Article

Vision Sensor-Based Road Detection for Field Robot Navigation

Keyu Lu *, Jian Li, Xiangjing An and Hangen He

College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, Hunan, China; E-Mails: lijian@nudt.edu.cn (J.L.); anxiangjing@gmail.com (X.A.); hangenhe@yahoo.com (H.H.)

* Author to whom correspondence should be addressed; E-Mail: keyu.lu@nudt.edu.cn; Tel./Fax: +86-731-8457-4383.

Academic Editors: Lianqing Liu, Ning Xi, Wen Jung Li, Xin Zhao and Yajing Shen

---

Abstract: Road detection is an essential component of field robot navigation systems. Vision sensors play an important role in road detection for their great potential in environmental perception. In this paper, we propose a hierarchical vision sensor-based method for robust road detection in challenging road scenes. More specifically, for a given road image captured by an on-board vision sensor, we introduce a multiple population genetic algorithm (MPGA)-based approach for efficient road vanishing point detection. Superpixel-level seeds are then selected in an unsupervised way using a clustering strategy. Then, according to the GrowCut framework, the seeds proliferate and iteratively try to occupy their neighbors. After convergence, the initial road segment is obtained. Finally, in order to achieve a globally-consistent road segment, the initial road segment is refined using the conditional random field (CRF) framework, which integrates high-level information into road detection. We perform several experiments to evaluate the common performance, scale sensitivity and noise sensitivity of the proposed method. The experimental results demonstrate that the proposed method exhibits high robustness compared to the state of the art.

Keywords: robot navigation; road detection; MPGA; GrowCut; conditional random field

---

1. Introduction

Road detection is a fundamental issue in field robot navigation systems, which have attracted keen attention in the past several decades. Vision sensors play an important role in road detection. Image

data captured by vision sensors contains rich information, such as luminance, color, texture, *etc*. Thus, vision sensors have a great potentiality in road detection [1]. Moreover, vision sensors are inexpensive compared to other popular road detection sensors, such as LiDAR and millimeter-wave radar [2].

For these reasons, many state-of-the-art field robot systems employ vision sensors for road detection. For example, Xu *et al.* [3] presented a mobile robot using a vision system to navigate in an unstructured environment. The vision system consisted of two cameras; one is used for road region detection, and the other is used for road direction estimation. Rasmussen [4] introduced a vehicle-based mobile robot system, which has achieved success in the DARPA Grand Challenge. Vision sensors mounted on the top of the windshield were used to detect the road vanishing point for steering control.

Vision sensor-based road detection is a binary labeling problem trying to label every pixel in the given road image with the category (road or background) to which it belongs [5]. However, vision sensor-based road detection is still a challenging job due to the diversity of road scenes with different geometric characteristics (varying colors and textures) and imaging conditions (different illuminations, viewpoints and weather conditions) [5].

The problem of vision sensor-based road detection has been intensively studied in recent years. Some methods are based on color and texture features, e.g., the method presented in [6] uses the HSI color space as the features for road detection, while the algorithm proposed in [7] combines texture and color features. However, in many off-road environments, the texture and color features of the road and its surroundings are quite complex and diverse, and sometimes, it is extremely difficult to distinguish road regions from the surroundings by using only texture and color features. Another approach for road detection is based on road boundaries; the proposed method in [8] used road boundaries to fit a road curvature model for road detection. Nevertheless, this kind of approach does not appropriately behave when there is no evident borders (e.g., unstructured roads). More recently, the vanishing point was used for road detection in [9]. This kind of method does not work well when there is no obvious road vanishing point or the road has curved boundaries [5]. To deal with curved boundaries, in [10], the authors proposed using the illuminant invariance to detect road regions. This approach is robust to illuminations, shadows and curved roads. However, it contain less information on road shape priors and is sensitive to noise. To make sensible use of prior information, in [11], road priors obtained from geographic information systems (GISs) are combined with the road cues estimated from the current image to achieve robust road segmentation. However, the method may fail when there is no GIS database. Without GIS or a map, Sotelo *et al.* [12] used road shape restrictions to enhance the road segmentation. To make better use of road shape priors, He *et al.* [5] proposed to use road shape priors for the road segmentation by encoding the priors into a graph-cut framework, but the method would be suboptimal when the features of the road and background are similar.

