IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2019, accepted September 18, 2019, date of publication September 27, 2019,
date of current version October 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944244

Fast and Robust Vanishing Point Detection Using Contourlet Texture Detector for Unstructured Road

GUOAN YANG, (Member, IEEE), YUHAO WANG, JUNJIE YANG, AND ZHENGZHI LU
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Guoan Yang (gayang@mail.xjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673314, Grant 61573273, and Grant 11504297, in part by the National Key Research and Development Program Project of China under Grant 2018YFB1700104, and in part by the Project of Henan Suda Electric Vehicle Technology Company Ltd.

ABSTRACT Fast and robust vision-based road detection in an unstructured environment is very challenging. In this paper, we focus on vanishing-point (VP) detection in unstructured roads and propose a response-modulated line-voting method based on a contourlet transform, followed by a voter selection process for VP detection. We first adopt the contourlet transform to estimate the dominant vector for each pixel, including orientation and its relevant response. The estimated dominant vector is then selected by a novel select function to retrieve approximately 40% of the pixels with a reliable dominant vector in the image to vote. Unlike previous methods, this method takes into account the magnitudes of response of the pixels to improve the efficiency of the voting process by suppressing possible interference by extreme and strong textures. The pixels are given a moderate response to vote. Finally, for situations where the road texture is likely to be selected as a criterion for voting by the line-voting scheme, we use this simple and fast scheme to vote for the VP. We conduct experiments on a public dataset of 1,003 different types of natural road images as well as on our own dataset of 400 such images. The results demonstrate that in our dataset, the proposed method is comparable to and outperforms the state-of-the-art methods.

INDEX TERMS Contourlet transform, line-voting method, reliable voter selection, response-modulated, unstructured road, VP detection.

I. INTRODUCTION

Automatic driving technology has been studied for many years. A vital part of it is detecting both well-paved roads and unstructured roads, which are usually in an area with many variations in color, illumination, texture, and weather conditions. Research shows that human vision is selective, acquiring information by efficiently distinguishing small amounts of important information from large amounts of visual stimulus [1]. Researchers are exploring how to give computers a similar ability to human visual perception in order to filter out redundant external signals and effectively represent the infinite information of nature. The use of brain studies and cognitive science in topics, such as the role of visual attention on decision-making and temporal pattern recognition of the lateral temporal lobe, have proven effective for developing computer simulations, particularly in the fields of computer vision and image understanding [2], [3]. Neurophysiological studies show that the visual cortex of primary mammalian has a sparse coding mechanism that can retrieve basic characteristic by a simple cell. [4].

Consequently, vision-based road detection by an efficient image information detector is an important research topic for autonomous driving technology. Over the past few decades, numerous methods have been proposed for detecting well-paved roads, but their application on unstructured roads has not worked well or has even failed. Therefore, researchers in recent years have gradually paid more attention to specialized algorithms to deal with the more challenging unstructured roads. A popular research topic has become the detection of a road's vanishing point (VP), a set of lines in the image plane that corresponds to a set of parallel surface lines in the 3D world space. These lines converge to a common point in the image space [5] and can be used for road following or for guidance through segmented or unstructured roads, an important part of automatic driving technology. For a straight road

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

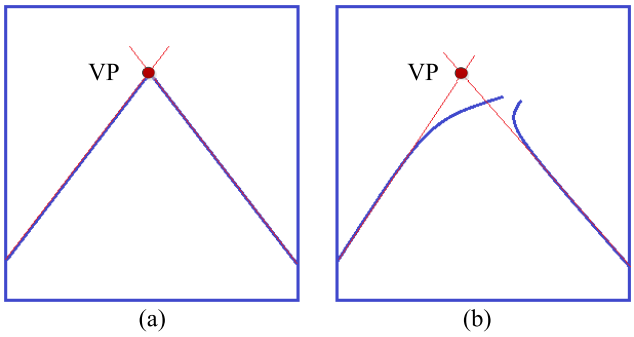

FIGURE 1. Vanishing point in (a) straight roads and (b) curved roads.

segment, as is shown in Figure 1 (a), the vanishing point is obtained as the intersection point of the lines that characterize the lane. For a curved road as shown in Figure 1 (b), the vanishing point is approximated by the lane borders, boundaries, and other features within the vicinity of the vehicle [6].

In this paper, we propose a novel algorithm to detect the VP in images of roads of various conditions, especially unstructured roads. Specifically, we propose first a contourlet texture detector (CTD) to speed up detection of pixels. The pixels with reliable dominant vectors are retrieved and used for vanishing point estimation. In addition, the magnitudes of the reliable dominant vectors of the retrieved pixels are then modulated by a response-modulated line-voting (RMLV) scheme to give each pixel a proper voting weight. The scheme uses an adaptive filter to eliminate the extreme magnitude of the small number and a modulator to repress the relatively salient magnitude of pixel vectors. All the modulated reliable pixels will later vote for the VP by the ling-voting scheme. In contrast to previous texture-based methods, which do not use the response magnitude, we take the texture response of the road pixels into account. This enhances the effect of the real texture to increase detection robustness while effectively suppressing road-independent texture responses. Reliable modulated voters with strong texture responses that are irrelevant to the image of the road are less likely to be selected as voters to the candidate VP than are reliable modulated voters relevant to the road image according to the line-voting scheme. Voters whose orientation is the same as that of the orientation of the line defined by candidate VPs and voters will use the line-voting scheme. The detected VP in the image is the candidate with the most votes.

The rest of this paper is organized as follows. We first review some relevant studies in Section II. The proposed CTD is described in Section III. The proposed vanishing point detection method is then detailed in Section IV. We evaluate the performance of the proposed algorithm in Section V. Finally, we draw some conclusions in Section VI.

II. RELATED WORK

Generally, there are two types of road images: structured roads and unstructured roads. An image of a structured road has distinctive edge contrast and can be divided into regions according to color contrast, texture distribution, or some other priori-known information. Because of these features, existing VP detection algorithms for structured roads can be classified into two main categories: edge-based and region-based. Edge-based methods, for example, that of Wang et al. [7], used the spline model; Rother [8] used Gaussian sphere mapping; Tuytelaars *et al.* [9] used the cascaded Hough transformation; and Liu *et al.* [10] designed a model motivated by a biological visual cortex for road detection. Other effective ways to find road boundaries and markings are B-snake [11], Hough transform and K-means [12], or steerable filter banks [13], [14]. More recently, Ding *et al.* [15] used the estimation envelopes of vertical lines to group straight lines by a path perspective triangle and line-length limits. Li *et al.* [16] proposed a priori-known dark-based image segmentation method. Wu *et al.* [17] segmented a road image by a line segment detector [18]. Region-based approaches [19]–[21] have classified the environment mainly by searching for similar structures and repeating patterns to determine the VP. In recent years, Alvarez *et al.* [22] proposed a machine learning-based method that used 3D road cues to learn and update the road region discriminator. Wang *et al.* [23] proposed a road boundaries region estimation-based method that was robust at detecting shadows and complex environments.

