

General road detection from a single image

Hui Kong^{1,2}, *Member, IEEE*, Jean-Yves Audibert^{1,2}, and Jean Ponce¹, *Fellow, IEEE*

¹Willow Team, Ecole Normale Supérieure / INRIA / CNRS, Paris, France

²Imagine team, Ecole des Ponts ParisTech, Paris, France

Email: tom.hui.kong@gmail.com, audibert@imagine.enpc.fr, ponce@di.ens.fr

Abstract—Given a single image of an arbitrary road, that may not be well-paved, or have clearly delineated edges, or some a priori known color or texture distribution, is it possible for a computer to find this road? This paper addresses this question by decomposing the road detection process into two steps: the estimation of the vanishing point associated with the main (straight) part of the road, followed by the segmentation of the corresponding road area based on the detected vanishing point. The main technical contributions of the proposed approach are a novel adaptive soft voting scheme based on a local voting region using high-confidence voters, whose texture orientations are computed using Gabor filters, and a new vanishing-point-constrained edge detection technique for detecting road boundaries. The proposed method has been implemented, and experiments with 1003 general road images demonstrate that it is effective at detecting road regions in challenging conditions.

Index Terms—vanishing point detection, road detection, soft voting, dominant edge detection.

I. INTRODUCTION

NUMEROUS image-based road detection algorithms have emerged as one of the components of fully automatic vehicle navigation systems [1]. Most of the early systems focused on following the well-paved road that is readily separated from its surroundings. More recently, triggered by the DARPA Grand Challenge [2], a competition between autonomous off-road vehicles, many algorithms have attempted to handle off-road conditions. Although significant advances have been made on specialized systems for detecting individual road types, little progress has been made in proposing a general algorithm to detect a variety of types of roads.

Given a road image as shown in Fig. 1, can the computer roughly determine where the road is? This paper answers this question by proposing a novel framework for segmenting the road area based on the estimation of the vanishing point associated with the main (straight) part of the road. The novelties of this paper lie in the following aspects: (1) In the estimation of texture orientation, we not only compute the texture orientation at each pixel, but also give a confidence to each estimation. The introduced confidence is then incorporated into the vanishing point estimation. (2) Observing that the higher image pixels tend to receive more votes than lower image pixels, which usually results in wrong vanishing point estimation for the road images where the true vanishing point of the road is not in the upper part of the image, a locally adaptive soft-voting (LASV) scheme is proposed to overcome this problem. The scheme uses a local voting region, in which pixels having low confidence texture orientation estimation are discarded. This vanishing point estimation method is quite



Fig. 1. Different types of roads with varying colors, textures and lighting.

efficient because only the selected pixels in the local voting region are used as voters. (3) To segment the road area, a vanishing-point constrained group of dominant edges are detected based on an Orientation Consistency Ratio (OCR) feature, and two most dominant edges are selected as the road borders by combining color cue. This road detection method integrates texture orientation and color information of the road, and it handles well changes of illumination and applies to general road images. In the preliminary version of this paper [3], we only use the OCR feature and a clustering method for road segmentation. We show through empirical results that the road segmentation accuracy is improved by combining the OCR and color features.

II. RELATED WORK

Generally, a road image can be classified into a structured (e.g., a road in urban area) or unstructured one (e.g., a road in rural area). For structured roads, the localization of road borders or road markings is one of the most commonly used approach. Color cue [4], [5], [6], Hough transform [7], [8], steerable filters [9], [10], and Spline model [11], [12], [13] etc. have been utilized to find the road boundaries or markings. The drawbacks of these methods is that they only consistently work for structured roads which have noticeable markings or borders. Methods based on segmenting the road using the color cue have also been proposed, but they do not work well for general road image, specially when the roads have little difference in colors between their surface and the environment. In addition, Laser [14], radar [15] and stereovision [16] have also been used for structured-road detection.

For unstructured roads or structured roads without remarkable boundaries and markings, Alon et al. [17] have combined the Adaboost-based region segmentation and the boundary detection constrained by geometric projection to find the “drivable” road area. However, it needs many different types of road images to train a region classifier, which might be onerous. Reverse optical flow technique [18] provides an adaptive segmentation of the road area, but the method does not work well on chaotic roads when the camera is unstable and the estimation of the optical flow is not robust enough. Stereo cameras [19], [20] are also used to determine terrain traversability. When there is little difference in color between the road and off-road areas, it is hard to find strong intensity change to delimit them. The one characteristic that seems to define the road in such situations is texture. The associated approaches [21], [22], [23] have attempted to define the forward “drivable” image region by utilizing the texture cue. They compute the texture orientation for each pixel, then seek the vanishing point of the road by a voting scheme, and finally localize the road boundary using the color cue. Our approach belongs to this line of research. Although multiple-sensor method [24] can handle unstructured road case, it is beyond the scope of this paper which only uses visual information.

The rest of this paper is organized as follows: a texture orientation estimation at each pixel for which a confidence level is provided (Section III), a voting scheme taking into account this confidence level and the distance from the voting pixel to the vanishing point candidate (Section IV), and a new vanishing-point constrained dominant edge detection technique for finding the boundaries of the road (Section V).

III. CONFIDENCE-RATED TEXTURE ORIENTATION ESTIMATION

Our texture orientation estimation relies on Gabor filters since they are known to be accurate (see [22, Section 2.1]). The kernels of the Gabor filters are similar to the 2D receptive field profiles of the mammalian cortical simple cells and exhibit desirable characteristics of spatial locality and orientation selectivity. For an orientation ϕ and a scale (radial frequency) ω , the Gabor wavelets (kernels, filters) are defined by [25]

$$\psi_{\omega,\phi}(x,y) = \frac{\omega}{\sqrt{2\pi}c} e^{-\omega^2(4a^2+b^2)/(8c^2)} (e^{ia\omega} - e^{-c^2/2})$$

where $a = x \cos \phi + y \sin \phi$, $b = -x \sin \phi + y \cos \phi$ and $c = 2.2$ (octave 1.7 in [25]). We consider 5 scales ($\omega = \omega_0 \times 2^k$, $\omega_0 = 2.1$, $k = 0, 1, 2, 3, 4$) on a geometric grid and 36 orientations (180 divided by 5). These parameters are similar to the ones in [22]. Figure 2 shows the real and imaginary parts of the Gabor kernels.

Let $I(x,y)$ be the gray level value of an image at (x,y) . The convolution of image I and a Gabor kernel of scale ω and orientation ϕ is defined as follows

$$\mathcal{G}_{\omega,\phi} = I \otimes \psi_{\omega,\phi} \quad (1)$$

The convolution result $\mathcal{G}_{\omega,\phi}(z)$ at pixel $z = (x,y)$ has two components, a real part and an imaginary part. To best characterize the local texture properties, we compute the square

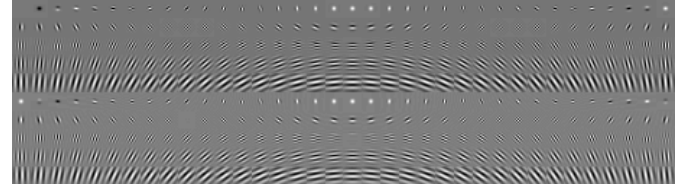


Fig. 2. Gabor kernels with 5 scales and 36 orientations: real part kernels (rows 1 to 5) and imaginary part kernels (rows 6 to 10).