In this paper, we introduce a hierarchical vision sensor-based road detection model to address this problem. More specifically, the proposed approach is depicted in Figure 1, which consists of three main components:

(1) Road vanishing point detection based on MPGA: We propose an efficient and effective road vanishing point detection method, which employed the multiple population genetic algorithm (MPGA) to search for vanishing point candidates heuristically. The value of the fitness function of MPGA is obtained by a locally-tangent-based voting scheme. In this way, we only need to

estimate the local dominant texture orientations and calculate voting values at the positions of vanishing point candidate. Thus, the proposed method is highly efficient compared to traditional vanishing point detection methods. In this paper, the road vanishing point is a key element of subsequent image processing tasks.

(2) GrowCut-based road segmentation: The initial road segments are obtained using GrowCut [13], which is an interactive segmentation framework based on cellular automaton (CA) theory [14]. The seed points of GrowCut are selected automatically by using the information of the road vanishing point, which makes GrowCut become an unsupervised process without an interactive property. Seed selection and GrowCut are performed at the superpixel level. Each superpixel is regarded as a cell with a label (road or background), the initial road segment is obtained when the proliferation of cells stops.

(3) Refinement using high-level information: In order to get rid of the shortcomings of the illuminant invariance-based method [11] and to ensure that the road segments are globally consistent, inspired by [5], we employ a conditional random field (CRF) [15] to integrate some high-level information into the road segments.
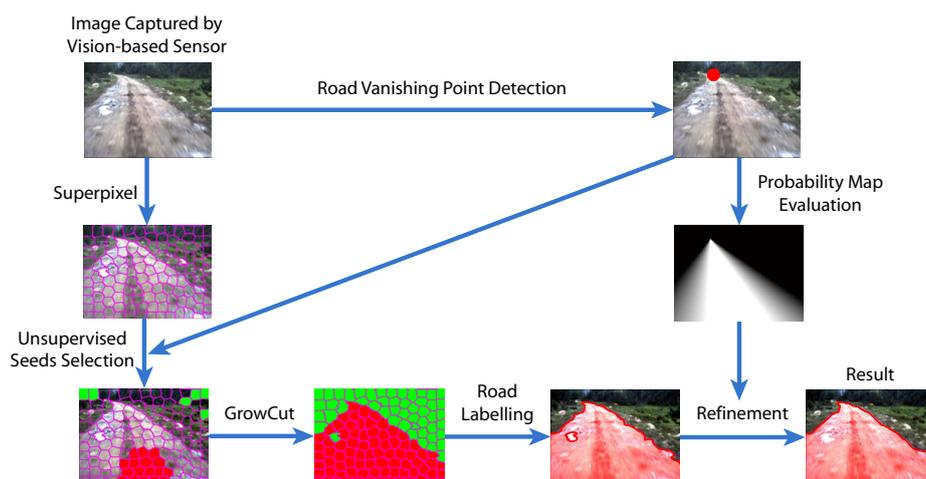

Figure 1. Diagram of our vision sensor-based road detection model.

The rest of this paper is organized as follows. In Sections 2–4, we introduce the proposed road detection method. Experimental results and a discussion are presented in Section 5, and conclusions are drawn in Section 6.

2. Road Vanishing Point Detection Based on MPGA

In a 2D image data captured by a vision sensor, the road vanishing point is the intersection of the projections of certain parallel lines (e.g., road edges, lane markings) in the real world. The road vanishing point plays an important role in road detection. The information of road direction can be obtained from the road vanishing point. Rasmussen *et al.* [16] use it for navigation directly. Moreover, the road vanishing point provides a strong clue to the localization of the road region. Hui Kong *et al.* [9] use the road vanishing point for road segmentation. In this paper, road vanishing point detection is a vital step

for the whole road detection task. This section presents a robust method for the fast estimation of the road vanishing point.