However, the above approaches failed to handle such sophisticated situations as unstructured roads with ruts, tire tracks left by vehicles, or complex road environments, blurred edges, and similar coloring. Those situations hamper the detection of apparent road cues. To deal with road images with intricate information, many specialized unstructured-road methods have been proposed. Huang *et al.* [24] proposed a hue, saturation, and value space and road feature-based method. Lookingbill *et al.* [25] proposed an optical flow and self-supervised learning method to determine the drivable region. Alvarez *et al.* [26] used a convolutional neural network for road scene segments. Li *et al.* [27] used a back-propagation neural network that learned color features to classify the pixels and heuristically fit the boundaries of the lanes. Those methods were either not robust enough or were time consuming in practice.

To better deal with complicated unstructured road images, textural cues can be used to detect the VP. The basic premise behind this method is that the textures in the image appear to converge into the VP. Accordingly, a Gabor filter is used to estimate the texture information to detect the Rasmussen [5] used 72 oriented Gabor filter banks to precisely estimate the orientation of each pixel and adopted a global hard-voting scheme to vote for the VP. The pixels of the image were first filtered by the Gabor filter banks, and the orientation corresponding to the pixel with the maximal response was chosen as the dominant orientation of that pixel. Then the VP of the road was the pixel with the most votes as selected by each pixel. However, the global hard-voting scheme favored the uppermost pixel in the image, and the voting process was redundant. To overcome the drawbacks of the hard-voting scheme, Kong *et al.* [28] proposed a confidence-rated function to select the candidate voters and a soft-voting scheme

to take into account the distance and the orientation between voters and the Miksik [29] expanded the Gabor wavelet into a linear combination of Haar-like box functions to speed the voting process, albeit with a loss of accuracy. To select voters with reliable texture orientation, Moghadam *et al.* [30] used the joint activities of only four Gabor filters to estimate the dominant orientation of each pixel by an optimal local dominant orientation method. Shi *et al.* [6] proposed a particle filter to reduce misidentification probability and computational complexity. In relation to the Gabor filter banks, Kong *et al.* [31] proposed a new generalized Laplacian of a Gaussian (gLoG) filter as a substitute for Gabor filter banks. Their filter was more accurate in estimating the texture information. Yang *et al.* [32] improved the Weber local descriptor [33] to obtain salient representative texture and orientation information about the road area. Our previous work proposed a contourlet based information processing mechanism of a vision nerve cell [1]. To improve the accuracy and time consumption of the previously discussed methods, we propose a texture-based RMLV method to estimate the dominant vector by a contourlet transform followed by a voter selecting process, a response modulation process, and a line-voting scheme to implement fast and robust VP detection.

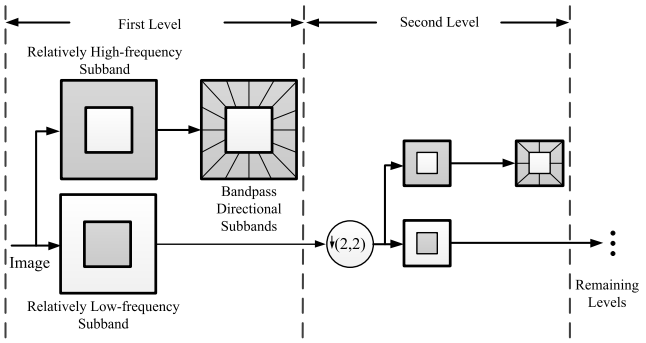

FIGURE 2. Two-level combined multiscale and multidirectional decomposition process.

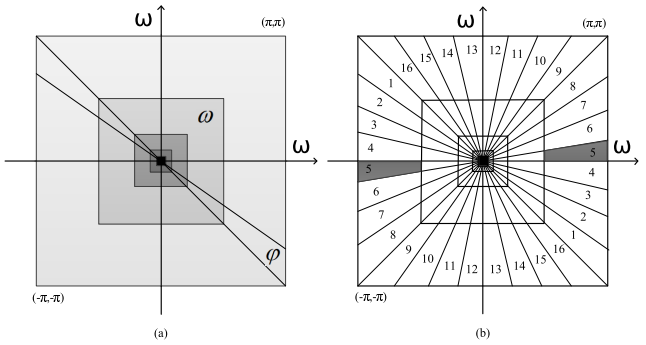

FIGURE 3. Orientation and frequency decomposition. (a) Four levels and 16 orientations decomposed by the contourlet transform. (b) Frequency decomposition of the image I (p) by a 4-level contourlet transform with 16 orientations at 4 relatively high scales.

III. CONTOURLET TEXTURE DETECTOR

This section describes our CTD, which consists of two major steps: (1) estimation of the dominant vector, including dominant orientation and its relevant response; and (2) detection of pixels by the dominant vector.

A. DOMINANT TEXTURE VECTOR ESTIMATION

A contourlet transform [34], in contrast to a wavelet transform, is known as an accurate feature extraction method with directionality and anisotropy, and it exhibits desirable orientation selectivity and spatial locality. The performance value in detecting the image by contourlet transform was proven by our former work, which proposed a contourlet-based information processing mechanism of a vision nerve cell [1]. The contourlet transform uses a Laplace pyramid frame and iterated directional filter banks (DFBs) to implement multiscale and directional decomposition and reconstruction of an image. Figure 2 shows a two-level multiscale and multidirectional contourlet decomposition process using a combination of a Laplace pyramid and a DFB. In the first level, the image is first decomposed into a relatively low-frequency subband and a relatively high-frequency subband by a See-May Phoong, Chai W. Kim, P. P. Vaidyanathan, Rashid Ansari (PKVA) wavelet filter. Then, the relatively high-frequency subband is decomposed into 2^n directional subbands by the DFB using the PKVA wavelet filter. In the second level, the relatively low-frequency subband in the first level is fed into the second level iteratively after down-sampling. The subsequent operations are the same as those of the first level. The decomposition of the remaining levels can be completed by the same process as that used in the second level. Figure 3 (a) shows the four levels and 16 orientations decomposed by the contourlet transform. Images can be decomposed into 65 subbands, where ω is the scale from a central low-frequency subband to surrounding relatively high-frequency subbands ($\omega = 0, 1, 2, 3, 4$). The φ is one of the 16 texture orientation numbers of each subband; the sequence of the orientation numbers is shown in Figure 3 (b). This leads to a superb frequency and orientation selectivity and can be reconstructed by a process that is the reverse of the process shown in Figure 3.