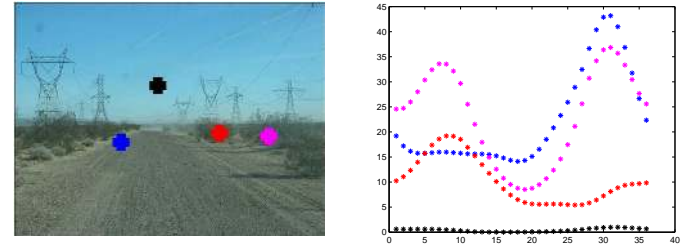


Fig. 3. Left: Four points on which the Gabor complex responses are evaluated. Right: The Gabor complex responses for the four points.

norm of this “complex response” of the Gabor filter for each 36 evenly spaced Gabor filter orientations:

$$I_{\omega,\phi}(z) = \text{Re}(\mathcal{G}_{\omega,\phi}(z))^2 + \text{Im}(\mathcal{G}_{\omega,\phi}(z))^2.$$

The response image for an orientation is then defined as the average of the responses at the different scales (see Fig. 3):

$$R_{\phi}(z) = \text{Average}_{\omega} I_{\omega,\phi}(z).$$

The texture orientation $\theta(z)$ is chosen as the filter orientation which gives the maximum average complex response at that location (the average is taken over the 5 scales):

$$\theta(z) = \text{Argmax}_{\phi} R_{\phi}(z)$$

The second row of Figure 4 shows the images overlaid with subsampled texture orientations estimated using Gabor filters.

From the convolution theorem applied to Eq. (1), we have

$$\mathfrak{F}\{\mathcal{G}_{\omega,\phi}\} = \mathfrak{F}\{I\}\mathfrak{F}\{\psi_{\omega,\phi}\},$$

hence

$$\mathcal{G}_{\omega,\phi} = \mathfrak{F}^{-1}\{\mathfrak{F}\{I\}\mathfrak{F}\{\psi_{\omega,\phi}\}\},$$

where \mathfrak{F} and \mathfrak{F}^{-1} denote the Fourier and inverse Fourier transform, respectively. The use of the fast Fourier transform

$$\begin{aligned} & \mathfrak{F}\{\psi_{\omega,\phi}\}(\xi,\nu) \\ &= \sqrt{8\pi} \frac{c}{\omega} \left(e^{-c^2((\alpha-\omega)^2+\beta^2)/(2\omega^2)} - e^{-c^2(\alpha^2+\omega^2+\beta^2)/(2\omega^2)} \right), \end{aligned}$$

with $\alpha = \xi \cos \phi + \nu \sin \phi$ and $\beta = -\xi \sin \phi + \nu \cos \phi$, allows fast computation of the response image.

Although the above solution for texture orientation estimation has been used by some previous researchers, the estimated texture orientation in this way is not guaranteed to be correct. To provide a confidence level to the texture orientation $\theta(z)$ at pixel z , we seek to a way which evaluates how peaky the function $\phi \mapsto R_{\phi}(z)$ is near the optimum angle $\theta(z)$. Let

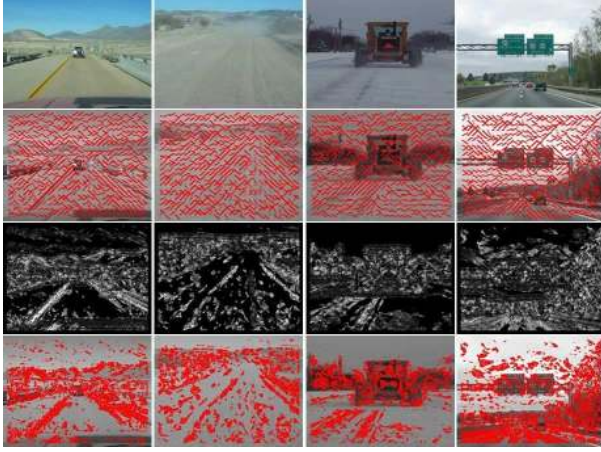


Fig. 4. First row: four road sample images. Second row: images overlaid with texture orientations estimated using Gabor filters. Third row: examples of the confidence map for the texture orientation estimation. The brighter the pixel, the higher confidence the orientation estimation. Fourth row: pixels with confidence larger than 0.3.

$r_1(z) > \dots > r_{36}(z)$ be the ordered values of the Gabor response for the 36 considered orientations (in particular, $r_1(z) = R_{\theta(z)}(z)$). If the global maximum response is significantly different from the other local maximum responses, the texture orientation estimation is reliable, otherwise, it is not. We have found that the local maximum responses usually fall between r_5 and r_{15} (r_2, r_3 and r_4 correspond to similar angles to the optimal one). Therefore, we choose the average of the responses from r_5 to r_{15} as the mean of the local maximum responses. The confidence in the orientation $\theta(z)$ is given by

$$\text{Conf}(z) = 1 - \frac{\text{Average}(r_5(z), \dots, r_{15}(z))}{r_1(z)}$$

We normalize Conf throughout the image to the range of 0 to 1. In our experiments, we discard the pixels having a confidence score smaller than T , and consider the remaining pixels as the “voting” pixels. T can be seen as a threshold put on the normalized confidence score. The optimal T is obtained by tuning T on our test image set, where $T = 0.3$ results in highest vanishing point detection accuracy.

We did not directly use the magnitude of the response of the Gabor filter, since it leads to worse results than the proposed method according to our tests. These negative results are mostly due to high magnitudes of the response in parts of the image that are not related to the road and low magnitudes of the Gabor response in the road area, which often happens with unstructured roads and bright sky.

IV. LOCALLY ADAPTIVE SOFT-VOTING

After having computed the texture orientation at each pixel of the image, one can make these pixels vote to obtain the vanishing point. Precisely, a pixel P for which the texture orientation is the vector \vec{O}_P can vote for all pixels V above P (we consider images in which the road is below the sky) such that the angle $\angle((PV), \vec{O}_P)$ between the direction (PV) and the vector \vec{O}_P is below some fixed threshold δ . This “hard-voting” strategy has been used in [22]. In our experiments, we

notice that this scheme tends to favor points that are high in the image, leading sometimes to large errors in the estimation of the vanishing point. A typical image for which this defect appears is given in Fig.5.

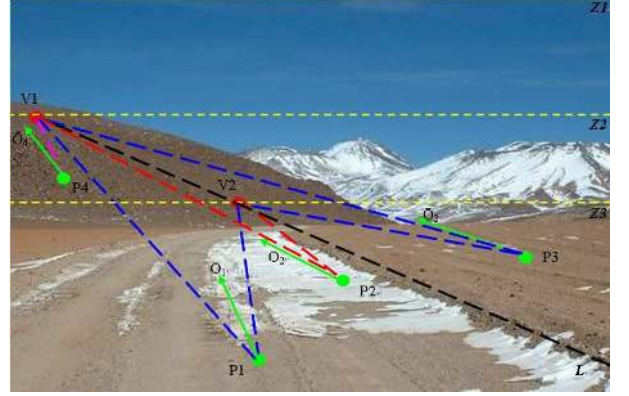


Fig. 5. Illustration of the problem in vanishing point estimation by conventional voting strategy. P1, P2, P3 and P4 are four possible voters. V1 and V2 are two vanishing point candidates (assuming that V2 is the true vanishing point). $\vec{O}_1, \vec{O}_2, \vec{O}_3$ and \vec{O}_4 are respectively the texture orientation vectors of the four voters. The two vanishing point candidates divide the whole image region into three zones, denoted as Z1, Z2 and Z3. Z1 does not vote for both candidates. Both Z2 and Z3 potentially vote for V1 while V2 receives votes only from Z3. Therefore, the higher vanishing point candidates tend to receive more votes than the lower candidates.