2.1. Searching Based on MPGA

Most texture-based vanishing point detection methods [9,16,17] obtain vanishing point candidates in a greedy way, which is computationally expensive. In the proposed method, the multiple population genetic algorithm (MPGA) is employed to search vanishing point candidates heuristically. After the vanishing point candidates are obtained, we just need to estimate the local dominant texture orientations of the vanishing point candidates and their voters, after which the voting value of the vanishing point candidate can be obtained by the local dominant texture orientations.

There are some modern optimization algorithms (such as the multiple population genetic algorithm (MPGA), particle swarm algorithm optimization (PSO), the simulated annealing algorithm (SA), *etc*.) that are mainly employed for difficult optimization problem solving. The work in [18] performed experiments to evaluate the performance of some modern optimization algorithms. It concluded that these modern optimization algorithms are effective, but the genetic algorithm (GA) is more efficient than PSO and SA. The genetic algorithm [19] is widely used in optimization problems. In the algorithm, problem parameters are encoded to chromosomes, and the solution can be obtained when the evolution of the chromosomes is stopped. However, the traditional genetic algorithm may lead to local optimal solutions, and its convergence speed is slow. The multiple population genetic algorithm (MPGA) was proposed to overcome the shortcomings of the traditional genetic algorithm. Instead of starting from one population, MPGA starts from two or more populations, and each population evolves in parallel. For this reason, we apply MPGA to search for vanishing point candidates.

In order to obtain chromosomes, the coordinates of the pixels are transformed to a binary encoding. Having obtained chromosomes, we next generate M initial populations, and each population contains N chromosomes. The fitness function $f(x, y)$ of the MPGA is defined as follows:

$$f(x, y) = \sum_{i=1}^{n} Vote(P_i, V_j) \tag{1}$$

where $Vote(P_i, V_j)$ represents the voting value of the vanishing point candidate V_j at (x, y), P_i is a voter of V_j and n denotes the number of voters in the voting region of V_j. The obtainment of voting values will be discussed in the next subsection.

2.2. Voting Scheme

After a vanishing point candidate is obtained, the local dominant texture orientation of the vanishing point candidate and its voters need to be estimated. Similar to the work in [16], Gabor filter banks are applied to estimate the local dominant texture orientation. For an orientation ϕ and a scale w, the Gabor kernels are defined by:

$$\psi_{w,\phi}(x, y) = \frac{w}{\sqrt{2\pi}c} e^{-w^2(4a^2+b^2)/(8c^2)} \left(e^{iaw} - e^{-c^2/2}\right) \tag{2}$$

where $a = x\cos\phi + y\sin\phi, b = -x\sin\phi + y\cos\phi$. The local dominant texture orientation of each vanishing point candidate is obtained by Gabor filter banks of 36 orientations (180 divided by five). The

estimation method is explained in detail in [9]. In our work, the local dominant texture orientation of a pixel is only computed once, and the value is saved for the subsequent computing.

After the local dominant texture orientation of the vanishing point candidate V_j is obtained, its voting value $Vote(P_i, V_j)$ can be computed. As shown in Figure 2, V is a vanishing point candidate, P is a voter of V and $\overrightarrow{PA}$ is the local dominant texture orientation vector of P. β equals $\angle VAP$. We define V_{proj} as the distance of V and A.
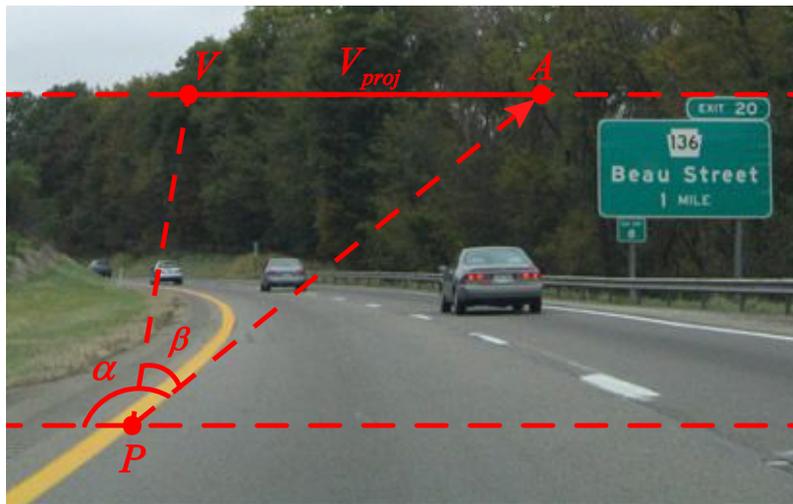

Figure 2. Illustration of the voting scheme.

