3D Scene Priors for Road Detection

Jose M. Alvarez
Computer Vision Center and
Computer Science Dpt.
Univ. Autonoma de Barcelona
jalvarez@cvc.uab.es

Theo Gevers
Faculty of Science
University of Amsterdam
th.gevers@uva.nl

Antonio M. Lopez
Computer Vision Center and
Computer Science Dpt.
Univ. Autonoma de Barcelona
antonio@cvc.uab.es

Abstract

Vision–based road detection is important in different areas of computer vision such as autonomous driving, car collision warning and pedestrian crossing detection. However, current vision–based road detection methods are usually based on low–level features and they assume structured roads, road homogeneity, and uniform lighting conditions.

Therefore, in this paper, contextual 3D information is used in addition to low–level cues. Low–level photometric invariant cues are derived from the appearance of roads. Contextual cues used include horizon lines, vanishing points, 3D scene layout and 3D road stages. Moreover, temporal road cues are included. All these cues are sensitive to different imaging conditions and hence are considered as weak cues. Therefore, they are combined to improve the overall performance of the algorithm. To this end, the low-level, contextual and temporal cues are combined in a Bayesian framework to classify road sequences.

Large scale experiments on road sequences show that the road detection method is robust to varying imaging conditions, road types, and scenarios (tunnels, urban and highway). Further, using the combined cues outperforms all other individual cues. Finally, the proposed method provides highest road detection accuracy when compared to state–of–the–art methods.

1. Introduction

Vision-based road detection is an important research topic in different areas of computer vision such as autonomous driving, car collision warning and pedestrian crossing detection. Detecting roads using a monocular vision–system is a very challenging problem as the detection algorithm must be able to deal with continuously changing backgrounds, the presence of different objects (vehicles, pedestrians), different environments (urban, highways, off–road), different road types (shape, color), and different imaging conditions (varying illumination, different viewpoints and changing weather conditions). In general, vision–based methods use low-level features for road detection [1, 2, 3, 4]. A forerunner system in outdoor navigation is the Natlab project developed by Thorpe *et al*. [3]. This system combines color and texture information to perform vision–based road tracking. The system is reinforced with a 3D vision based obstacle detection algorithm. More recently, in [1], Sotelo *et al*. consider a sequence of images and combine color information with the shape of the road detected in previous frames. In [4], Lombardi *et al*. use texture to perform pixel classification for road detection.

Color appearance information has been widely used as the main cue for road detection, since color provides powerful information of the road to be detected in the absence of reliable shape information. In addition, color imposes less physical restrictions, leading to more versatile systems. The two most popular color spaces, that have proved to be robust to minor illuminant changes, are *HSV* [1, 5] and normalized *RGB* [2]. However, algorithms based on these color spaces fail under severe lighting variations (strong shadows and highlights among others) and these algorithms show dependency on highly structured roads, road homogeneity, simplified road shapes. Further, these approaches only consider pixel level information to reinforce the system or synthetic data which can not be extracted from a single image. Therefore, in addition to appearance cues, we introduce

Figure 1. Vision–based road detection in real–world scenarios must be robust to extreme situations such as shadows, reflections, crowed scenarios or direct light source incident to the camera.

978-1-4244-6983-3/10/$26.00 ©2010 IEEE

contextual information such as 3D road geometry and shape for road detection. It is increasingly being recognized in the vision community that context information is necessary for reliable extraction of image regions and objects [6, 7, 8]. However, contextual information has not been used for road detection before.

Therefore, in this paper, contextual 3D information is used in addition to low-level cues. Low-level photometric invariant cues are derived from the appearance of roads. Contextual cues used include horizon line, vanishing point, 3D scene layout and 3D road stages. Moreover, temporal information is included. These cues are robust to different imaging conditions and hence are considered as weak cues. Therefore, they are combined to improve the overall performance of the algorithm. To this end, the low-level, contextual and temporal cues are combined in a Bayesian framework to classify road sequences.