In this study, we used the contourlet transform to estimate the dominant texture vector at each pixel, including the dominant orientation and its relevant response. To estimate the dominant orientation $\phi(p)$ and its corresponding response $E(p)$ at each pixel location $p(x, y)$ in the image, the gray image $I(p)$ was decomposed by a four-level contourlet decomposition process. Consequently, as shown in Figure 3 (b), the image $I(p)$ was decomposed into one scale of the lowest-frequency subband S_0 (the central black square in Figure 3 (b)) and four scales and 16 orientations in each scale of relatively high-frequency subbands $S_{(\omega,\varphi)}$) ($\omega = 0, 1, 2, 3, 4, \varphi = 1, 2, \ldots, 16$). This would maintain a multiple resolution on scale and orientation for the texture in the image. To ensure a reliable estimation of texture response $E_{(\omega,\varphi)}$ for each subband $S_{(\omega,\varphi)}$ with orientation number φ if the edge of each subband $S_{(\omega,\varphi)}$ is aliased with adjacent subbands, we chose the middle angle as the texture orientation

for the corresponding subband. For each texture orientation number φ, we have

$$\phi_\varphi = (\varphi - 0.5) \times \frac{180^\circ}{16} \tag{1}$$

where ϕ_φ is the real texture orientation for the texture orientation number φ. For example, as shown in Figure 3 (b), the subband $S_{(4,5)}$, where $\omega = 4$ and $\varphi = 5$, had the real texture orientation $\phi_5 = (5 - 0.5) \times \frac{180^\circ}{16} = 50.625^\circ$. Eventually, we obtained subband S_0 and 64 subbands $S_{(\omega,\varphi)}$ with the texture orientation ϕ_φ.

The texture response $E_{\phi_\varphi}(p)$ for each pixel $p(x, y)$ of each texture orientation ϕ_φ was the reconstruction of the subbands $S_{(\omega,\varphi)}$ of different scales with the same texture orientation (the response of subband S_0 and subband $S_{(\omega,\varphi)}$ of the other texture orientation was set to 0). From the reconstruction process, we have

$$|E_{\phi_\varphi}(p)| = C^{-1}(S_{(1,\phi_\varphi)}, S_{(2,\phi_\varphi)}, S_{(3,\phi_\varphi)}, S_{(4,\phi_\varphi)}) \tag{2}$$

where C^{-1} is the reconstruction process of the contourlet transform as the reverse of the process shown in Figure 2. Thus, we have decomposed each pixel p of the image into 16 texture orientations with angle $\phi_\varphi(p)$ and a corresponding absolute response $|E_{\phi_\varphi}(p)|$. Considering the aliasing effect between subbands, a texture orientation number of the pixels through contourlet transform greater than 16 would not maintain a reliable estimation of each subband. From the frequency spectrum of the filters, the subband decomposition of 32 orientations is narrower and steeper than the subbands with 16 orientations. Consequently, the subbands of 32 orientations are more coefficient in the transition zone, leading to more inaccuracy coefficients in the subbands of 32 orientations. Thus, a texture orientation number of decomposition greater than 16 has less accuracy coefficients according to the aliasing between subbands and is more coefficient in the transition zone. Moreover, if a texture orientation number of the decomposition is less than 16, it would not guarantee a good angular resolution of texture orientation for each pixel, which is further discussed in Section V.

The vector of each pixel p at texture orientation ϕ_φ is defined as

$$V^{\phi_\varphi}(p) = |E_{\phi_\varphi}(p)|e^{j\phi_\varphi} \tag{3}$$

where ϕ_φ ($\varphi = 1, 2, \ldots, 16$) is the texture orientation, and $|E_{\phi_\varphi}(p)|$ is the relevant absolute response to ϕ_φ. The dominant vector $V^{\phi_d}(p)$ is the vector with the strongest texture response $|E_{\phi_d}(p)|$ and its relevant orientation ϕ_d.

B. DETECTION OF PIXELS

In earlier methods, to accurately select pixels with reliable orientations, the pixel that was chosen as the candidate VP voter had an orientation corresponding to the strongest response across all orientations with high confidence. The candidate VP voter could also be chosen from pixels with orientations summed by two linearly independent pseudo vectors whose orientation was the orientation of the two strongest responses and whose two responses were the difference between the two strongest responses and other responses individually. However, this caused considerable time consumption without ensuring the accuracy of the orientation.

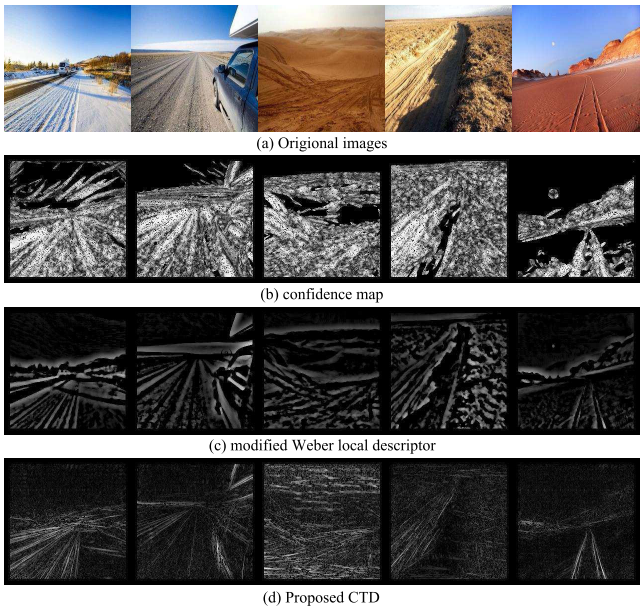

FIGURE 4. Comparison of different image texture detectors.

To overcome these drawbacks, we propose a novel detection method on each pixel. Each pixel whose vectors, adjacent to the dominant vector, has an approximate response to the dominant vector will be considered as the unreliable pixel. The remaining pixels will be considered as reliable pixels with reliable dominant vectors.

The detect function of each pixel p is defined as

$$D(p) = \begin{cases} 0, & \text{if } \dfrac{V^{\phi_d}(p) - V^{\phi_{(d\pm1) \bmod 16}}(p)}{V^{\phi_d}(p)} \le 0.3 \\ 1, & \text{others} \end{cases} \tag{4}$$

where $V^{\phi_d}(p)$ is the dominant vector at pixel p, $V^{\phi_{(d\pm1) \bmod 16}}$ is the left and right vectors adjacent to the dominant vector. When the ratio of the difference between the dominant vector and the left or right adjacent vectors to the dominant vector is less than T, the pixel p is considered as an unreliable pixel and will be eliminated from the set of voters. T is set to 0.3 experimently, which will maintain a better pixel selectivity and guanrantee the remaining pixels have enough information to the image for voting process.

From the detect function, we have

$$I_{CTD}(p) = I(P) \times D(p) \tag{5}$$

where $I_{CTD}(p)$ is the detected pixels by CTD, which will be chosen as a reliable voter to vote for the VP.