As shown in Figure 2, (V_x, V_y) and (P_x, P_y) are defined as the coordinates of V and P, respectively. V_{proj} can be obtained as follows:

$$V_{proj} = P_x - V_x + \frac{V_y - P_y}{\tan(180 - \alpha)} \tag{3}$$

After getting the value of V_{proj}, our voting scheme can be defined as follows:

$$Vote(P, V) = \begin{cases} 0 & V_{proj} > \dfrac{ImageWide}{2} \\ \dfrac{1}{1 + V_{proj}^2} & otherwise \end{cases} \tag{4}$$

where $ImageWide$ represents the width of the image, V is a vanishing point candidate and P is a voter of V. The definition of the local voting region is similar to the soft-voting scheme in [9], which defines the voting regions as the half-disk below V. All of the pixels in the local voting region are used as voters.

For better comparison, like traditional texture-based vanishing point detection methods, we compute $Vote(P, V)$ at every pixel in the test image using Equation (4), where V_{proj} is obtained from Equation (3). The results are shown in Figure 3b, which illustrates that the vanishing point is the maximum value in the voting space. For MPGA-based voting, we only compute the voting value at the vanishing point candidate using Equation (4) and then search the next vanishing point candidate based on MPGA according to the result of Equation (4). Each vanishing point candidate is shown in Figure 3c (marked with red color). From Figure 3, we can see that every population experiences the processes of selection, reproduction and mutation by the fitness function; the maximum value could be found at last, and the road vanishing point is estimated at the location of the maximum value.
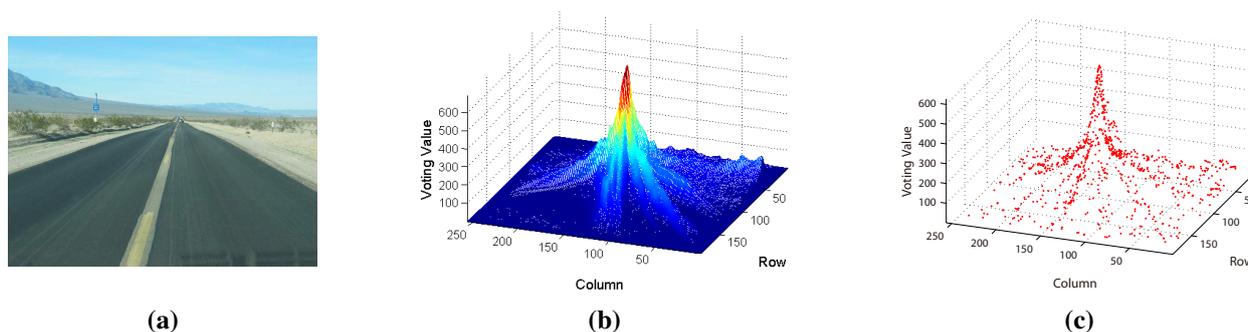
Figure 3. Illustration of voting: (**a**) the original road image; (**b**) the greedy voting map; (**c**) the MPGA voting map.

As we just need to estimate the local dominant texture orientations and compute voting values at the positions of the vanishing point candidate, the proposed method is more efficient than traditional texture-based methods. For a quantitative comparison, let MP and NI denote the number of initial populations and the number of chromosomes in each population, respectively. The number of vanishing point candidates is denoted by NV. For the road images of size 320×240, the NV of traditional texture-based methods is about:

$$NV = 320 \times 240 = 76,800 \tag{5}$$

We perform an experiment to validate the efficiency of our MPGA-based voting strategy. As shown in Table 1, NV of our voting strategy is much less than that of traditional texture-based methods (76,800). This means that our voting strategy spends less time on voting. Thus, the proposed method is more efficient than the traditional one.