To deal with this problem, we propose a soft-voting scheme where the voting score received by a vanishing point candidate from a voter is a value taking into account the distance between the vanishing point candidate and the voter. We treat as a candidate vanishing point each pixel in the top 90% portion of the whole image (although we might only consider fewer by sub-sampling the image), which, to our knowledge, is a realistic assumption for general road images. For each point V of the image, we define a voting region R_V as the intersection of the Gabor response image with a half-disk below V centered at V (see Fig.6). The radius of this half-disk is set to be $0.35 \times \Upsilon$, where Υ is the length of the image diagonal (see our empirical validation part).

Each pixel P inside R_V , for which the texture orientation \vec{O}_P has been confidently estimated (see end of Section III), will vote for the candidate vanishing point V all the more as P is close to V and the orientation of its texture coincides with the direction (PV) . Specifically, we introduce the ratio $d(P, V)$ equal to the distance between P and V divided by

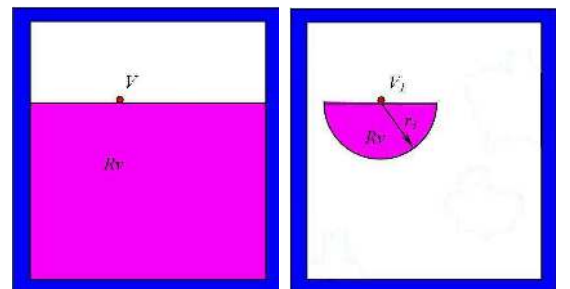


Fig. 6. Left: Global R_V . Right: local R_V . The blue belt in the images is the border pixels excluded from voting owing to the Gabor kernel size.

TABLE I
LOCALLY ADAPTIVE SOFT-VOTING (LASV) SCHEME

1. For each pixel of the input image, compute the confidence in its texture orientation estimation.
2. Normalize the confidence to the range of 0 to 1, and only keep as voters the pixels whose confidence is larger than 0.3.
3. Only the pixels in the top 90% portion of the image are selected as vanishing point candidates.
4. Create the local voting region R_V for each vanishing point candidate, and only the voters within R_V vote for it.
5. Vote for each vanishing point candidate based on Eq. 2.
6. The pixel who receives the largest voting score is selected as the initial vanishing point.

the diagonal length of the image, and let $\gamma = \angle((PV), \bar{O}_P)$ be the angle in degrees between the direction (PV) and the texture orientation at P .

$$\text{Vote}(P, V) = \begin{cases} \frac{1}{1 + [\gamma \frac{1}{d(P, V)}]^2} & \text{if } \gamma \leq \frac{5}{1 + 2d(P, V)} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

It is worth noting that γ itself is independent of $d(P, V)$. However, the threshold of γ depends on $d(P, V)$ so that points that are far away (but still within R_V) are taken into account only if the angle γ is very small (typically less than 3°), while points closer to V will be taken into account up to $\gamma \leq 5^\circ$. For example, if $d(P, V)=0.3$, P votes for V only when γ is smaller than $5/(1 + 0.6)$ (approximately 3°). In contrast, if $d(P, V)=0.03$, P votes for V when γ is smaller than $5/(1+0.06)$ (approximately 5°). In this way, γ 's threshold can also be viewed as a penalty coefficient that penalizes the vanishing point candidates in the top end of the image who has an advantage of receiving more votes than the lower vanishing point candidates. In addition, this also allows to improve the computational efficiency. At the end, the vanishing point is detected as the candidate that receives the largest voting score. The LASV process is briefly described in Table I.

The advantages of the proposed LASV method over the conventional global hard-voting method lie in three-fold when the true vanishing point does not lie at the very top end of the image. First, the soft-voting strategy suppresses the support to the false vanishing point (i.e., those vanishing point candidates above the true vanishing point) by making the voting score far less than one (unless γ is very small). For example, it reduces the support received by $V1$ from those voters in $Z2$ and $Z3$ in Fig.5. Second, it increases the ratio of the support received by the true vanishing point to that received by the higher false vanishing point, e.g., the support to $V2$ is larger than that to $V1$ if $P1$ votes for both $V1$ and $V2$, while the support to $V1$ and $V2$ is equal when using hard-voting method even if $P1$ votes for both $V1$ and $V2$. To discard pixels far away from the vanishing point candidate, or with low confidence in the texture orientation estimation, or with γ not small enough results in a significant computational speed-up. Our empirical results show that LASV is more than five times faster than the slow version of the global hard-voting method [22].

V. ROAD SEGMENTATION

The correctly detected vanishing point provides a strong clue to the localization of the road region. Therefore, we propose a vanishing-point constrained dominant edge detection method to find the two most dominant edges of the road. Based on the two dominant edges, we can roughly segment the road area and update the vanishing point estimated by LASV with the joint point of the two most dominant edges.

In [23], a similar straight road segmentation method is given to detect both road borders simultaneously. It is achieved by optimizing a criterion, which is the difference between the average values of some characteristic (e.g., R,G,B color cues) within the image road region and that characteristic in the region outside the road. It may work when the road and off-road regions have different characteristics. However, it usually fails for both cases where there is little difference in color between road and off-road regions, and where the color is not homogeneous in road region.

We also need to point out the distinction between the road support region segmentation method proposed in [26] and ours. The main difference is that they obtain the middle line of the road by using the imaginary ‘‘road support ray’’. This technique is well adapted to desert (unpaved) roads where there usually is a clear trace left by previous vehicles and these rays exhibit a even distribution. However, it may not work as well on paved roads whose texture is usually sparser, and, therefore, finding the middle line may prove more difficult than road borders. In contrast, our method finds the road borders by optimizing a criterion, which is a combination of a pre-defined feature, called ‘‘Orientation Consistency Ratio’’ (OCR), and a measure related to color cue.

The proposed road segmentation strategy is to find the two most dominant edges by initially locating the first one and the other based on the first one. Because we utilize both texture and color cues, the proposed method exhibits good merits in handling very general road detection tasks, e.g., for some unpaved roads where there is very subtle or no change in colors between the road and its surrounding areas (road covered by snow or desert road), or for some roads where color in road region is not homogeneous (road after rain), or for well-paved roads where painted markings are present.

The definition of ‘‘Orientation Consistency Ratio’’ is given in the top left image of Fig.7: l is a line consisting of a set of discrete oriented points/pixels (the orientation of these points denoted by a black arrow in the figure). For each point, if the angle between the point's orientation and the line's direction is smaller than a threshold, this point is viewed to be orientationally consistent with the line. OCR is defined as the ratio between the number of orientationally consistent points and the number of total points on the line.