Figure 4 compares various reliable pixels obtained by Kong *et al.* [28] Gabor based detector, Yang *et al.* [32] modified Weber local descriptor, and our CTD. The previous methods, especially Kong's method, contained more pixels irrelevant to the real texture. An improvement on this,

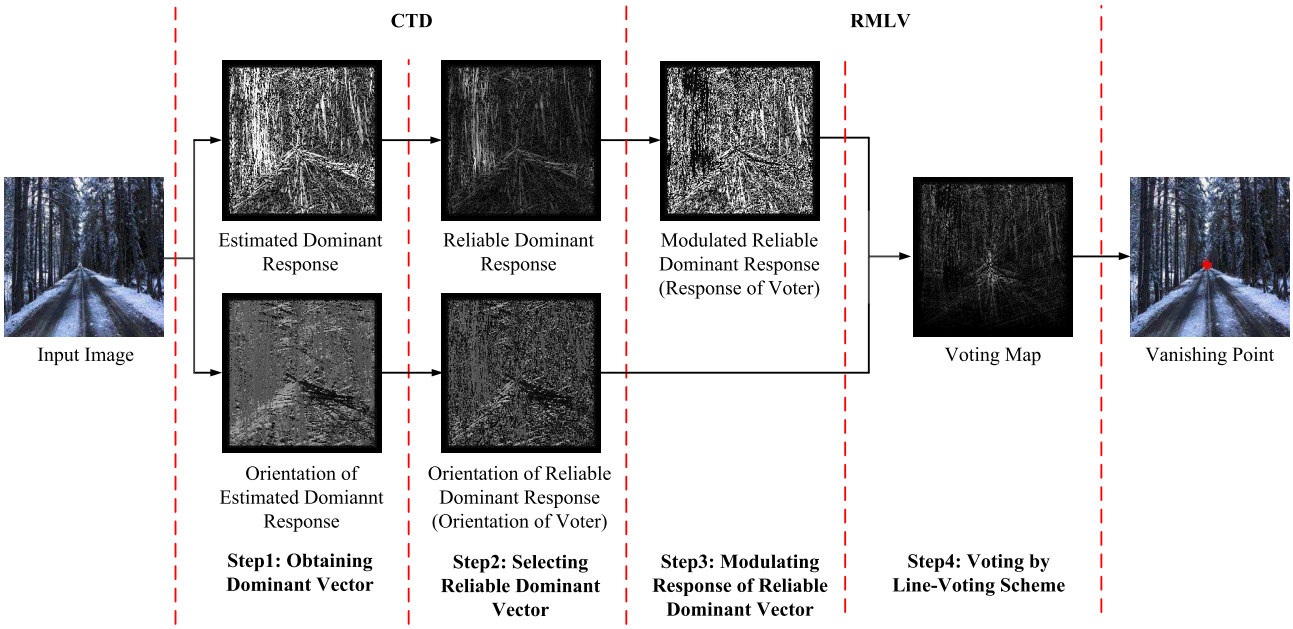

FIGURE 5. Proposed vanishing point approach.

our detected method eliminated the most pixels without apparent texture information and that contained little important image information. The extreme texture information that could interfere with the voting process will be eliminated by our proposed response modulate method introduced in section IV. Thus, we can use fewer pixels with different orientation as the voters, saving much time.

IV. VANISHING POINT DETECTION METHOD

In this section, we describe our VP voting method, which consists of two steps: (1) response modulation of the dominant vector of detected pixels; and (2) VP voting. Figure 5 shows an overview of the framework.

A. RESPONSE MODULATION OF RELIABLE VOTER

The previous voting scheme does not directly adopt the response of each pixel because that could lead to worse results if a high response in parts of the image is unrelated to the road area. To deal with this problem, we further modulated the response of the reliable dominant vector to eliminate the strong influence of extreme texture in the image.

The response $|E_R(p)|$ of the reliable dominant vector is first normalized to the range of 0 to 1 and divided into 256 intervals. Then, the number of the response $|E_R(p)|$ in each interval is counted.

The modulation progress can be defined as follows:

$$|E_R(p)| = \begin{cases} 0, & \text{if} |E_R(p)| > |E_R(p)|_{T\%\times N} \\ |E_R(p)|_{0.8N}, & \text{if} |E_R(p)|_{0.8N} < |E_R(p)| < |E_R(p)|_{T\%\times N} \end{cases} \quad (6)$$

where N stands for the maximum statistical number of each interval. $|E_R(p)|_{0.8N}$ denotes the median response with statistical number 80% of N, and $|E_R(p)|_{0.2N}$ denotes the median response $T\%$ of N, which is larger than the response with statistical number N. To retain more road clues here, and as a tradeoff between keeping valid texture and removing noise, T is set to 20, experimentally, which is discussed further in Section V.

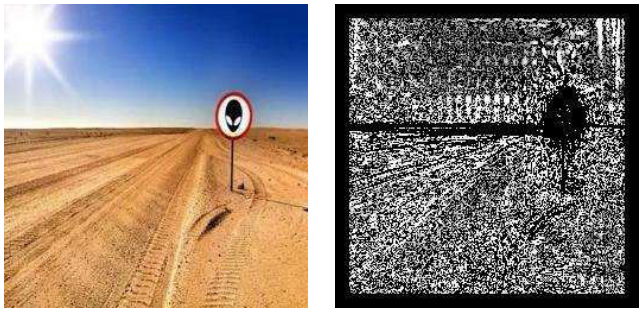

FIGURE 6. A road image and a modulated response. (a) Road image with extreme texture unrelated to the road. (b) The modulated response of image (a).

The response $|E_R(p)|$ larger than the response $|E_R(p)|_N$ with maximum statistical number was modulated by two steps: (1) the response $|E_R(p)|$ larger than $|E_R(p)|_{0.2N}$ was set to 0, which effectively removes the interference by extreme texture unrelated to the road with small numbers of selected pixels; and (2) the response $|E_R(p)|$ larger than $|E_R(p)|_{0.2N}$ and less than $|E_R(p)|_{0.8N}$ was set to $|E_R(p)|_{0.8N}$. This can, to some extent, suppress the influence of the large response providing a moderate response to vote on.

Figure 6 (b) shows that as a consequence of the modulation process, a street sign with extreme texture in contrast to the rest of the image is removed, and the remaining strong texture related to the road is appropriately modulated to gain a fast

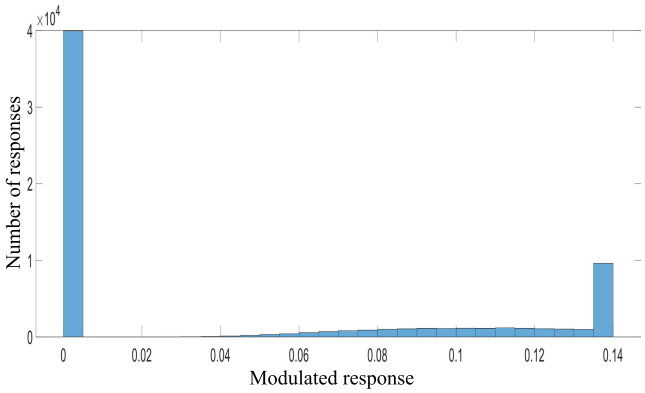

FIGURE 7. Histogram of the image in figure 6.