Table 1. The result of efficiency validation.

NV \ NI MP	10	20	30	40	50
10	418	1009	1553	2139	2674
20	731	1767	2716	3738	4594
30	1004	2439	3761	5132	6313
40	1290	3098	4717	6427	7797
50	1537	3681	5597	7571	9177

Figures 4 and 5 are some examples of MPGA vanishing point detection in structured roads and unstructured roads, respectively. The red points in the first rows of the figures are the obtained road vanishing point. The second rows are the corresponding MPGA voting maps.
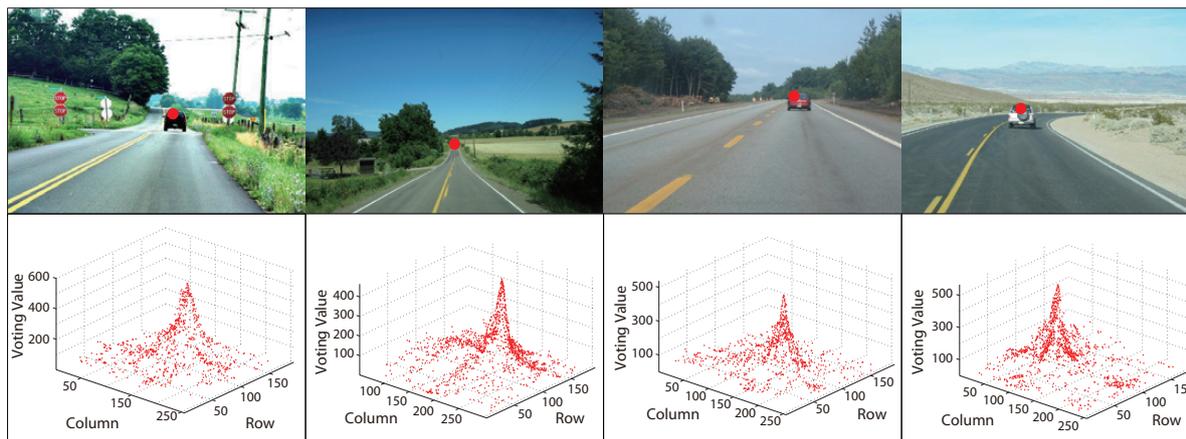
Figure 4. Examples of MPGA-based vanishing point detection in structured roads.
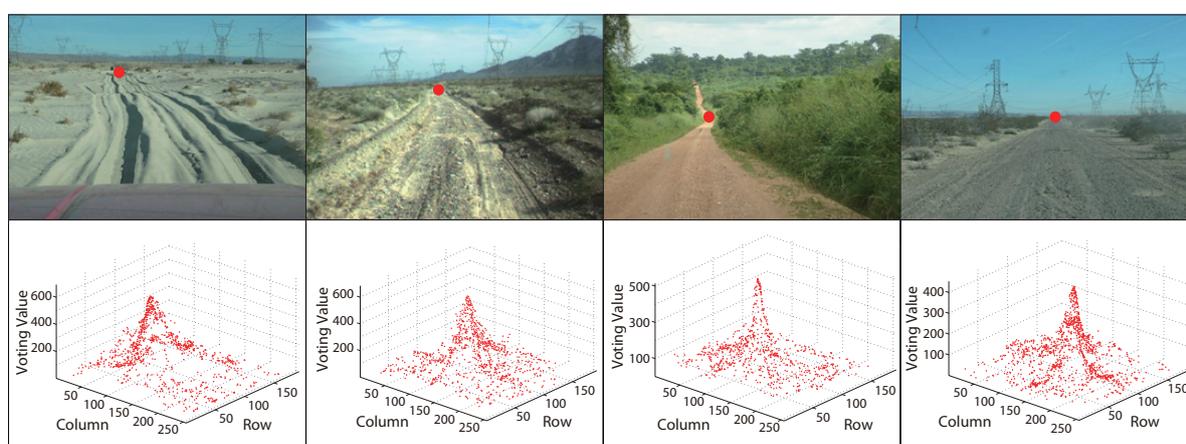

Figure 5. Examples of MPGA-based vanishing point detection in unstructured roads.