We find that the initially estimated vanishing point ($vp0$) coincides with the joint point of a few dominant edges of the road if this vanishing point is a correct estimation, while it usually falls on the extension of one of the most dominant boundaries if it is a wrong estimation. Therefore, we propose to use the initial vanishing point estimation as a constraint to find the first most dominant road boundary. Specifically,

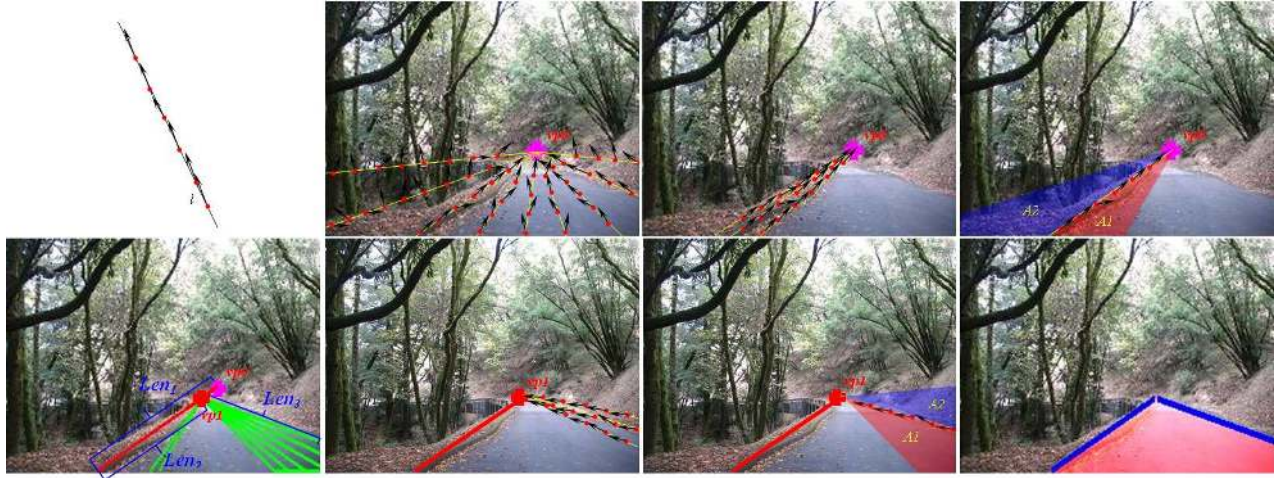


Fig. 7. Illustration of detection of the two most dominant edges. Top row: (1) line segments consisting of discrete oriented points. (2) Some sample rays originating from the initially detected vanishing point $vp0$. (3) Illustration of computing the sum of the OCRs of each ray and its two direct neighbors. (4) Illustration of computing the color difference between each ray's two neighboring regions, $A1$ and $A2$. Bottom row: (1) The first most dominant border (red line), updated vanishing point $vp1$ and the n rays (green lines) which have the largest S_J . (2), (3) and (4) Among the green lines, the second road border is selected as the one which maximizes Eq.(3).

we will search this boundary from a set of imaginary rays which originate from the initially estimated vanishing point. We only consider 29 evenly distributed rays (excluding those rays whose angle relative to horizon is smaller than 20° or larger than 160°) with the angle between two neighboring of them being 5° . The second image in the top row of Fig. 7 shows some of these imaginary rays, with each of them consisting of a set of oriented points whose orientations have been estimated by Gabor filters. Two measures are computed: the sum of the OCRs of each ray and its two direct neighbors (illustrated in the third image of the top row of Fig. 7), and the color difference between the two neighboring regions of each ray (illustrated as the color difference of $A1$ and $A2$ in the fourth image of the top row of Fig. 7). Accordingly, the first most dominant boundary is selected as the ray which maximizes the product of the above two measures,

$$b = \arg_j \max(\text{diff}(A1, A2) \times \sum_{j=i-1}^{i+1} OCR_j) \quad (3)$$

where $A1$ and $A2$ is the two direct neighboring regions on either side of the i -th ray, and $\text{diff}(A1, A2)$ is the color difference of $A1$ and $A2$. Specifically, $\text{diff}(A1, A2)$ is defined as the largest of the color difference for each channel, $\text{diff}(A1, A2) = \max\{\text{diff}(A1, A2)_c | c = R, G, B\}$ ¹, and $\text{diff}(A1, A2)_c$ is given by

$$\text{diff}(A1, A2)_c = \frac{|\text{mean}(A1) - \text{mean}(A2)|}{\sqrt{\text{var}(A1) + \text{var}(A2)}} \quad (4)$$

where $\text{mean}()$ and $\text{var}()$ are the mean and variance of pixel values in a region for a single color channel.

¹The main purpose of including colors in this section is to show that color information helps to boost road segmentation accuracy (compared with our preliminary results by vanishing point [3]). Although we choose RGB as exemplar colors, the use of the other colors, e.g., HSV, might also achieve equivalent or even better performance.

TABLE II
VANISHING POINT CONSTRAINED ROAD BORDER DETECTION

1. Starting from the initially estimated vanishing point $vp0$, construct a set of evenly distributed imaginary rays.
2. For each ray, compute two measures, i.e., OCR and color difference.
3. Select the ray as the first road border if it satisfies the following two conditions:
 - 3.a It maximizes the criterion, i.e., Eq.(3).
 - 3.b Its length is no smaller than one third of the image height.
4. Update the vanishing point estimation:
 - 4.a Regularly sample some points (with a four-pixel step) on the first road border, denoted as p_s .
 - 4.b Through each point of p_s , respectively construct a set of evenly distributed rays, denoted as L_s .
 - 4.c From each L_s , find a subset of n rays such that their OCRs rank top n among the 29 rays.
 - 4.d The new vanishing point $vp1$ is selected from p_s as the one which maximizes the sum of the top n OCRs.
5. Starting from $vp1$, detect the second road border in a similar way as the first border, with a constraint that the angle between the road borders is larger than 20° .

Note that the area of $A1$ and $A2$ is controlled by their wedge angle respectively, which is set to be 20° in our experiment. The red line, E , in the bottom left image of Fig. 7, is detected as the first most dominant edge and its length is denoted as Len_1 . To avoid possible false detection caused by short edges, the smallest Len_1 is set to be one third of the height of image.

Once the first border of the road E is found, we will update the initial vanishing point by looking at the points on E where several dominant edges converge according to the OCR. For this, through each (regularly) sampled pixel J on E , we construct a set of line segments (L_s) such that the angle between any two neighboring lines of L_s is fixed ($\beta = 5^\circ$ in our experiments). We also set the angle between E and any one of L_s is larger than 20° (motivated by the assumption that the vanishing angle between the two road borders is generally larger than 20°). We compute the OCR for each line of L_s (we count the number of lines whose OCR is larger than 0.02 and denote this number by τ), and for each new vanishing

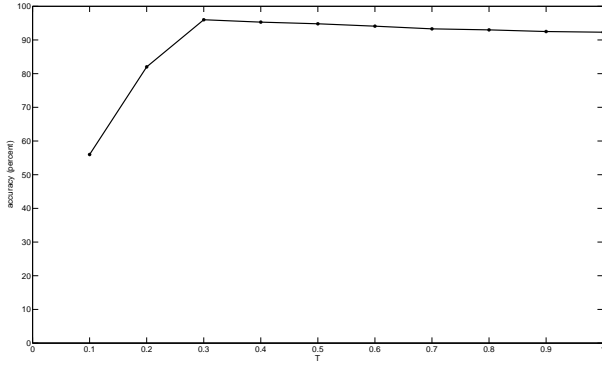


Fig. 8. The effect of T , the threshold set on the confidence of texture orientation estimation, on the vanishing point detection accuracy.

point candidate J , we consider the sum S_J of the top n OCR ($n = \min\{\tau, 8\}$ in our experiments). The green line segments in Fig. 7 are the n lines starting from J receiving the top n OCR. The new vanishing point is then estimated as the point J maximizing S_J . We try other points along E besides the initial vanishing point since the initial vanishing point estimation may not be accurate (i.e., it is not the joint point of the most dominant edges of the roads). Some of the updated vanishing points can be observed in Fig. 12.