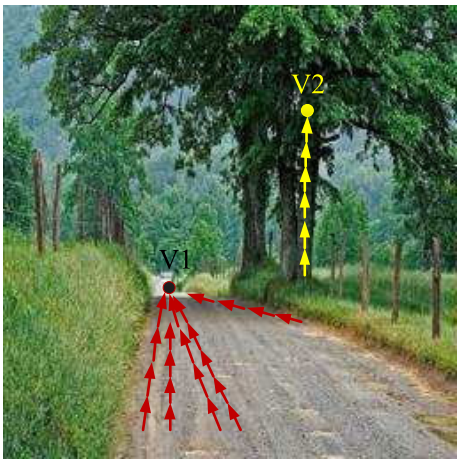

FIGURE 8. The ground truth vanishing point V1 receives more votes from the response to the road boundaries, road markings, and tire tracks left by previous vehicles (red arrows) than the possible interference vanishing point V2 receives from the voters unrelated to the road (yellow arrows).

and accurate result in the voting process. Using the process above, we can achieve pixel $p_{(CTD,M)}$ with a reliable dominant vector, with its orientation and modulated response, to vote for the VP according to the voting scheme introduced in Section IV.

Figure 7 shows an example of the histogram of the modulated response for the image in Figure 6 (a). From the histogram, we can see that almost 4×10^4 pixels were eliminated. The nearly 40% remaining pixels will vote for the VP, which greatly saves time.

B. VANISHING POINT VOTING

After the modulation of the response, the reliable modulated voters $p_{(CTD,M)}$ can be used to vote for the VP in an adopted line-voting scheme. The previous method [28] took the voters closely oriented to the candidate VP into account to slightly improve the accuracy of the VP detection. However, this cost much more time than did the line-voting scheme. The simple line-voting scheme allows all the voters with the same orientation to the line defined by voters and candidate VP vote, regardless of the distance between the voters and the candidate VP. In contrast, the distance-related voting scheme caters to the locally strong off-road texture.

Consequently, as is shown in Figure 8, in the line-voting scheme, when pixels with strong responses irrelevant to the road in the image are less likely to be selected as voters than pixels relevant to the road, all pixels in the image will be voted by the selected voter introduced above according to their response and orientation. Eventually, the pixel with the maximum accumulated responses is treated as the estimated VP.

V. EXPERIMENTAL RESULTS AND ANALYSIS

To assess the performance of the proposed VP detection method, we tested the method on 400 road images downloaded from the Internet using Google Images and from a well-known public road image dataset [31], consisting of 1003 images. The images vary in color, illumination, texture, and surrounding weather conditions. All images were normalized to the size of 256 × 256 because the contourlet transform can decompose only square images.

To evaluate the performance of the algorithm, we invite 20 volunteers aware of the concept of the VP to manually mark the location of the VP based on their perceptions. We use a median filter to calculate the median of the manually marked vanishing point (for x and y coordinates) as the initial ground truth position. To remove the effect of the subjectivity of each individual in marking the vanishing point, the five farthest manually marked positions from the initial ground truth position are removed. The ground truth is calculated as the mean of the other 15 locations.

Figure 9 and Figure 10 show a variety of images from our dataset and the public dataset, including structured roads, unstructured roads, and roads with various illuminations and weather conditions. Each image is overlaid with its estimated VP, shown as a red dot. The first, third, and fifth rows are outputs of the proposed approach (black dots are the ground truths and red dots are the estimated VP locations) and the second, fourth, and sixth rows are the voting maps. The images show that the proposed method successfully estimates each VP through a single image.

We compare our approach with some state of the art texture-based vanishing point detection methods such as Rasmussen [5], Kong *et al.* (Gabor) [28], Kong (gloG) [31], Moghadam *et al.* [30], Yang *et al.* [32], respectively. Figure 11 shows some images with estimated vanishing points in many challenging unstructured road images. We can see that compared with our proposed method, the other methods cannot deal well the images having ambiguous texture in front of the road or images with little or complicated and multidirectional textures, such as roads with ruts and vehicle tire tracks. The possible reasons are that the contourlet, tasked with transforming and selecting the texture to be the reliable voter, effectively filters out much of the dominant texture from the image. The proposed RMLV scheme can unleash the important influence of the comparatively salient texture in the road.

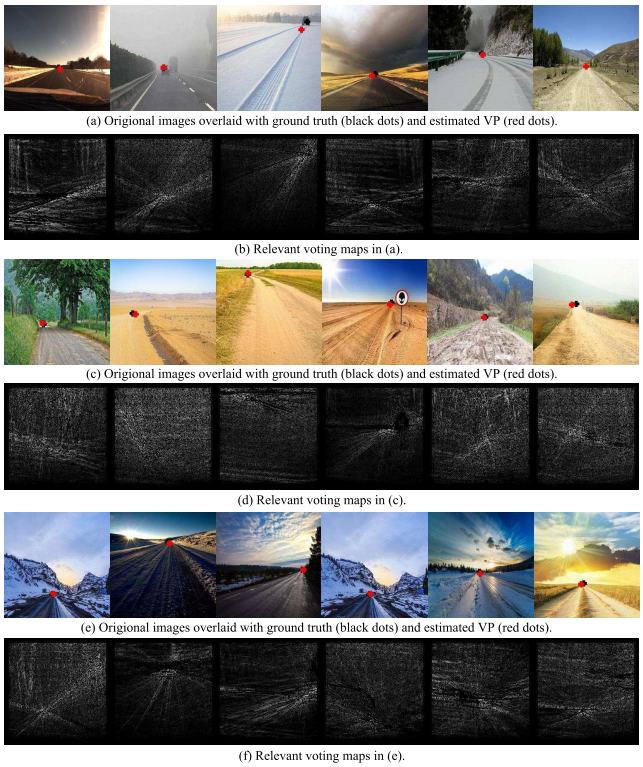

FIGURE 9. Examples of VP detection in our dataset.