3. GrowCut-Based Road Segmentation

The initial road segments are obtained using GrowCut [13], which is an interactive segmentation framework based on cellular automaton (CA) theory [14]. GrowCut starts with a set of seed points, and the seed points iteratively try to occupy their neighbors according to cellular automaton, until convergence. Because of its robustness, GrowCut is widely used in image segmentation [13], object detection [20], *etc*. In this paper, the seed points of GrowCut are selected automatically by using the information of the road vanishing point, which makes GrowCut become an unsupervised process without an interactive property.

3.1. Seed Selection at the Superpixel Level

Seeds are the starting points of GrowCut. In our work, the seeds are selected automatically by using the K-means clustering algorithm [21]. The road and the background clustering region are defined by using the information of the road vanishing point, which can be obtained efficiently by the method introduced in Section 2. As shown in Figure 6a, *V* denotes the road vanishing point of the given image

and AB is the horizontal line crossing the road vanishing point V (called the vanishing line). Let C and D denote the bottom left corner and the bottom right corner of the given image, respectively. The lines between V, C and D divide the area below AB into three regions: VAC, VCD and VBD.

For images captured by on-board cameras, most pixels in the region VCD belong to the road surface, and most pixels in the region VAC and VBD belong to the background. The region VCD is defined as the road clustering region, while VAC and VBD are the background clustering region. As shown in Figure 6b, the red area is the road clustering region; the green area is the background clustering region; and the blue area is the sky region (the region above vanishing line AB).
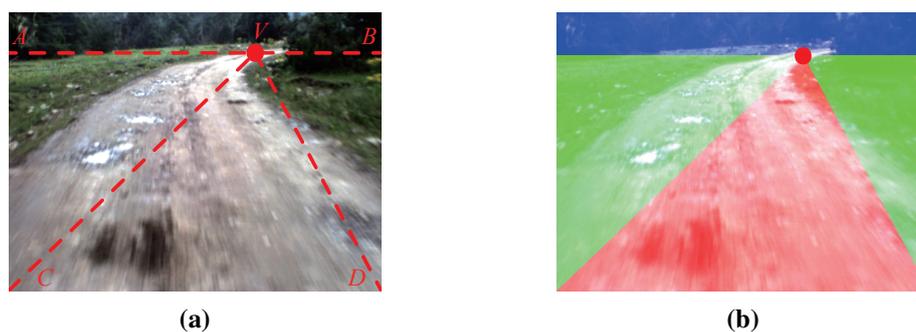

Figure 6. Clustering region: (**a**) illustration of the clustering region definition; (**b**) example of the clustering region definition.

We use three channels (R, G and B) of the given color road image as the features of clustering. After the clustering region is defined, the K-means clustering algorithm [21] is applied to the road and the background clustering regions, respectively (the initial points of clustering are selected randomly). Road and background region clustering is illustrated in Figure 7, where Figure 7a is the original road image and Figure 7b,c are the results of RGB feature-based K-means clustering. For the result of road region clustering in Figure 7b, the proportion of the red category is larger than that of the green category. This means that in the result of road region clustering, the category with a larger proportion can be seen as the representative of the road surface. According to this assumption, road seeds of GrowCut are selected from the category with a larger proportion in the result of the road region clustering (the red category in Figure 7b). Similarly, background seeds are selected from the category with a larger proportion in the result of the background region clustering (the green category in Figure 7c).
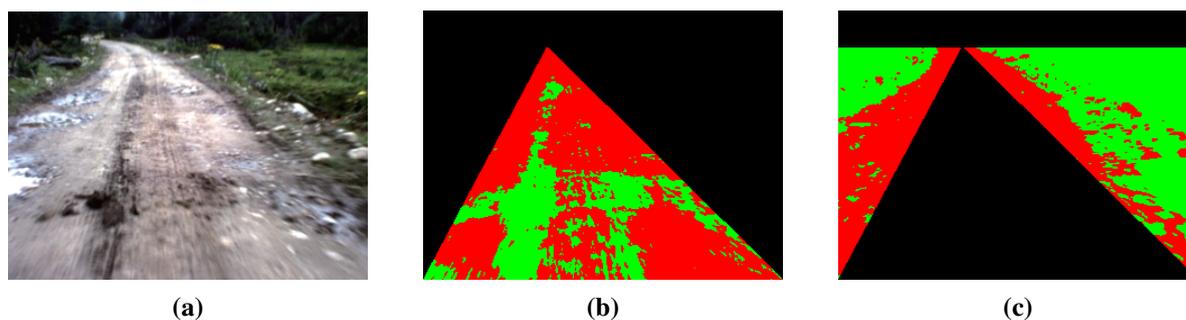