In this paper, real–world scenarios are considered with unpredictable situations and imaging conditions such as extreme shadows, illumination, and complex road shapes (Fig. 1). The proposed algorithm exploits 3D information available in a single image to detect the drivable road surface ahead of the target vehicle (Fig. 2). In this way, we define different 3D contextual cues such as horizon line (road should be below of it), vanishing point (where roads are aimed at), 3D layout (side walks, buildings and sky), and 3D stages (road models).

The novelty of the approach is the introduction of 3D contextual cues and combining them to obtain a diversified ensemble of road cues. In general, combining multiple classifiers is a powerful technique to improve the performance of single classifiers [9, 7]. The improvement is even higher when the method uses diversified cues, *i.e*., cues which are robust or sensitive to different artifacts present in an image. In this way, the proposed method extracts information at scene, image and pixel–level. Further, the proposed method exploits the sequential nature of the data by considering the existing correlation between detected roads in consecutive frames.

Figure 2. The proposed algorithm exploits all the information available in a single image to detect the drivable road surface ahead the target vehicle.

The rest of the paper is organized as follows. First, in Sect. 2, 3D information available from a single image is discussed. Then, in Sect. 3, information available from image sequences is exploited. The framework for combining all the cues is outlined in Sect. 4. Next, in Sect. 5, experiments are presented and results are discussed. Finally, conclusions are drawn in Sect. 6.

2. 3D Road Cues from Still Images

The goal is to combine diversified road features taken from a single image to perform road detection in real–world driving situations. Diversity is imposed by extracting information using two different approaches: top–down (3D scene cues) and bottom–up (pixel classification). The former imposes 3D knowledge and expectations. The latter reinforces the visual measurements improving the accuracy of the results. In this section, three different 3D cues are discussed: horizon line, vanishing point and road geometry. Furthermore, constraints are imposed at pixel–level (color) and mixture of both (layout).

2.1. Horizon Estimation

The horizon line is important information for inferring where the road is located in each image, *i.e*., the road is below the horizon line. To estimate the position of the horizon line, the approach by Sivic *et al*. [10, 11, 12] is used. This method estimates the horizon line by applying non–linear mixtures of linear regressors to the description of an image obtained using gist descriptors [13]. After the horizon line is computed, a pixel–wise confidence map H is generated. A fuzzy labeling approach is used for those pixels close to the horizon line (Fig. 3). Note that detecting roads using the horizon line estimation is robust to lighting variations.

(a) (b) (c)

Figure 3. Horizon line cue computed from a single image. (a) Input image. (b) Estimated horizon line. (c) Pixel–wise road probability map from the horizon line estimation. White pixels are those exhibiting higher probability to be road pixels.

2.2. Vanishing Point

The road area can be detected based on vanishing points [8, 14]. The perspective effect of an image can be exploited to estimate the vanishing point and then, detect

where the road is heading. In particular, the soft voting approach in [8] is used. First, edges are detected using the maximum averaged response of a Gabor filter. Then, the vanishing point is estimated using soft voting [8] and dominant and minor edges are located to infer the location of the different lanes in the road. Finally, a pixel–wise confidence map V is computed assigning higher confidence to those regions on the right side of the main road–lane and lower confidence to those lanes on the left side (Fig. 4). Different confidences are used because left lanes may contain vehicles driving in the opposite direction.

(a) (b) (c)

Figure 4. Given the input image (a), vanishing point is estimated using line segments (b). Then, a pixel–wise confidence map is computed (c). The main lane receives a higher probability than side lanes.

Detecting the road using the vanishing point is robust to global lighting variations, different road types, damaged roads and the presence of other vehicles in the scene. However, it is not robust against curved roads, heavy traffic and when strong shadow edges are present.

2.3. 3D Scene Layout

Another 3D cue is the layout of the scene. The layout is analyzed using three major parts of the image: (1) sky pixels, (2) vertical surface pixels and (3) ground pixels. With these 3D cues the road is limited to ground, non-sky image regions. Further, regions are avoided which are vertically orientated (*i.e.*, buildings, vehicles, pedestrians or any other object present in the scene). The segmentation of the image in these 3D cues is computed by the method proposed by Hoiem *et al*. [15]. This method provides, for each pixel, a label and a confidence map for each class (Fig. 5). Hence, a pixel–wise road confidence map L is obtained considering those pixels which have higher probability of being ground areas than being vertical surfaces or being sky.