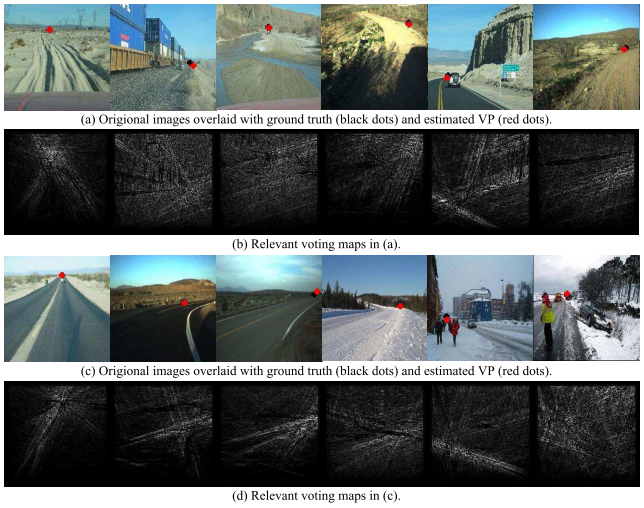

FIGURE 10. Examples of VP detection in public dataset.

Furthermore, we evaluate the performance of the proposed VP detection approach quantitatively. For this comparison, we adopt the normalized Euclidean distance proposed in [30] to measure the estimation error between the detected VP and the ground truth manually determined through the perspective of human perception. The normalized Euclidean distance is defined as

$$NormDist = \frac{||P_v - \hat{P}_v||}{Diag(I)} \tag{7}$$

where P_v and $\hat{P}_v$ are the ground truths of the VP and the estimated VP, respectively. $Diag(I) = \sqrt{2} \times 256$ is the length

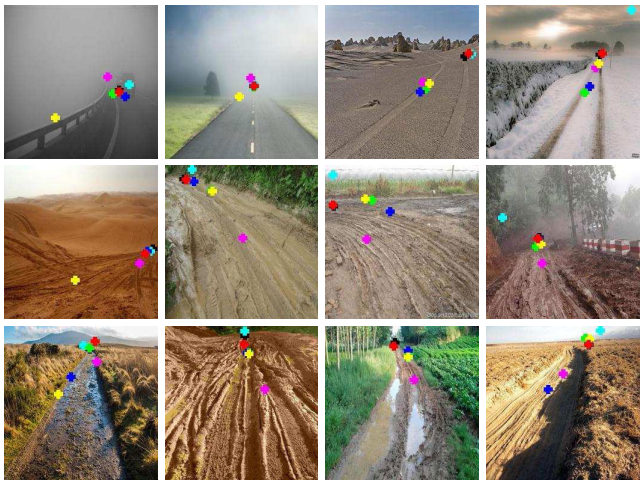

FIGURE 11. Experimental results of some road images with intricate or ambiguous texture. Red crosses show the results of our proposed method. Other colors are cyan, Rasmussen; green, Kong (Gabor); blue, Kong (gLoG); magenta, Moghadam; orange, Yang; and black, the ground truth.

TABLE 1. Mean error and mean running time for various methods in our dataset.

Methods	Error	Time(s)
Rasmussen [5]	0.051128	107.2732
Kong(Gabor) [28]	0.038313	84.22885
Kong(gloG) [31]	0.033494	89.70143
Moghadam [30]	0.045870	**0.701715**
Yang [32]	0.040211	0.873004
Proposed	**0.031513**	0.791989

TABLE 2. Mean error and mean running time for various methods in public dataset.

Methods	Error	Time(s)
Rasmussen [5]	0.077097	108.3658
Kong(Gabor) [28]	**0.040639**	85.9482
Kong(gloG) [31]	0.051556	90.2845
Moghadam [30]	0.063407	**0.725841**
Yang [32]	0.045931	0.868948
Proposed	0.056143	0.786584

of the diagonal of image. The width of the road in the image is 256 and the resolution is 256×256. The closer the *NormDist* is to 0, the closer the estimated VP is to the ground truth. A *NormDist* greater than 0.1 is set to 0.1, which is considered to be a failure of the corresponding method. To evaluate the performance on our own dataset and on the public dataset, the methods were implemented on a core i7-6700 3.4 GHz computer using MATLAB.

Table 1 shows the numerical results in terms of the mean error and the mean running time (in seconds) for our dataset obtained by the various methods. It is clear that our proposed method outperforms all the other VP detection approaches; is much faster than the methods of Rasmussen, Kong (Gabor), and Kong (gLoG); is comparable with that of Moghadam and Yang; but is slightly more expensive than the Moghadam method in computation time.

Table 2 shows the numerical results in terms of the mean error and the mean running time (in seconds) for the public

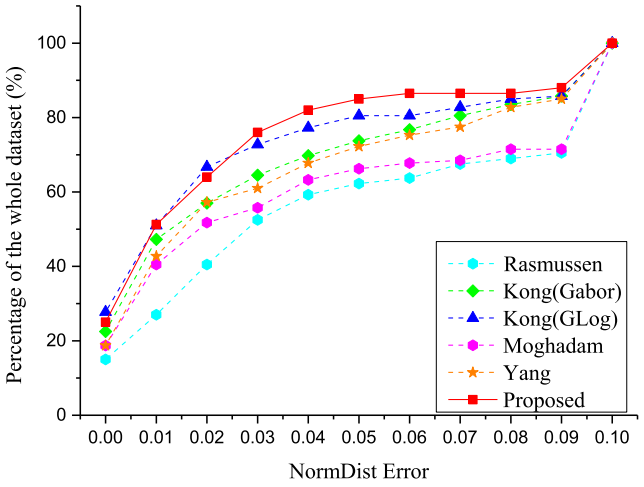

FIGURE 12. Accumulated error distribution of various VP detection methods in our dataset. On the x-axis, 0 stands for NormDist in [0, 0.01), 0.01 stands for *NormDist* in [0.01, 0.02)..., and 0.1 stands for *NormDist* in [0.1, 1].

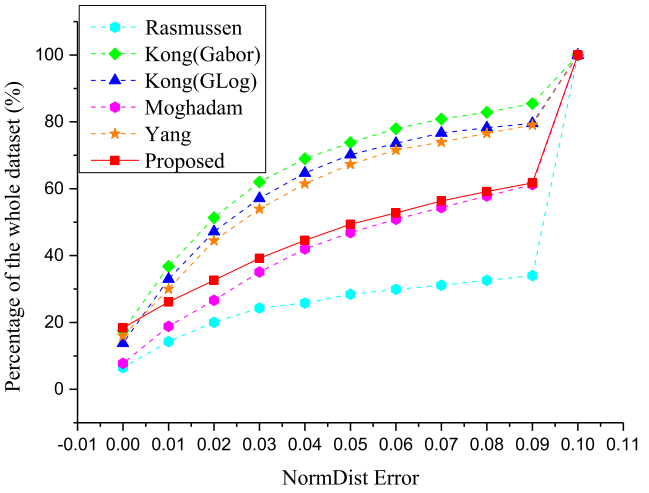

FIGURE 13. Accumulated error distribution of various VP detection methods in public dataset. On the x-axis, 0 stands for NormDist in [0, 0.01), 0.01 stands for *NormDist* in [0.01, 0.02)..., and 0.1 stands for *NormDist* in [0.1, 1].

dataset obtained by various methods. It can be seen in the Table that our proposed method outperforms Rasmussen's and Moghadam's VP detection approaches; is much faster than that of Rasmussen, Kong (Gabor), and Kong (gLoG); is comparable with that of Moghadam and Yang; but is slightly more expensive than Moghadam's method in computation time.