Figure 7. Road and background region clustering: (**a**) test road image; (**b**) road region clustering; (**c**) background region clustering.

In order to produce spatially more appealing road segments, inspired by [22], we select seeds and apply GrowCut at the superpixel level. We perform a preliminary over-segmentation of the given road image into superpixels using the SLICalgorithm [23]. Rather than using the pixel grids and rectangular patches, superpixels are likely to be uniform in color and texture and tend to preserve boundaries, so they are more perceptually meaningful and representationally efficient [24]. Furthermore, using superpixels may dramatically reduce the computational complexity of subsequent image processing tasks (such as GrowCut) [22].

In the given road image, for the k-th superpixel S_k, Cr_k denotes the ratio of the number of pixels belonging to road seeds and the total number of pixels in this superpixel and takes the form:

$$Cr_k = \frac{\sum\limits_{i \in S_k} l_i}{N_k} \tag{6}$$

where $l_i \in \{0(background), 1(road)\}$ denotes the label of the i-th pixel and N_k is the total number of pixels in S_k.

Let (x_k, y_k) denote the mean coordinates of the pixels in S_k and (x_m, y_m) be the coordinates of the pixel in the bottom center of the given road image. For the selection of road seeds, the normalized distance between S_k and (x_m, y_m) is denoted by Dr_k, which can be defined as follows:

$$Dr_k = \frac{\sqrt{(x_m - x_k)^2 + (y_m - y_k)^2}}{\sqrt{x_m^2 + y_m^2}} \tag{7}$$

Let Pr_k be the probability of S_k being selected as a road seed:

$$Pr_k = \frac{Cr_k + (1 - Dr_k) \cdot Tr}{1 + Tr} \tag{8}$$

where Tr is the factor controlling the weight of $(1 - Dr_k)$ in computing Pr_k ($Tr = 0.01$ in this work). Due to the fact that superpixels closer to the bottom center of the on-board road image have a higher probability of being road seeds, Dr_k should be taken into consideration in computing the probability of being road seeds.

For the result of the road region clustering in Figure 8a, the probability of each superpixel being selected as a road seed can be obtained by Equation (8) (see Figure 8b, which is called the probability map). The result of superpixel-level road seed selection is illustrated in Figure 8c.
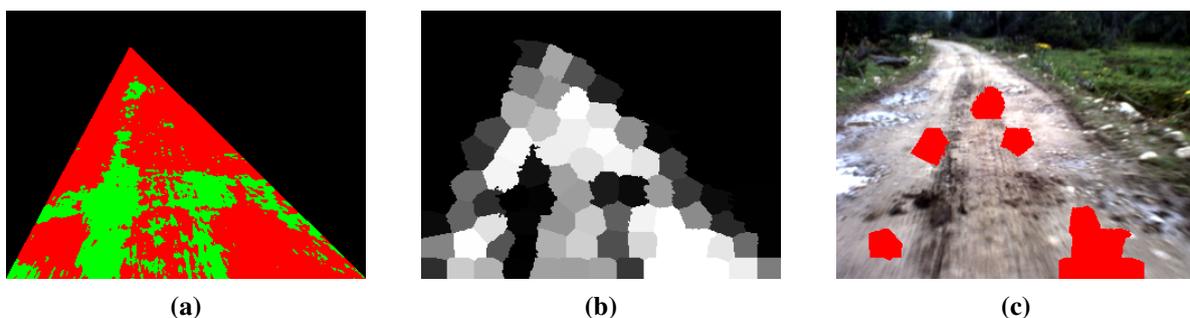

(a)	(b)	(c)

Figure 8. Illustration of superpixel-level road seed selection: (**a**) the result of road region clustering; (**b**) the probability of being road seeds (the brighter the superpixel, the higher the probability); (**c**) the result of superpixel-level road seed selection.