Road detection using scene layout is robust to different types of asphalts, lane markings and pedestrian crossings. However, scene-layout for road detection may be sensitive to shadows as the algorithm uses superpixel segmentation.

2.4. 3D Road Geometry

Another important cue for detecting the road is its 3D geometry. This road geometry can be inferred using a scene

(a) (b) (c) (d)

Figure 5. The input image (a) is partitioned in three 3D geometry classes: sky, vertical surface and ground. Using the approach of Hoiem *et al*. [15], a pixel–wise confidence is estimated for each class (sky areas (b), vertical surfaces (c) and ground areas (d)). The whiter the pixels, the higher the probability of a class.

(road) classification algorithm where each class represents typical 3D road geometries such as left turn, straight road and junctions [16].

Road detection via scene classification is performed in two steps. First, the road geometry is obtained using the image features and then the corresponding road probability map for that geometry is selected (Fig. 6). In this paper, a simple one–vs–all classifier approach is used where specific classifiers for each geometry are trained using features extracted from class–specific training sets. Further, a road probability map is computed off–line using the manual segmentation of training images. These segmentations are averaged to obtain a pixel–wise confidence map G containing the probability of the corresponding input image pixel depicting road surface.

Following the approach in [16], images are described using opponent SIFT descriptor with dense sampling [17] (10 pixels sampling grid). K-means is used for dimensionality reduction. This descriptor is invariant to image scale and rotation, and robust to changes in illumination, noise, and minor changes in viewpoint. Then, for each class, a SVM classifier is trained and the one–vs–all approach is used for classifying each new image.

(a) (b) (c)

Figure 6. Geometric information for road detection. (a) Input image, (b) road geometry via scene classification, (c) pixel–wise confidence map which is learned off–line using training images.

This cue provides 3D information of the shape of the road from a single image. It is robust to local illumination effects such as shadows and highlights. Further, the result (detected road) can not degenerate since it is forced to be in one of the 3D geometry classes.

2.5. Road Appearance

Color is a powerful cue for road detection. It provides additional information to the previous shape cues. However, photometric invariant information is needed to provide robustness to lighting conditions. Algorithms based on transformed color spaces (*i.e.*, *HSV* [5, 1] or *rg* [2]) are, to a certain degree still sensitive to lighting variations such as strong shadows and highlights. Therefore, in this paper, the photometric invariant road detection approach in [18] is used. The algorithm exploits the lighting invariant benefits of a physics–based color space [19] combined with a model–based region growing algorithm in a frame-by-frame framework. The model is built at each frame using the normalized histogram of several seeds placed at the bottom part of the image. The output of the algorithm is a pixel–wise similarity C with the road model (Fig. 7).

(a) (b) (c)

Figure 7. Photometric information for road detection. (a) Input image, (b) Illuminant–invariant image, (c) Confidence map is computed based on the similarity between pixels and road model.

This method is robust to varying illumination conditions and shadows and provides pixel–level accuracy. However, the algorithm may fail when images are overexposed and in the presence of lane–markings, pedestrian crossings and heavy traffic.

3. Road Features from Image Sequences

In the previous section, different constraints are computed from still images for the detection of roads. In addition, the sequential nature of the data is considered. Therefore, temporal cues are extracted to constrain the algorithm to newly obtained sequential data. In this way, the algorithm considers the temporal road consistency, *i.e.*, the road ahead of a vehicle can not change drastically from one frame to the next one. This temporal coherence is exploited at two different levels: featured–based (local) and image–based (global).

3.1. Feature–Based

The feature–based approach enables the inclusion of temporal coherence to each feature independently. A straightforward approach is including the temporal dynamics to track points along the horizon location or the vanishing point. To achieve this, two different Kalman filters are used for temporal smoothing of the position of the horizon line (HL) and the vanishing point (VP). HL and VP are the state variables of the filters and the observations are the output of the respective algorithms. Then, under the assumption of constant speed of the vehicle, the dynamics of each filter is $x_{t+1} = Ax_t + w$ and the observation model is $y_t = Hx_t + v$. A is the dynamics matrix, H refers to the linear relationship between the state and its observations and w and v are noise biasing the model and the disturbances corrupting the measurements respectively.