From the updated vanishing point and more precisely from the n dominant edges which have voted for it, we deduce the position of the second border of the road in a similar way as explained for the first road border detection. The length of the obtained second most dominant edge is denoted Len_3 and the length of the first dominant edge is updated to Len_2 (see Fig. 7). The smallest Len_2 and the smallest Len_3 are set to be one third of the image height to avoid possible false detections. The process to detect the road borders is summarized in Table II.

VI. EXPERIMENTAL RESULTS

A. Vanishing point estimation

Vanishing point estimation is tested on 1003 general road images. These road images exhibit large variations in color, texture, illumination and ambient environment. Among them, about 430 images are from the photographs taken on a scouting trip along a possible Grand Challenge route in the Southern California desert and the other part is downloaded from internet by Google Image. Some image samples are shown in Fig. 1. All images are normalized to the same size with height of 180 and width of 240. To assess the algorithm's performance vs. human perception of the vanishing point location, we request 5 persons to manually mark the vanishing point location after they are trained to know the vanishing point concept. To remove the effect brought by the subjectivity of each individual in marking vanishing point, a median filter is applied to these human recorded results (for x and y coordinates, respectively) and the median is used as the initial ground-true position. The two farthest manually marked locations to the initial ground-true position are removed as outliers. Finally, the ground-truth location is computed as the mean of the other three locations.

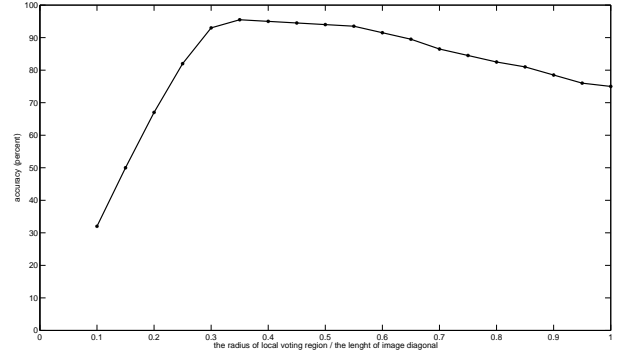


Fig. 9. The effect of the radius of the local voting region on the vanishing point detection accuracy (note that the detected vanishing point is deemed to be correct if the error between the detected vanishing point position and the ground truth one is no larger than 10 pixels).

For brevity, the soft voting strategy defined in Eq. 2 is denoted by “Soft” and the hard voting strategy (by replacing $\frac{1}{1+[\gamma \frac{1}{d(P,V)}]^2}$ with 1 in Eq. 2) is denoted as “Hard”. The voting strategy based on global voting region (left image of Fig. 6) is denoted by “Global” and the one based on local voting region (right image of Fig. 6) is denoted by “Local”. We compare the “Hard” v.s. “Soft” and “Global” v.s. “Local” schemes. We also compare different combination of them with/without introducing the confidence factor.

Before comparing these combination schemes, we firstly introduce how the threshold (T), which is set to the confidence of texture orientation estimation, affects the vanishing point detection accuracy. Using the “Soft”+“Local” strategy (the radius of the local region set to $0.35 \times Y$), we tune T from 0 to 1 with an interval of 0.1, and the result is shown in Figure 8. Note that the detected vanishing point is deemed to be correct if the error between the detected vanishing point position and the ground truth one is no larger than 10 pixels. The optimal vanishing point detection result is obtained when T is set to be 0.3. Similarly, the size of the local voting region also plays a role in detecting vanishing point. In Fig. 9, the vanishing point detection accuracy is obtained based on the “Soft”+“Local” strategy where the radius of local voting region is tuned from 0 to Y and only the image pixels whose texture orientation estimation confidence is larger than 0.3 are used for voting. From Fig. 9, we obtain the best vanishing point detection results when the radius of the local voting region is about $0.35 \times Y$, and this size is fixed in all the subsequent experiments which are based on local voting region.

Figure 10 visually gives us the comparison of vanishing point estimation on some sample images. The estimation using the “Hard” and “Soft” voting based on global R_V are shown in (a) and (b) respectively, while some results using “Hard” and “Soft” voting based on local R_V are shown in (d) and (e) respectively. Figure 10 (c) and (f) shows some samples voted from those image pixels whose confidence score is larger than 0.3. By comparing (a) with (b) and comparing (d) with (e), it can be observed that “Soft” voting scheme is better than “Hard” voting scheme. By comparing (a) with (d) and comparing (b) with (e), we find that local voting region scheme is more accurate than global voting region one. The examples

based on the “Soft” voting from those highly confident texture orientations in the global R_V are shown in row (c), and the estimations based on LASV are shown in row (f). Comparing (c) with (a) and (b), and comparing (f) with (d) and (e), we find that it does improve the vanishing point estimation accuracy by introducing the confidence measure.

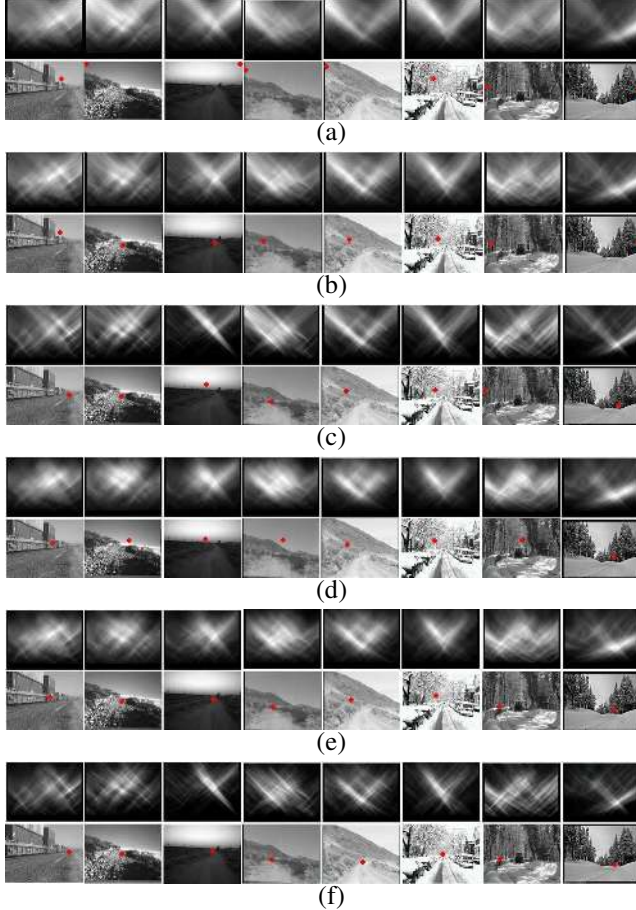


Fig. 10. Comparison of vanishing point estimation based on different combinations: (a) ‘Hard’+‘Global’; (b) ‘Soft’+‘Global’; (c) ‘Soft’+‘Global’ ($T>0.3$); (d) ‘Hard’+‘Local’; (e) ‘Soft’+‘Local’; (f) ‘Soft’+‘Local’ ($T>0.3$)

Figure 11 lists some statistics of the above different combinations. Based on the ground truth positions, we compute the L2 norm distance, i.e., $\text{dist}_{L2} = \sqrt{(x_d - x_g)^2 + (y_d - y_g)^2}$, where (x_d, y_d) is the detection position and (x_g, y_g) is the ground truth position, of the results produced by the above different combinations to the ground truth positions, and put these distances into a 15-bin histogram. The horizontal axis of Fig. 11 represent the dist_{L2} .