For in-depth analysis and a detailed comparison, we evaluated the methods by putting the normalized distance *NormDist* into an 11-bin histogram, as shown in Figure 12 and Figure 13. The y-axis shows the percentage of the entire dataset in each histogram bin, and the x-axis shows the normalized distance error (*NormDist*).

The histogram shows that, in our dataset, our proposed method has the lowest percentage of failure cases. For all the images, 48 (12%) had a normalized Euclidean distance error *NormDist* greater than 0.1. Image numbers for each of the methods are as follows: Rasmussen, 118 images (29.5%); Kong (Gabor), 57 images (14.25%); Kong (gLoG), 57 images (14.25%); Moghadam, 114 images (28.5%); and Yang, 60 images (15%). For a *NormDist* error of less than 0.01, image numbers for each of the methods are as follows: Rasmussen, 60 images (15%); Kong (Gabor), 90 images (22.5%); Kong (gLoG), 111 images (27.75%); Moghadam, 75 images (18.75%); and Yang, 75 images (18.75%). In contrast, our proposed method, with 100 images (25%) has a normalized distance error (*NormDist*) of less than 0.01.

The Figure 13 shows that, in the public dataset, our proposed method has the highest percentage of error NormDist in [0, 0.01). For all the images, 185 (18.35%) have a normalized Euclidean distance error NormDist in [0, 0.01). Image numbers for each of the methods are as follows: Rasmussen, 65 images (6.48%); Kong (Gabor), 175 images (17.55%); Kong (gLoG), 138 images (13.76%); Moghadam, 78 images (7.77%); and Yang, 160 images (15.95%). For a NormDist error of greater than 0.1, image numbers for each of the methods are as follows: Rasmussen, 662 images (66%); Kong (Gabor), 160 images (15.96%); Kong (gLoG), 209 images (20.84%); Moghadam, 400 images (39.89%); and Yang, 210 images (20.94%). In contrast, our proposed method, with 383 images (38.27%) has a normalized distance error (*NormDist*) of greater than 0.1.

TABLE 3. Comparison of mean error and mean running time for different level of orientation in both datasets.

Methods	Error	Time(s)
our dataset(8)	0.036485	0.501547
our dataset(16)	**0.031513**	0.791989
our dataset(32)	0.034851	1.248541
public dataset(8)	0.060152	0.512486
public dataset(16)	**0.056143**	0.786584
public dataset(32)	0.058467	1.245123

We investigate the decomposing level of orientation as mentioned in Section III. By selecting different levels of 8,16, and 32, the mean error and mean running time on our dataset and public dataset are shown in Table 3. When selecting 8 level, the mean error in the two datasets are 0.036485 and 0.060152, respectively. When selecting level 32, the mean error in the two datasets are 0.034851 and 0.058467, respectively. Obviously, the mean error when selecting 16 level is smaller than the other two levels (8 level and 32 level). In addition, although the mean time is little bit faster than the other two when selecting 8 levels, the mean running time of the three levels are all at a real-time running time. Therefore, the mean error is more important and the decomposing level of orientation is 16 in our experiment.

Furthermore, we investigate the role of the parameter T as mentioned in Section IV. By setting T to different values, the mean error on both datasets are shown in Table 4. When T is set to 20, it means that the extreme texture with a statistical number less than 20% of N is set to 0. It can eliminate the possible noise from strong textures. When T is set to other values in either dataset, the error increases.

TABLE 4. Mean error with different parameter T values in both datasets.

Methods	Error
our dataset(10)	0.033658
our dataset(15)	0.032548
our dataset(20)	**0.031513**
our dataset(25)	0.031847
our dataset(30)	0.032014
public dataset(10)	0.054841
public dataset(15)	0.055486
public dataset(20)	**0.056143**
public dataset(25)	0.056984
public dataset(30)	0.057458

VI. CONCLUSION

We have proposed a novel framework for VP detection in a single image for unstructured roads based on a contourlet transform. The contourlet transform is an accurate and fast texture estimation approach for achieving real-time VP detection. The images are first estimated by the contourlet transform to determine the dominant vectors of each pixel. Then, the algorithm uses the selection step of the pixels to reduce the computation cost and maintain stability in the voting process. The response of the dominant vector of selected pixels is further modulated to eliminate the influence of extreme textures in the image and give the pixels a moderate response to vote on. After the modulation process, all the pixels in the image are voted on by the selected and modulated pixels according to a line-voting scheme. Pixels with the maximum number of votes are considered as the estimated VP. Furthermore, a series of quantitative and qualitative analyses was conducted using a set of natural images downloaded from Google Images and from a public roads dataset. The proposed method was comparable to and outperformed the state-of-the-art VP detection methods in terms of time and accuracy for both our own and the public datasets. However, the proposed method may have difficulties in identifying whether or not an extreme texture is interference. In future work, this problem may be addressed by deep learning methods or by a method that can classify image textures into road and background noise.

REFERENCES

[1] G. Yang, Z. Lu, J. Yang, and Y. Wang, "An adaptive contourlet HMM–PCNN model of sparse representation for image denoising," *IEEE Access*, vol. 7, no. 1, pp. 88243–88253, Dec. 2019. Accessed: Jun. 24, 2019. doi: 10.1109/ACCESS.2019.2924674.

[2] K. D. Harris and T. D. Mrsic-Flogel, "Perceptual training continuously refines neuronal population codes in primary visual cortex," *Nature*, vol. 503, pp. 51–58, Nov. 2013.

[3] Y. Yan and T. D. Mrsic-Flogel, "Cortical connectivity and sensory coding," *Nature*, vol. 17, no. 10, pp. 1380–1387, Sep. 2014.

[4] B.-H. Liu, P. Y. Li, Y. J. Sun, Y.-T. Li, L. I. Zhang, and H. W. Tao, "Intervening inhibition underlies simple-cell receptive field structure in visual cortex," *Nature Neurosci.*, vol. 13, pp. 89–96, Nov. 2010.

[5] C. Rasmussen, "RoadCompass: Following rural roads with vision + ladar using vanishing point tracking," *Auton. Robot.*, vol. 25, no. 3, pp. 205–229, 2008.

[6] J. Shi, J. Wang, and F. Fu, "Fast and robust vanishing point detection for unstructured road following," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 970–979, Apr. 2016.