3.2. Image–based

The key idea is to include temporal information, considering that the road geometry in the current frame is not much different from the road in previous frames. However, due to the dynamic nature of data sequences, recent observations should be taken into account more prominently than distant ones. In this way, temporal information is added using time series analysis to predict the expected values of observations rather than simple averages over views. This is specially relevant when other vehicles are present in the scene.

To this end, an exponentially weighted moving average (EWMA) [20, 21] is used to express the dynamic structure of the data (previously detected road). This process is able to cope with sudden changes in the data. Further, EWMA assumes that the road detected in the current frame is correlated (similar) to the road detected in previous frames. EWMA uses a decay factor that weighs the influence of each past result. Thus, more recent results receive higher weights than older ones. Using EWMA, the weights are computed as follows:

$$E[p(x_i = R)^t] = \frac{1}{\sum_{j=1}^{T} \lambda^{j-1}} \sum_{j=1}^{T} \lambda^{j-1} p(x_i = R)^{(t-j)}, \quad (1)$$

where $E[p(x_i = R)^t]$ is the expected probability of a pixel being a road at discrete instant time t (current frame) and $p(x_i = R)^{(t-j)}$ is the probability of a pixel being a road j frames before the frame being analyzed. Further, λ is the decay factor. This factor determines both the degree of weighting of recent observations. A lower decay gives a higher weighting to recent values. Parameter T can be set to infinity since the weighting procedure will rapidly reduce to zero for distant results. Since $0 < \lambda < 1$, $\lambda^n \longrightarrow 0$ when $n \longrightarrow \infty$, the process will eventually place a zero weight on observations far in the past.

4. 3D Scene Cues for Road Detection

In previous sections, different confidence maps are computed based on road cues. In this section, a method is dis-

cussed to merge these confidence maps for combined road classification.

4.1. Feature Combination and Classification

Feature combination is divided in two different parts: per frame combination and temporal adaptation. First, given a single input image I, the probability is obtained for a pixel x_i depicting the road surface R, $p(x_i = R|I)$. Then, from this image, different cues P_j are computed as detailed in Sect. 2. In fact, the five different cues considered are (1) horizon H, (2) vanishing point V, (3) scene layout L, (4) road geometry G and (5) road appearance C: $\mathbf{P} = \{H, V, L, G, C\}$. Then, the confidence maps obtained for each cue is interpreted as the prior–conditional probability of a pixel being road, $p(x_i = R|P_j)$. Further, pixels are assumed to be conditionally independent and that the cues (*i.e.*, $p(P_j)$) are uniformly distributed over their respective domains. Then, the probability of a pixel being road is $p(x_i = R|P_1, P_2, \ldots, P_5)$. Then, Bayes's rule is applied twice to obtain:

$$p(x_i = R|P_1, P_2, \ldots, P_5) \propto \prod_{j=1}^{5} p(x_i = R|P_j) p(x_i = R), \quad (2)$$

where $p(x_i = R)$ is the probability of a pixel to belong to a road. More than 5000 different images (including different scenarios, road widths, and types) have been manually segmented and their ground truth is averaged to obtain $p(x_i = R)$.

In addition, the image–based temporal information is incorporated. To this end, the expected road probability map for the current frame is combined with the estimation of the road for that frame (Sect. 3) using an adaptive model [22]. Thus, prior road results and the current estimated road confidence are integrated pixel–by–pixel into the final probability map as follows:

$$p(x_i = R)^a = (1 - \alpha) E[p(x_i = R)^t] + \alpha p(x_i = R)^t, \quad (3)$$

where $p(x_i = R)^a$ is the (final) probability of a pixel being a road pixel, $E[p(x_i = R)^t]$ is the estimated pixel probability from Eq. (1) and $p(x_i = R)^t$ is the probability of a pixel being road in the current frame using road cues in Sect. 2. Further, α is an adaptation parameter. The lower α, the more persistent the model is. The result is a highly adaptive model, robust to sudden changes and variability in road pixels due to noise, shadows and road texture.

