evaluations, we can conclude that, in general, the overall performance is improved when cues are combined.

Finally, Fig. 16 shows the average ROC curves of eight different ensemble configurations.¹ For comparison, we also include the ROC curves of four state-of-the-art methods: ST_1 is the ensemble consisting of road shape models and illuminant invariant cue, as proposed in [28]; ST_2 is the illuminant invariant method proposed in [20]; ST_3 is the layout algorithm in [27]; and ST_4 is the lane markings algorithm in [26]. The summary of performance is listed in Table II, and Fig. 17 shows some additional qualitative results of our method compared to the layout algorithm [27] and the road detection method proposed in [11]. As shown, common methods relying on single cues fail to detect the road in many images. However, combining multiple diversified cues improves the output of the

¹ E_1 : color, contextual cues, and online and temporal priors; E_2 : cues and online priors; E_3 : contextual and color cues; E_4 : best leave-one-cue-out ensemble configuration; E_5 : color, horizon line, vanishing point, and road shape; E_6 : 3-D scene layout, horizon line, vanishing point, and road shape; E_7 : color, horizon line, and vanishing point; E_8 : color and 3-D scene layout.

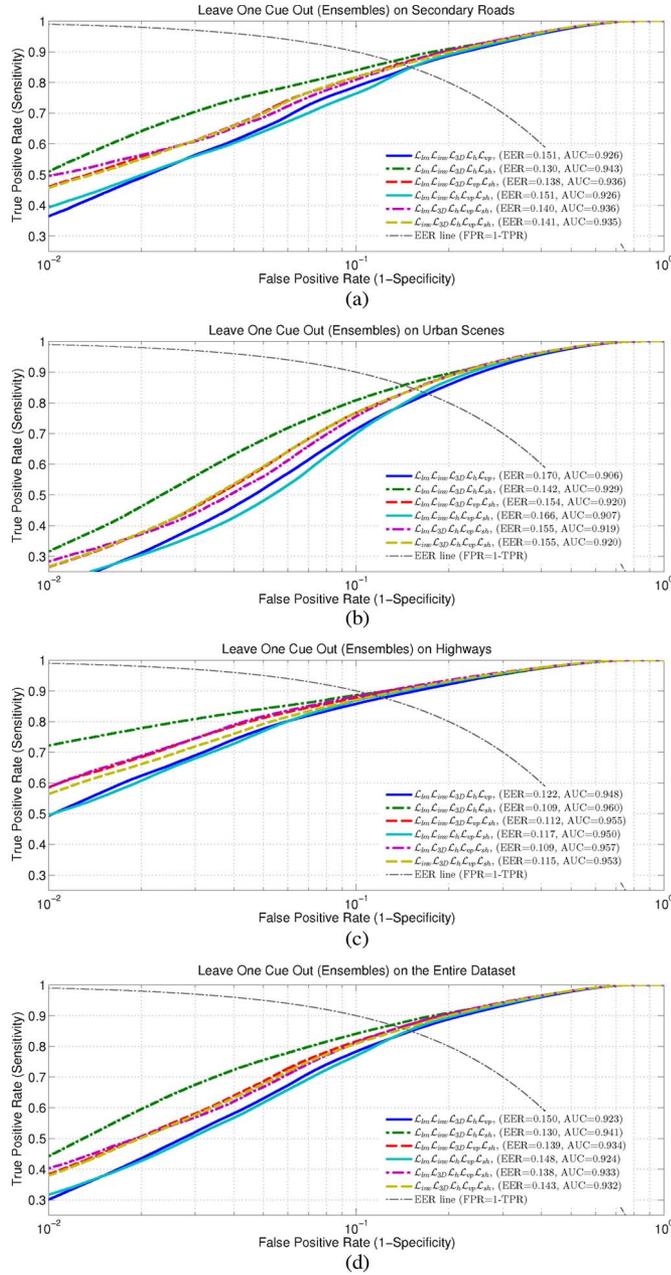


Fig. 15. Average ROC curves using leave-one-cue-out ensembles without considering priors (see Fig. 14 for cue description). (a) Curves for secondary structured roads. (b) Curves for urban scenes. (c) Curves for highways. (d) Curves considering the complete data set. Best viewed in color.

algorithm. As shown, the overall performance increases when all cues are taken into account. This improvement is even higher when road priors are included generating smoother curves. Higher performance is achieved when contextual cues and priors are added to the ensemble (see the performance of E_6 , E_4 , and E_3 , respectively). From these results, we can conclude that contextual cues provide relevant information to infer the location of the road. Nevertheless, low-level information is essential to achieve reliable road detection results. The highest area under ROC value (highest performance) is provided by combining appearance, contextual cues, and priors (online and temporal); see Fig. 16.