[7] Y. Wang, D. Shen, and E. K. Teoh, "Lane detection using spline model," *Pattern Recognit. Lett.*, vol. 21, no. 8, pp. 677–689, 2000.

[8] C. Rother, "A new approach to vanishing point detection in architectural environments," in *Proc. BMVC*, Bristol, U.K., 2000, pp. 382–391.

[9] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons, "The cascaded Hough transform as an aid in aerial image interpretation," in *Proc. IEEE-ICCV*, Bombay, India, Jan. 1998, pp. 67–72.

[10] X. Liu, Z. Cao, N. Gu, S. Nahavandi, C. Zhou, and M. Tan, "Intelligent line segment perception with cortex-like mechanisms," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 12, pp. 1522–1534, Dec. 2015.

[11] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using B-snake," *Image Vis. Comput.*, vol. 22, no. 4, pp. 269–280, Apr. 2004.

[12] R. Ebrahimpour, R. Rasoolinezhad, Z. Hajiabolhasani, and M. Ebrahimi, "Vanishing point detection in corridors: Using Hough transform and K-means clustering," *IET Comput. Vis.*, vol. 6, no. 1, pp. 40–51, Jan. 2012.

[13] M. Nieto and L. Salgado, "Real-time vanishing point estimation in road sequences using adaptive steerable filter banks," in *Proc. ACIVS*, Delft, The Netherlands, 2009, pp. 28–31.

[14] J.-P. Tardif, "Non-iterative approach for fast and accurate vanishing point detection," in *Proc. IEEE-ICCV*, Kyoto, Japan, Oct. 2009, pp. 1250–1257.

[15] W. Ding and Y. Li, "Efficient vanishing point detection method in complex urban road environments," *IET Comput. Vis.*, vol. 9, no. 4, pp. 549–558, Aug. 2015.

[16] Y. Li, W. Ding, X. Zhang, and Z. Ju, "Road detection algorithm for autonomous navigation systems based on dark channel prior and vanishing point in complex road scenes," *Robot. Auton. Syst.*, vol. 85, pp. 1–11, Nov. 2016.

[17] Z. Wu, W. Fu, R. Xue, and W. Wang, "A novel line space voting method for vanishing-point detection of general road images," *Sensors*, vol. 16, no. 7, p. 948, 2016.

[18] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, Apr. 2012.

[19] Y. Alon, A. Ferencz, and A. Shashua, "Off-road path following using region classification and geometric projection constraints," in *Proc. IEEE-CVPR*, New York, NY, USA, Jun. 2006, pp. 17–22.

[20] H. Kogan, R. Maurer, and R. Keshet, "Vanishing points estimation by self-similarity," in *Proc. IEEE-CVPR*, Miami Beach, FL, USA, Jun. 2009, pp. 20–25.

[21] H.-Y. Cheng, C.-C. Yu, C.-C. Tseng, K.-C. Fan, J.-N. Hwang, and B.-S. Jeng, "Environment classification and hierarchical lane detection for structured and unstructured roads," *IET Comput. Vis.*, vol. 4, no. 1, pp. 37–49, Mar. 2010.

[22] J. M. Alvarez, T. Gevers, and A. M. Lopez, "3D scene priors for road detection," in *Proc. IEEE-CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 13–18.

[23] E. Wang, A. Sun, Y. Li, X. Hou, and Y. Zhu, "Fast vanishing point detection method based on road border region estimation," *IET Image Process.*, vol. 12, no. 3, pp. 361–373, 2018.

[24] J. Huang, B. Kong, B. Li, and F. Zheng, "A new method of unstructured road detection based on HSV color space and road features," in *Proc. IEEE-IA*, Seogwipo-si, South Korea, Jul. 2007, pp. 08–11.

[25] A. Lookingbill, J. Rogers, D. Lieb, J. Curry, and S. Thrun, "Reverse optical flow for self-supervised adaptive autonomous robot navigation," *Int. J. Comput. Vis.*, vol. 74, no. 3, pp. 287–302, Sep. 2007.

[26] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *Proc. ECCV*, Florence, Italy, 2012, pp. 376–389.

[27] T. Li, C. Xu, and Y. Cai, "A fast and robust heuristic road detection algorithm," *Inf. Technol. J.*, vol. 13, no. 8, pp. 1555–1560, 2014.

[28] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2211–2220, Aug. 2010.

[29] O. Miksik, "Rapid vanishing point estimation for general road detection," in *Proc. IEEE-ICRA*, Saint Paul, MN, USA, May 2012, pp. 14–18.

[30] P. Moghadam, J. A. Starzyk, and W. S. Wijesoma, "Fast vanishing-point detection in unstructured environments," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 425–430, Jan. 2012.

[31] H. Kong, S. E. Sarma, and F. Tang, "Generalizing Laplacian of Gaussian filters for vanishing-point detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 408–418, Mar. 2013.

[32] W. Yang, B. Fang, and Y. Y. Tang, "Fast and accurate vanishing point detection and its application in inverse perspective mapping of structured road," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 5, pp. 755–766, May 2018.

[33] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
[34] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

GUOAN YANG received the B.S. degree in industrial automation and control engineering from Jilin University, Changchun China, in 1986, the M.S. degree in system and automatic control engineering from Tokyo Metropolitan University, Tokyo, Japan, in 1993, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2006. He started his professional career with Hertz Corporation, Tokyo, as a Research Fellow, and where he was involved in research on electronics and mobile communication, from April 1993 to May 2001. Since May 2001, he has been with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, China, where he became affiliated in image processing, computer vision, and pattern recognition. He is currently an Associate Professor with the School of Electronic and Information Engineering, Xi'an Jiaotong University. He has published close to 40 articles. He holds four patents and two books. His research interests include image compression, vision computing, wavelet analysis, multiscale geometric analysis, compressed sensing, neural networks, and deep learning. He is a Reviewer of several famous journals such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON MULTIMEDIA.

YUHAO WANG received the B.S. degree in control science and engineering from Southwest Jiaotong University, Chengdu, China, in 2013. He is currently pursuing the M.S. degree in control science and engineering with Xi'an Jiaotong University. From 2013 to 2016, he was a full-time Engineer in autonomous driving research with Xi'an Railway Company. His research interests include image processing, computer vision, road detection, and autonomous driving.

JUNJIE YANG received the B.S. degree in control science and engineering from Jilin University, Changchun, China, in 2018. He is currently pursuing the M.S. degree in control science and engineering with Xi'an Jiaotong University. His research interests include image compression, vision computing, multiscale geometric analysis, neural networks, and deep learning.

ZHENGZHI LU received the B.S. degree in control science and engineering and the M.S. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2017 and 2019, respectively., where he is currently pursuing the Ph.D. degree in control science and engineering. His research interests include image processing, computer vision, wavelets, and multiscale geometric analysis.