Once the combined confidence map is computed, the classifier assigns a road or background label to each pixel. In this paper, this is achieved by thresholding the confidence maps. An important aspect of this approach is considering only positive examples (road confidences) in the classification step. This is an advantage due to the complexity and diversity of the background class (different scenarios, vehicles, buildings, pedestrian, sky and so on).

5. Experiments

A large scale dataset of different image sequences is used in the experiments. These sequences are acquired using an on-board color camera by the Sony ICX084 CCD sensor (640x840 and 8 bits per pixel) at 15fps. Images are taken at 10 different days, different daytime (e.g. morning, noon and afternoon) and for three different scenarios (urban, highways and secondary structured roads). Thus, images exhibit different backgrounds, different lighting and weather conditions and the presence of other objects such as vehicles or pedestrians (Fig. 8). The dataset is divided into three different non-overlapping subsets, S1, S2 and S3. S1 consists of 500 images containing 10 different road appearances (straight, left turn, right turn, strong left turn, strong right turn, near and distant car in the same lane, near and distant car in the opposite lane and the last one refers to other road appearances such as intersections and traffic circle). This subset is used for training the stage classifier. S2 consists of 5000 images with different road shapes and is used for generating the road prior as described in Sect. 4.1. Finally, S3 consists of more than 10000 images from different scenarios, days and daytime and is used for evaluating the proposed algorithm.

Figure 8. Example images from the database. Top row: images showing highways. Middle row: images showing urban scenarios. Bottom row: images showing secondary structured roads.

Quantitative evaluation is performed on a subset of S3 consisting of 1000 images selected from every 8 frames. All these images are manually segmented to generate the ground–truth. Performance evaluation is provided using pixel–wise measures from which four error measures are computed: quality $\hat{g}$, detection rate DR, detection accuracy DA and effectiveness F, see Table 1. Further, a valid road result index VRI is used. A result is useful ($VRI = 1$), when at least 80% of non road boundary pixels are correctly classified. Otherwise, the result is useless ($VRI = 0$). Boundary pixels are discarded to reduce the inherent error when images are segmented manually to generate the ground–truth [23].

The performance of the proposed method is validated and compared to different road detection algorithms. The first algorithm is based on the vanishing point estimation [8]. The second is based on the layout of the scene [15].

The third is based on the geometry of the road [16]. The last two algorithms are photometric–based approaches: the illuminant–invariant algorithm in [18] and the *HSI* based algorithm proposed in [1] and used in [5]. For fair comparison, the parameters of these algorithms are obtained using a exhaustive learning approach. In this way, a set of images is processed and evaluated using all possible values within the range of each parameter. The proper set of parameter values is the one which maximizes the average performance. Finally, the performance of our method using temporal information is included.

Contingency Table		Ground–truth	
		Non–Road	Road
Result	Non–Road	TN	FN
	Road	FP	TP

Pixel–wise measure	Definition
Quality	$\hat{g} = \frac{TP}{TP+FP+FN}$
Detection rate	$DR = \frac{TP}{TP+FP}$
Detection accuracy	$DA = \frac{TP}{TP+FN}$
Effectiveness	$F = \frac{2PR}{P+R}$

Table 1. Right table describes pixel–wise measures used to evaluate the performance of detection results. These measures are defined using the entries of a contingency table (left).

Example results are shown in Fig. 9. The performance of the algorithms is outlined in Table 2. Further, the dataset is divided in different scenarios (highways, urban like and secondary structured roads, see Fig. 8) and the performance per scenario is outlined in Table 3. This partition provides more insight in the relevance of each final ensemble cue. For instance, the use of vanishing points is relevant for highways, because of the lane markings. However, its performance may be negatively influenced for urban scenarios due to the diversity of edges in these images.

From the results, it can be derived that combining cues improves the overall performance of road detection. This improvement is even higher when temporal information is included. In this case, the improvement is at expense of a slight loss of accuracy in preserving objects present in the scene (vehicles or pedestrians). Finally, the results obtained using the combination framework do not degenerate. A large portion of the road is detected despite the acquisition conditions. There is a decrease in performance in urban like scenarios where the image contains large sidewalk areas. In most cases, these areas have the same appearance as the road.