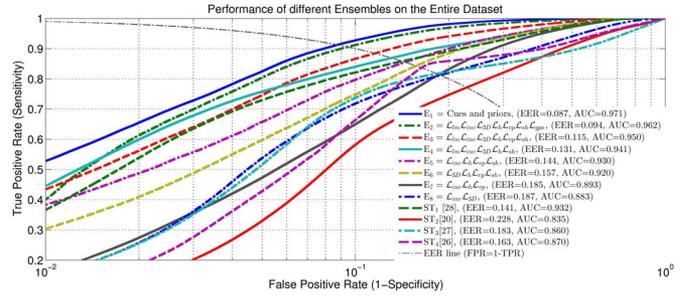


Fig. 16. Average ROC curves using the entire data set for different ensemble configurations and state-of-the-art methods. E_1 : color, contextual cues, online and temporal priors; E_2 : cues and online priors; E_3 : contextual and color cues; E_4 : best leave-one-cue-out ensemble configuration; E_5 : color, horizon line, vanishing point, and road shape; E_6 : 3-D scene layout, horizon line, vanishing point, and road shape; E_7 : color, horizon line, and vanishing point; E_8 : color and 3-D scene layout; ST_1 is the ensemble proposed in [28]; ST_2 is the illuminant invariant approach proposed in [20]; ST_3 is the layout algorithm in [27]; and ST_4 is the lane markings algorithm in [26]. Best viewed in color.

TABLE II
AREA UNDER ROC AND EER FOR THE DIFFERENT ENSEMBLE CONFIGURATIONS AND STATE-OF-THE-ART ALGORITHMS. SEE FIG. 16 FOR DETAILS

	EER	AUC
E_1 : Cues and priors	0.087	0.971
E_2 : $\mathcal{L}_{lm}\mathcal{L}_{inv}\mathcal{L}_{3D}\mathcal{L}_h\mathcal{L}_{vp}\mathcal{L}_{sh}\mathcal{L}_{gps}$	0.094	0.962
E_3 : $\mathcal{L}_{lm}\mathcal{L}_{inv}\mathcal{L}_{3D}\mathcal{L}_h\mathcal{L}_{vp}\mathcal{L}_{sh}$	0.115	0.950
E_4 : $\mathcal{L}_{lm}\mathcal{L}_{inv}\mathcal{L}_{3D}\mathcal{L}_h\mathcal{L}_{sh}$	0.131	0.941
E_5 : $\mathcal{L}_{inv}\mathcal{L}_h\mathcal{L}_{vp}\mathcal{L}_{sh}$	0.144	0.930
E_6 : $\mathcal{L}_{3D}\mathcal{L}_h\mathcal{L}_{vp}\mathcal{L}_{sh}$	0.157	0.920
E_7 : $\mathcal{L}_{inv}\mathcal{L}_h\mathcal{L}_{vp}$	0.185	0.893
E_8 : $\mathcal{L}_{inv}\mathcal{L}_{3D}$	0.187	0.883
ST_1 : [28]	0.141	0.932
ST_2 : [20]	0.228	0.835
ST_3 : [27]	0.183	0.860
ST_4 : [26]	0.163	0.870

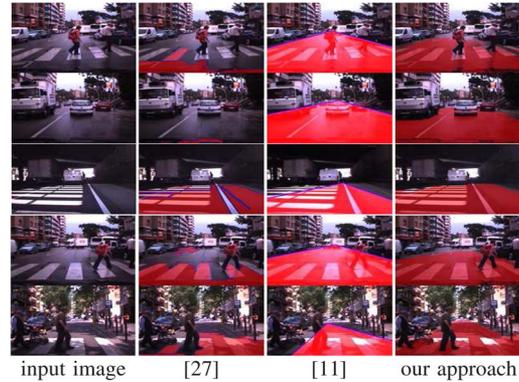


Fig. 17. Qualitative example results for our approach and related methods. Best viewed in color.

From these quantitative and qualitative results, it can be derived that combining cues (color and contextual) improves the overall performance of road detection. This improvement is even higher when temporal information is included. In this case, the improvement is at the expense of slight loss of accuracy in preserving objects present in the scene (vehicles or pedestrians). Hence, we can conclude that combining cues and priors improves the performance of current road detection algorithms, enabling a reliable road segmentation robust against different road scenarios, illumination conditions, and road shapes.

Currently, we are using a nonoptimized MATLAB code to run our experiments. The processing time to process a 320×240 image using a 32-bit DualCore 2.4GHz is given as follows: horizon line 0.16 s; vanishing point 0.2 s; lane markings 0.015 s; road shape 0.10 s; 3-D layout 6 s; \mathcal{I} 0.10 s and negligible to build the ensemble. Nevertheless, the system allows cue parallelization; hence, the real-time processing rate could be easily reached by the same machine.