If the distance is larger than or equal to 15, it is put into the 15th bin of the histogram. The seven histograms are shown in (a) of Fig. 11. From Fig. 11 (a), we may find that the vanishing point estimation from the pixels with high confidence is much better than the estimation without considering the confidence factor. Local voting-region based method produces more accurate estimation than the corresponding global voting-region based method. Based on these histograms, we also compute the percentage of the images whose error distance is smaller than a number. The best results come from the “Soft” voting based on

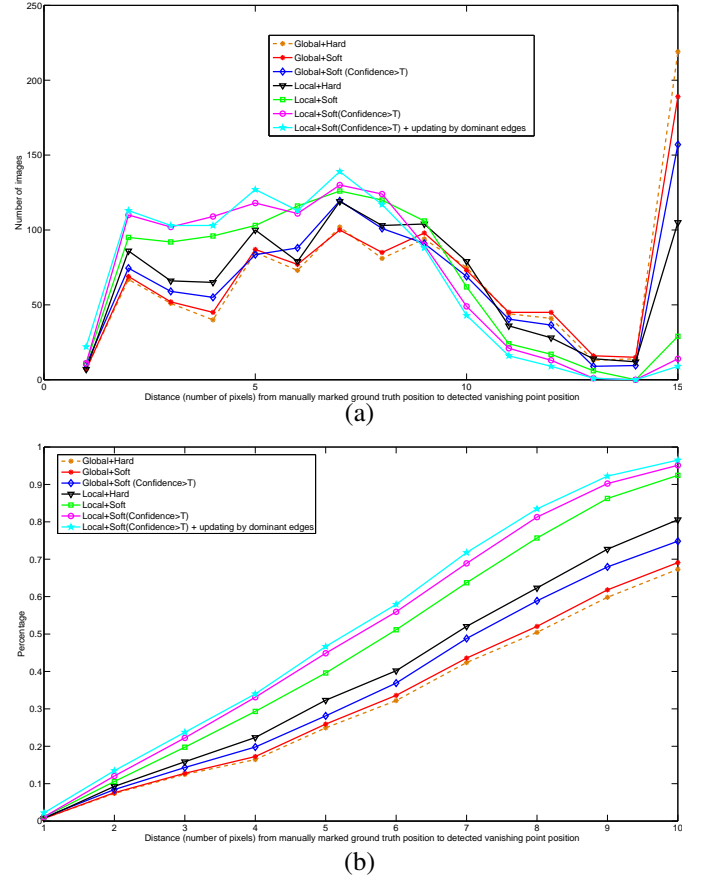
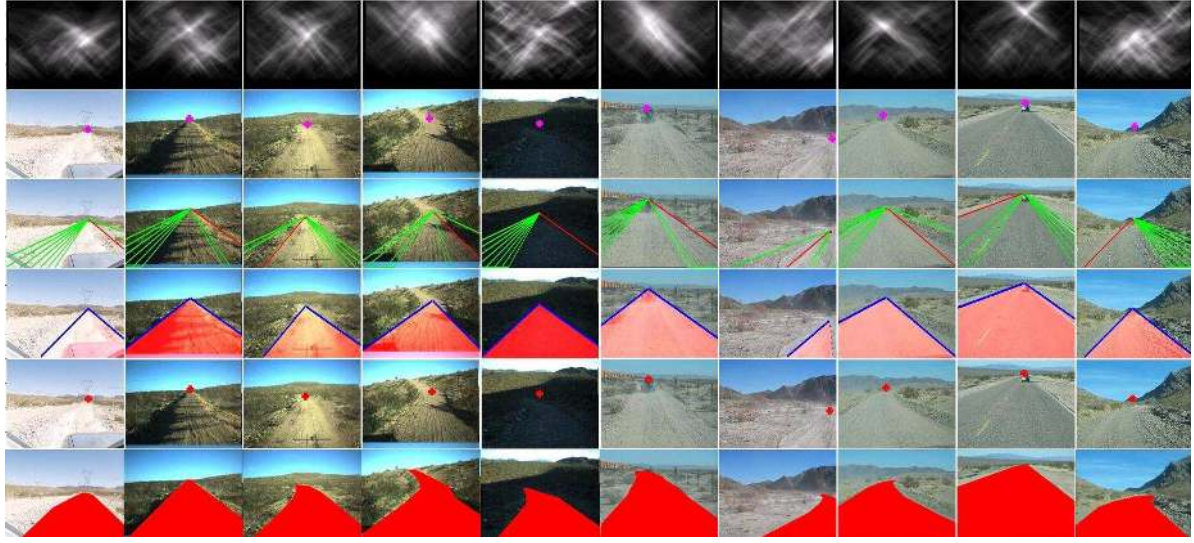


Fig. 11. Comparison of vanishing point estimation accuracy: (a) At a certain error distance, the number of images whose road vanishing point detection is viewed to be correct. (b) The percentage of images whose vanishing point detection error is smaller than a threshold.

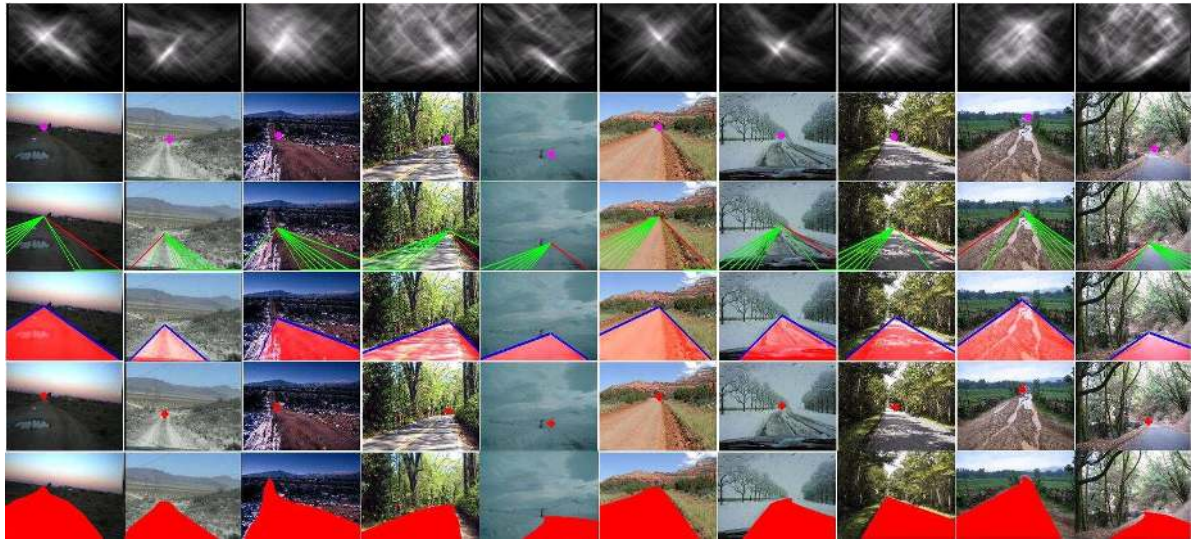
the high-confidence image pixels of the local R_V (confidence value is larger than 0.3) plus updating by the joint point of the two most dominant edges. About 96% of all images have an error distance no bigger than 10 pixels. The method described in [22] belongs to the “Global”+“Hard” scheme. Based on our experiment, our algorithms perform much better: applying the “Global”+“Hard” scheme to our data, totally 112 images produce an error of more than 50 pixels, where 87 of them have very low ground-truth vanishing points. In contrast, such a large error occurs in only 33 images for the weakest variant (Global + Soft) of our method. On average, on our test data, our method gives a 9-pixel instead of 14-pixel error for the method in [22]. Note that, for curved road, the vanishing point produced by our method is located at the joint point of the most immediate straight road borders.