The algorithm is currently implemented in Matlab code. However, parallel computing can be applied to reach real–time requirements since all visual cues can be executed separately.

6. Conclusions

In this paper, contextual 3D information is used in addition to low-level cues. Low-level photometric invariant cues are derived from the appearance of roads. The contextual cues used are 3D road cues including horizon lines, vanishing points, 3D scene layout and 3D road stages. Moreover, temporal road cues are included. The low-level, contextual and temporal cues are combined in a Bayesian framework to classify road sequences.

Large scale experiments on road sequences have shown that the road detection method is robust to varying imaging conditions, road types, and scenarios (tunnels, urban and highway). Further, using the combined cues outperforms all other individual cues. Finally, the proposed method provided highest road detection accuracy when compared to state–of–the–art methods.

	Complete database				
	$\hat{g}$	*DA*	*DR*	*F*	*VRI*
Vanishing [8]	0.63 ± 0.16	0.70 ± 0.19	0.87 ± 0.13	0.76 ± 0.14	70%
Layout [15]	0.56 ± 0.44	0.59 ± 0.46	0.69 ± 0.42	0.59 ± 0.45	61%
Geometry [16]	0.45 ± 0.04	0.62 ± 0.08	0.74 ± 0.05	0.70 ± 0.04	59%
Illuminant–invariant [18]	0.74 ± 0.34	0.76 ± 0.35	0.88 ± 0.25	0.78 ± 0.34	73%
HSI based [1]	0.68 ± 0.16	0.57 ± 0.14	0.69 ± 0.10	0.63 ± 0.10	65%
Our method[a]	0.75 ± 0.16	0.92 ± 0.07	0.82 ± 0.19	0.85 ± 0.11	82%
Our method[b]	0.84 ± 0.23	0.87 ± 0.3	0.90 ± 0.22	0.86 ± 0.25	92%

Table 2. Performance of different road detection algorithms using the complete database.

[a] using 3D cues from still images.
[b] including features from image sequences.

Acknowledgements

This work was supported by the Spanish Government (projects TRA2007-62526/AUT and Consolider Ingenio 2010: MIPRCV (CSD200700018)) and the Catalan Generalitat (project CTP-2008ITT00001).

References

[1] 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.

[2] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," *Procs. IEEE ITSC*, pp. 939–944, 2006.

[3] C. Thorpe, M. Hebert, T. Kanade, and S. Shafer, "Vision and navigation for the carnegie-mellon navlab," *IEEE Trans. on PAMI*, vol. 10, no. 3, pp. 362 – 373, May 1988.

[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.