REFERENCES

- [1] C. Thorpe, M. Hebert, T. Kanade, and S. Shafer, "Vision and navigation for the Carnegie-Mellon Navlab," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 3, pp. 362–373, May 1988.
- [2] M. A. Sotelo, F. J. Rodriguez, and L. Magdalena, "Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 2, pp. 69–83, Jun. 2004.
- [3] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 24, pp. 309–318, Dec. 2004.
- [4] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in *Proc. IEEE CVPR*, 2004, pp. 470–477.
- [5] P. Sturgess, K. Alahari, L. Ladicky, and P. H. S. Torr, "Combining appearance and structure from motion features for road scene understanding," in *Proc. BMVC*, 2009, pp. 1–11.
- [6] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," in *Proc. IEEE ITSC*, 2006, pp. 939–944.
- [7] Y. Alon, A. Ferencz, and A. Shashua, "Off-road path following using region classification and geometric projection constraints," in *Proc. IEEE CVPR*, 2006, pp. 689–696.
- [8] Z. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 16–26, Mar. 2008.
- [9] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.
- [10] A. M. López, J. Serrat, C. Cañero, F. Lumbreras, and T. Graf, "Robust lane markings detection and road geometry computation," *Int. J. Autom. Technol.*, vol. 11, no. 3, pp. 395–407, Jun. 2010.
- [11] H. Kong, J. Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Trans. Image Processing*, vol. 19, no. 8, pp. 2211–2220, Aug. 2010.
- [12] J. M. Alvarez, T. Gevers, and A. M. Lopez, "Vision based road detection using road models," in *Proc. IEEE ICIP*, 2009, pp. 2073–2076.
- [13] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, Oct./Dec. 2008.
- [14] C. Wang, Z. Hu, N. Hamada, and K. Uchimura, "An improved road network modeling and map matching for precise vehicle localization," in *Proc. ICIRA*, Dec. 2009, vol. 5928, pp. 174–184, Springer-Verlag.
- [15] Z. Hu and K. Uchimura, "Real-time data fusion on tracking camera pose for direct visual guidance," in *Proc. IEEE IV*, 2004, pp. 842–847.
- [16] I. Heywood, S. Cornelius, and S. Carver, *An Introduction to Geographical Information Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
- [17] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edition ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [18] A. Myronenko, X. Song, and M. Carreira-Perpinan, "Non-rigid point set registration: Coherent point drift," in *Proc. NIPS*, Cambridge, MA, USA, 2007, pp. 1009–1016, MIT Press.
- [19] J. M. Alvarez, F. Lumbreras, T. Gevers, and A. M. Lopez, "Geographic information for vision-based road detection," in *Proc. IEEE IV*, 2010, pp. 621–626.
- [20] J. M. Alvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 184–193, Mar. 2011.
- [21] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.
- [22] J. Sivic, B. Kaneva, A. Torralba, S. Avidan, and W. T. Freeman, "Creating and exploring a large photorealistic virtual space," in *Proc. IEEE CVPRW*, Jun. 2008, pp. 1–8.
- [23] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, May/Jun. 2001.
- [24] A. Torralba, "Statistical context priming for object detection," in *Proc. IEEE ICCV*, 2001, vol. 1, pp. 763–770.
- [25] T. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.
- [26] D. Ponsa, J. Serrat, and A. M. López, "On-board image-based vehicle detection and tracking," *Trans. Inst. Meas. Control*, vol. 33, no. 7, pp. 783–805, Oct. 2011.
- [27] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 151–172, Oct. 2007.
- [28] J. M. Alvarez, T. Gevers, F. Diego, and A. M. Lopez, "Road geometry classification by adaptive shape models," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 459–468, Mar. 2013.
- [29] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 862–877, Jul. 2004.
- [30] C. Stauffer and E. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.



Jose M. Álvarez (M'06) is currently a Researcher with the NICTA, Canberra, A.C.T., Australia, and also a Research Fellow with the Australian National University, Acton, A.C.T., Australia. Previously, he was a Postdoctoral Researcher with the Computational and Biological Learning Group, New York University, New York, NY, USA. During his postdoctoral studies, he was a Visiting Researcher with Volkswagen Research and the University of Amsterdam, Amsterdam, The Netherlands. His main research interests include road detection, color, photometric invariance, machine learning, and fusion of classifiers.



Antonio M. López (M'10) received the Ph.D. degree from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2000.

In 1996, he participated in the foundation of the Computer Vision Center, UAB, where he has held different institutional responsibilities, presently being responsible for the research group on advanced driver assistance systems by computer vision. He is currently an Associate Professor with the UAB. He has been responsible for public and private projects, and he is a coauthor of more than 100 papers in the

field of computer vision.



Theo Gevers (M'12) is an Associate Professor of computer science with the University of Amsterdam, Amsterdam, The Netherlands, where he is also a Teaching Director of the M.Sc. program in artificial intelligence. His main research interests are in the fundamentals of content-based image retrieval, color image processing, and computer vision, specifically in the theoretical foundation of geometric and photometric invariants.

He was a recipient of the Vici award (for excellent researchers) from the Dutch Organisation for Scientific Research. He serves as Chair for various conferences, and he is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Felipe Lumbreras received the Ph.D. degree in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2001.

He is currently an Associate Professor with the Department of Computer Science, UAB, where he is a member of the Computer Vision Center. His research interests include texture analysis, 3-D reconstruction, and computer vision for automotive applications.