B. Dominant edge detection and road segmentation

Among the 1003 images, about 300 images are from well paved roads with painted markers. Excluding the 430 desert images, the rest images corresponding to the rural roads have no painted lines although part of them are also well paved. For over 90% of the rural roads, the two road borders are detected as the two most dominant edges. For the desert images, the road can be correctly detected as long as the vanishing point estimation is close to the true position. For curved roads, the



(a)



(b)

Fig. 12. Vanishing point detection and road segmentation. (a) sample images from Mojave desert and (b) sample images downloaded from Google image. For (a) and (b), the first rows show the voting images based on ‘Local’+‘Soft’ scheme. The second rows are the initially detected vanishing points based on the voting images. The third rows show the detected dominant road edges based on the OCR and color features (note that the detected red dominant edges correspond to the first most dominant road borders). The fourth rows are the segmented road regions based on the two detected road borders. The fifth rows display the updated vanishing points. The sixth rows are the ground-truth road segmentation.

detected road region is the most immediately drivable area although part of the road surface cannot be fully encompassed by the two dominant edges.

Figure 12 (a) corresponds to the desert road images and (b) comes from the downloaded images. Note that some initially detected vanishing point locations are improved by the two dominant edges. The initial vanishing points by LASV are shown in the second rows respectively. The detected dominant edge candidates are shown in the third rows respectively, where the red lines are the first detected road borders. The two most dominant edges are detected and shown in the fourth rows respectively. The updated vanishing points by dominant edges are shown in the fifth rows. By checking the

vanishing point detection results, we find that some failed cases are caused by extreme illumination conditions (e.g., intensity saturation or strong edge of shadow casted by trees, like the images shown in the seventh and eighth columns of Fig.13). The vanishing point detection tends to fail when the vehicle goes up or down the mountain and there is not enough supporting voting region for the vanishing point (the fifth column of Fig.13). But if there is enough supporting voting region, the vanishing point can be correctly detected even when the vehicle is not running on the flat road (the sixth column of Fig.13). Similarly, the vanishing point detection is accurate during turning the vehicle if there is a large supporting voting region available in the image (the first and

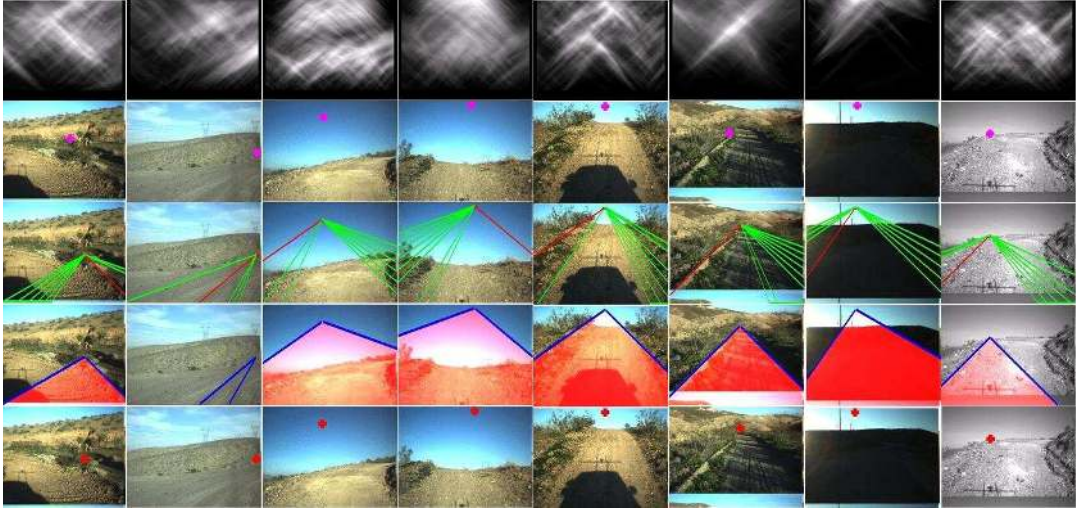


Fig. 13. Examples of failed detections under extreme illumination conditions, or when the vehicle is going up or down the mountain, or making a turn.

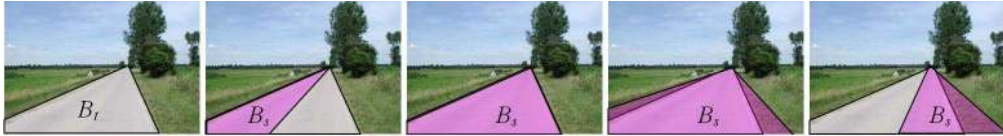


Fig. 14. Illustration of the ‘recall’ concept: the ground truth road segmentation is represented by the gray area in the first image, and the light purple areas in the other images represent the detected road regions. Except for the third image, the ‘recall’ for the other images is smaller than ‘1’.

third columns of Fig. 13), and vice versa (the second, fourth and last column of Fig. 13). To deal with the above failed situations, we might have to seek the other ways instead of only relying on vanishing point detection for road detection. The alternative solution might be the road tracking strategy based on the detected road appearance by our method in previous frames.

To quantitatively show the road segmentation accuracy, we manually labelled the 1003 road images. Some of the labelled road images are shown in the last rows of Fig. 12 (a) and (b). Let B_t and B_s denote the binarized ground-truth and detected road regions of one image respectively, the ‘recall’ is computed as

$$\text{recall} = \frac{B_t \cap B_s}{B_t \cup B_s} \quad (5)$$

where the road regions in B_t and B_s are set to be ‘1’ and the off-road regions are set to be ‘0’. Based on this definition, we may find that the ‘recall’ reaches its maximum value, ‘1’, only when the detected road region coincides with the ground truth one. Figure 14 illustrates the concept of ‘recall’, where the ground truth road segmentation is represented by the gray area in the first image, and the light purple areas in the other images represent the detected road regions. Except for the third image, the ‘recall’ for the other images is smaller than ‘1’.

We change the recall rate from 0 to 1 and calculate the statistics of how many road images are correctly segmented, which is displayed in Fig. 15, where the ‘recall’ is represented in percentage as the horizontal axis. We compare the road segmentation method proposed in this paper with the one in [3]. Because we combine texture (OCR) and color features for road segmentation in this paper, we can observe a large

improvement over [3] where only a clustering method based on OCR features is used.

Our method is efficient and can be run in real time. This is attributed to the sparse number of voters in the local voting region during the vanishing point detection, and the efficient dominant edge detection (the most heavy computation being in the calculation of the OCR for each constructed edge). We run our implementation under Windows OS with a CPU of 1.8GHZ and 1G memory, it takes about 62 seconds for our 1003 240×180 images (i.e., about 17 frames per second). In addition, there is still much room in improving the efficiency. For example, the running speed can be significantly improved by subsampling the vanishing point candidates (e.g., with a even step of 2 pixels), since, in the current version, we consider every pixel as a vanishing point candidate in the top 90% portion of image. For the memory space requirement, our method is economic where the largest memory usage is less than 9M (in texture orientation computation by Gabor filters).