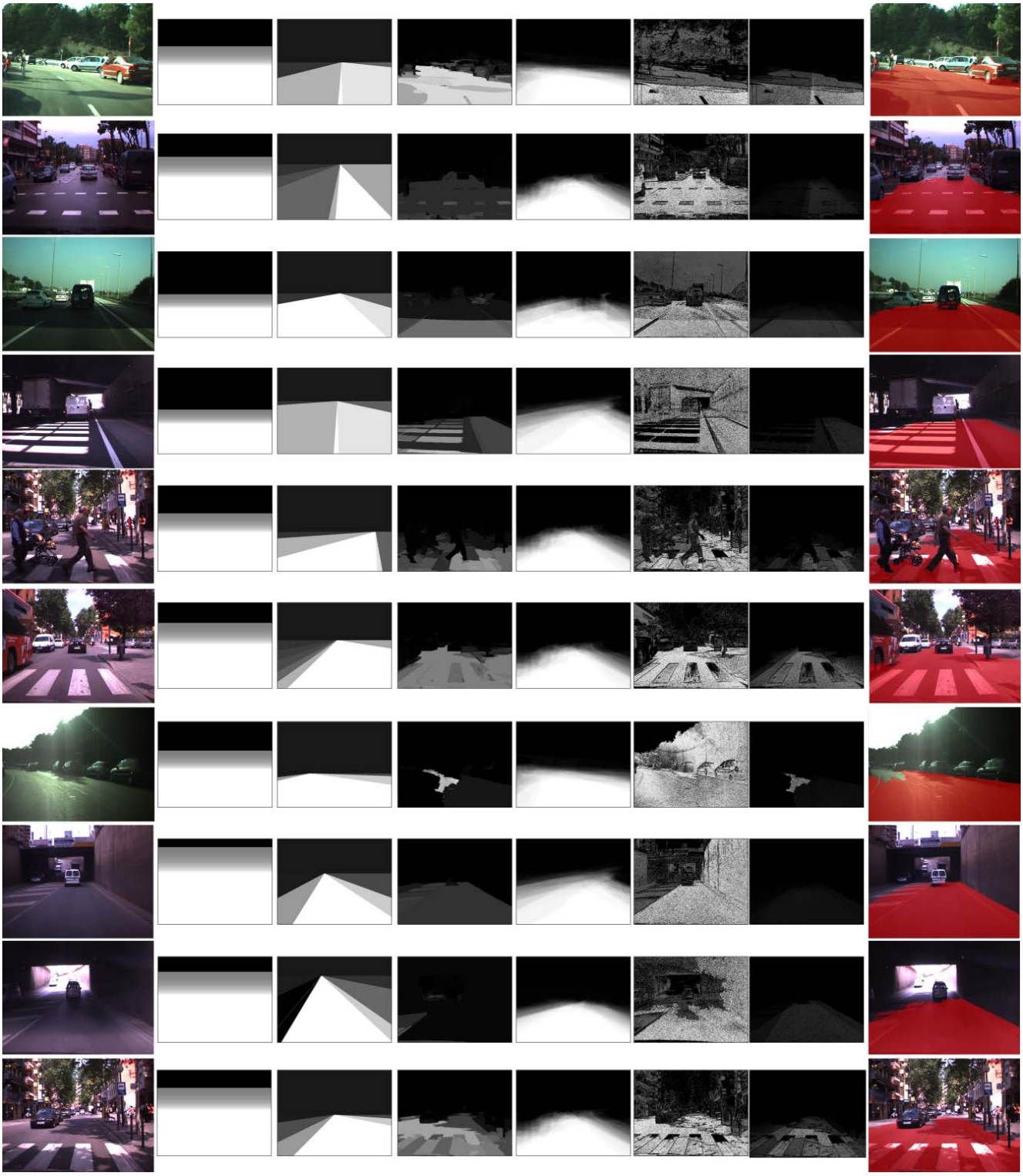

Figure 9. Example results of the proposed road detection algorithm. Given an image (a) the road region is estimated using different cues: (b) horizon line, (c) vanishing point, (d) 3D scene layout, (e) 3D road geometry and (f) road appearance. The pixel probabilities from each cue is then combined using a bayesian framework (g). Finally, temporal information is included. Then, the result (h) is obtained binarizing the probability map using a fixed threshold.

	Highways				
	$\hat{g}$	*DA*	*DR*	*F*	*VRI*
Vanishing [8]	0.70 ± 0.15	0.75 ± 0.18	0.93 ± 0.11	0.81 ± 0.12	81%
Layout [15]	0.37 ± 0.40	0.38 ± 0.41	0.80 ± 0.38	0.42 ± 0.41	40%
Geometry [16]	0.67 ± 0.11	0.68 ± 0.12	0.91 ± 0.15	0.76 ± 0.09	75%
Illuminant–invariant [18]	0.78 ± 0.20	0.87 ± 0.17	0.91 ± 0.16	0.85 ± 0.14	82%
HSI based [1]	0.65 ± 0.11	0.67 ± 0.12	0.95 ± 0.07	0.75 ± 0.16	78%
Our method[a]	0.82 ± 0.18	0.83 ± 0.19	0.99 ± 0.02	0.88 ± 0.11	96%
Our method[b]	0.91 ± 0.04	0.92 ± 0.03	0.99 ± 0.02	0.95 ± 0.01	99%
	Secondary structured roads				
	$\hat{g}$	*DA*	*DR*	*F*	*VRI*
Vanishing [8]	0.62 ± 0.16	0.69 ± 0.19	0.87 ± 0.13	0.74 ± 0.15	77%
Layout [15]	0.61 ± 0.43	0.64 ± 0.45	0.70 ± 0.41	0.63 ± 0.44	66%
Geometry [16]	0.73 ± 0.1	0.78 ± 0.05	0.83 ± 0.08	0.80 ± 0.2	73%
Illuminant–invariant [18]	0.78 ± 0.16	0.91 ± 0.07	0.84 ± 0.19	0.86 ± 0.11	75%
HSI based [1]	0.64 ± 0.11	0.73 ± 0.15	0.88 ± 0.19	0.74 ± 0.09	72%
Our method[a]	0.76 ± 0.34	0.78 ± 0.31	0.89 ± 0.25	0.80 ± 0.33	83%
Our method[b]	0.84 ± 0.28	0.87 ± 0.29	0.91 ± 0.22	0.87 ± 0.29	90%
	Urban Scenes				
	$\hat{g}$	*DA*	*DR*	*F*	*VRI*
Vanishing [8]	0.59 ± 0.15	0.78 ± 0.22	0.76 ± 0.18	0.73 ± 0.13	55%
Layout [15]	0.38 ± 0.42	0.39 ± 0.43	0.69 ± 0.42	0.41 ± 0.44	43%
Geometry [16]	0.60 ± 0.07	0.55 ± 0.12	0.68 ± 0.23	0.63 ± 0.21	38%
Illuminant–invariant [18]	0.65 ± 0.32	0.93 ± 0.35	0.72 ± 0.13	0.80 ± 0.33	69%
HSI based [1]	0.57 ± 0.12	0.70 ± 0.18	0.81 ± 0.16	0.72 ± 0.11	60%
Our method[a]	0.69 ± 0.32	0.75 ± 0.35	0.90 ± 0.11	0.73 ± 0.34	80%
Our method[b]	0.80 ± 0.23	0.88 ± 0.25	0.92 ± 0.09	0.86 ± 0.23	93%

Table 3. Performance of different road detection algorithms differentiated by the scenarios.

[a] using 3D cues from still images.
[b] including features from image sequences.

[5] C. Rotaru, T. Graf, , and J. Zhang, "Color image segmentation in HSI space for automotive applications," *Journal of Real-Time Image Processing*, pp. 1164–1173, 2008.

[6] A. B. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," in *ICCV*, 2003, pp. 273–280.

[7] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[8] H. Kong, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *CVPR '09*. Miami, FL, USA: IEEE Computer Society, 2009, pp. 96–103.

[9] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on PAMI*, vol. 20, no. 3, pp. 226–239, March 1998.

[10] J. Sivic, B. Kaneva, A. Torralba, S. Avidan, and W. T. Freeman, "Creating and exploring a large photorealistic virtual space," in *Procs. of the First IEEE Workshop on Internet Vision, Ancorage, Alaska, USA*, June 2008.

[11] D. Hoiem, "Seeing the world behind the image: Spatial layout for 3d scene understanding," Ph.D. dissertation, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, Aug 2007.

[12] A. Torralba and P. Sinha., "Statistical context priming for object detection," in *ICCV'01*, vol. 1, 2001, p. 763.

[13] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[14] C. Rasmussen, "Grouping dominant orientations for ill-structured road following." in *CVPR'04*, 2004, pp. 470–477.

[15] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 151–172, 2007.

[16] J. Alvarez, T. Gevers, and A. 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.

[17] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *CVPR*, June 2005, pp. 524–531.

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

[19] G. Finlayson, S. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Trans. on PAMI*, vol. 28, no. 1, 2006.

[20] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *CVPR '00*, vol. 2, 2000, pp. 160–167 vol.2.

[21] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, 1999, pp. –252 Vol. 2.

[22] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Trans. on PAMI*, vol. 26, no. 7, pp. 862–877, 2004.

[23] E. D. Gelasca, T. Ebrahimi, M. C. Q. Farias, M. Carli, and S. K. Mitra, "Towards perceptually driven segmentation evaluation metrics," in *CVPRW'04*, Washington, DC, USA, 2004, p. 52.