VII. CONCLUSION

A novel framework for segmenting the general road region from one single image is proposed based on the road vanishing point estimation using a novel scheme, called Locally Adaptive Soft-Voting (LASV) algorithm. Then the estimated vanishing point is used as a constraint to detect two dominant edges for segmenting the road area. To remove the effect caused by noisy pixels, each Gabor texture orientation is estimated with a confidence score. In voting, only the pixels of a local voting region whose confidence is high are used, which reduces the computational complexity and improves the accuracy.

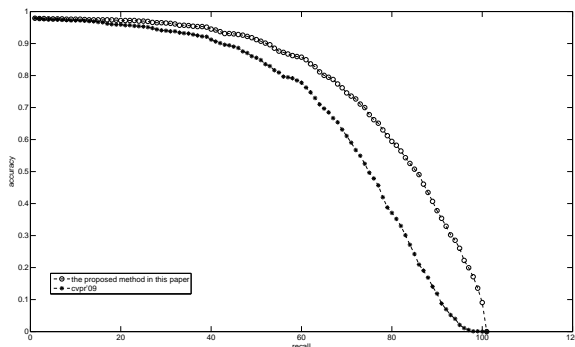


Fig. 15. Road segmentation accuracy: combination of texture and color features improves the accuracy over texture-feature based method.

VIII. ACKNOWLEDGEMENTS

This work was supported by a Postdoctoral Fellowship from Evitech and DGA of France. Thanks to Steven Johnson for providing the images of the Mojave desert.

REFERENCES

- [1] J. C. McCall and M. M. Trivedi, "Video based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. on Intelligent Transportation Systems*, pp. 20–37, 2006. 1
- [2] D. A. R. P. A. (DARPA), "Darpa grand challenge," *Online source: http://www.darpa.mil/grandchallenge*. 1
- [3] H. Kong, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," *CVPR*, 2009. 1, 5, 9
- [4] S.-J. T. Tsung-Ying Sun and V. Chan, "Hsi color model based lane-marking detection," *IEEE Intelligent Transportation Systems Conference*, pp. 1168–1172, 2006. 1
- [5] K.-Y. Chiu and S.-F. Lin, "Lane detection using color-based segmentation," *IEEE Intelligent Vehicles Symposium*, 2005. 1
- [6] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. on Intelligent Transportation Systems*, 2004. 1
- [7] J. B. Southall and C. Taylor, "Stochastic road shape estimation," *ICCV*, pp. 205–212, 2001. 1
- [8] B. Yu and A. K. Jain, "Lane boundary detection using a multiresolution hough transform," *ICIP*, vol. 2, pp. 748–751, 1997. 1
- [9] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *PAMI*, vol. 13, no. 9, pp. 891–906, 1991. 1
- [10] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *IEEE Transactions on Intelligent Transportation Systems*, 2006. 1
- [11] A. Kaskas, R. Husson, and D. Wolf, "Chi-square fitting of deformable templates for lane boundary detection," *IAR Annual Meeting*, 1995. 1
- [12] C. R. Jung and C. R. Kelber, "A robust linear-parabolic model for lane following," *17th Brazilian Symposium on Computer Graphics and Image Processing*, 2004. 1
- [13] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *Image and Vision Computing*, pp. 269–280, 2004. 1
- [14] J. Sparbert, K. Dietmayer, and D. Streller, "Lane detection and street type classification using laser range images," *IEEE Proceedings in Intelligent transportation Systems*, pp. 456–464, 2001. 1
- [15] B. Ma, S. Lakshmanan, and A. O. Hero, "Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion," *IEEE Trans. on Intelligent Transportation Systems*, 2000. 1
- [16] M. Bertozzi and A. Broggi, "Gold: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998. 1
- [17] Y. Alon, A. Ferencz, and A. Shashua, "Off-road path following using region classification and geometric projection constraints," *CVPR*, 2006. 2
- [18] A. Lookingbill, J. Rogers, D. Lieb, J. Curry, and S. Thrun, "Reverse optical flow for self-supervised adaptive autonomous robot navigation," *IJCV*, vol. 74, no. 3, pp. 287–302, 2007. 2
- [19] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," *IEEE International Workshop on Machine Vision for Intelligent Vehicles*, 2005. 2
- [20] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous Robots*, 2005. 2
- [21] M. Nieto and L. Salgado, "Real-time vanishing point estimation in road sequences using adaptive steerable filter banks," *Advanced Concepts for Intelligent Vision Systems, LNCS*, pp. 840–848, 2007. 2
- [22] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," *CVPR*, 2004. 2, 3, 4, 7
- [23] C. Rasmussen, "Texture-based vanishing point voting for road shape estimation," *BMVC*, 2004. 2, 4
- [24] C. Rasmussen, "Roadcompass: Following rural roads with vision + lidar using vanishing point tracking," *Autonomous Robots*, 2008. 2
- [25] T. Lee, "Image representation using 2d gabor wavelets," *PAMI*, vol. 18, pp. 959–971, 1996. 2
- [26] C. Rasmussen and T. Korah, "On-vehicle and aerial texture analysis for vision-based desert road following," *IEEE International Workshop on Machine Vision for Intelligent Vehicles*, 2005. 4

PLACE
PHOTO
HERE

Hui Kong received the PhD degree in computer vision from Nanyang Technological University, Singapore, in 2007. He was a postdoctoral fellow in the Willow team of Ecole Normale Supérieure and the Imagine team of Ecole Nationale des Ponts from 2008 to 2009 in Paris. From 2005 to 2007, he worked as a full-time R&D engineer on digital consumer electronics in the Panasonic Singapore Labs. His research interests lie in computer vision, pattern recognition, and image processing.

PLACE
PHOTO
HERE

Jean-Yves Audibert received the PhD degree in mathematics from the University of Paris VI in 2004. Since then, he is a researcher in the Computer Science department at Ecole des Ponts ParisTech. Since 2007, he is also a research associate in the Computer Science department at Ecole Normale Supérieure in a joint INRIA/ENS/CNRS project. His research interest and publications range from Statistics to Computer Vision, including theoretical properties of learning procedures, boosting algorithms, kernel machines, object recognition, image segmentation, content-based image retrieval.

PLACE
PHOTO
HERE

Jean Ponce received the Doctorat de Troisième Cycle and Doctorat d'Etat degrees in Computer Science from the University of Paris Orsay in 1983 and 1988. He has held Research Scientist positions at the Institut National de la Recherche en Informatique et Automatique (1981–1984), the MIT Artificial Intelligence Laboratory (1984–1985), and the Stanford University Robotics Laboratory (1985–1989), and served on the faculty of the Dept. of Computer Science at the University of Illinois at Urbana-Champaign from 1990 to 2005. Since 2005, he has been a Professor at Ecole Normale Supérieure in Paris, France. Dr. Ponce is the author of over 150 publications, including the textbook "Computer Vision: A Modern Approach", in collaboration with David Forsyth. Dr. Ponce is a member of the Editorial Boards of the International Journal of Computer Vision (for which he served as Editor-in-Chief from 2003 to 2008), the SIAM Journal on Imaging Sciences Foundations, and Trends in Computer Graphics and Vision. He was also an Area Editor of Computer Vision and Image Understanding (1994–2000) and an Associate Editor of the IEEE Transactions on Robotics and Automation (1996–2001). He was Program Chair of the 1997 CVPR and served as General Chair of the year 2000 edition of this conference. He also served as General Chair of the 2008 ECCV. In 2003, he was named an IEEE Fellow for his contributions to Computer Vision, and he received a US patent for a robotic parts feeder.