Compared to most traditional seed selection methods, the proposed method is highly robust. With the common assumption [25], traditional methods define a "safe" window in the road image and assume that the pixels in the "safe" window belong to the road pattern (see Figure 9; the "safe" window is a semi-circular region at the center-bottom of the on-board road image). As shown in Figure 9, traditional methods may not work well when the features of the road and its surroundings are complex and diverse, where they are likely to select seeds in some incorrect regions, such as shadow and some parts of the car.
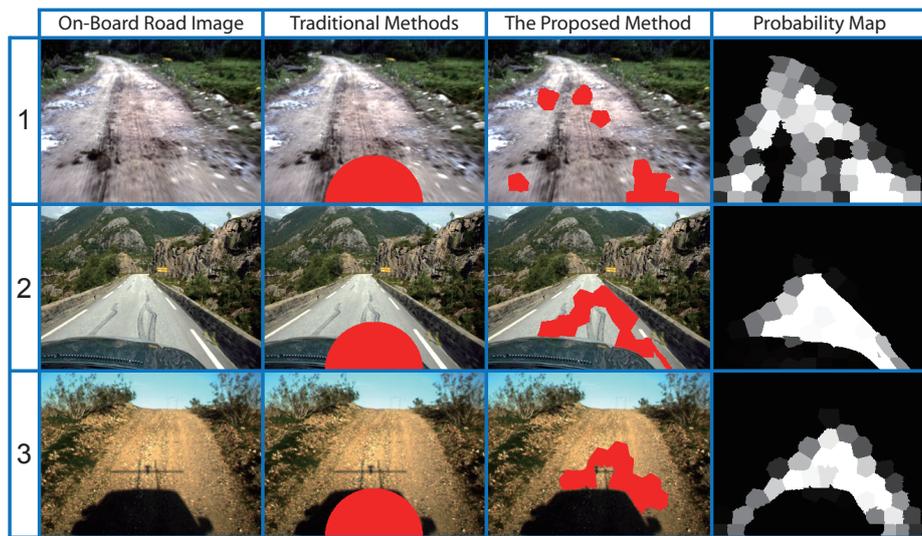

Figure 9. Comparisons with the traditional seed selection method.

Similarly, for background seeds, let Cg_k be the ratio of the number of pixels belonging to background seeds and the total number of pixels in the superpixel S_k:

$$Cg_k = \frac{\sum\limits_{i \in S_k} |l_i - 1|}{N_k} \tag{9}$$

In order to make background seeds be equally distributed on the left and right side of the road image, the background clustering region is divided into two parts, denoted by Bl and Br, which represent the parts on the left and right side of the road clustering region, respectively. Let (x_l, y_l) and (x_r, y_r) be the coordinates in the top left corner and top right corner of the background clustering region, respectively. $(x_{\max}, y_{\max})$ denote the coordinates in the bottom right corner of the given road image. For the selection of background seeds, the normalized distance Dg_k takes the form:

$$Dg_k = \begin{cases} \dfrac{\sqrt{(x_l - x_k)^2 + (y_l - y_k)^2}}{\sqrt{x_{\max^2} + y_{\max^2}}} & k \in Bl \\[4mm] \dfrac{\sqrt{(x_r - x_k)^2 + (y_r - y_k)^2}}{\sqrt{x_{\max^2} + y_{\max^2}}} & k \in Br \end{cases} \tag{10}$$

The probability of S_k being selected as a background seed can be defined as follows:

$$Pg_k = \frac{Cg_k + (1 - Dg_k) \cdot Tg}{1 + Tg} \tag{11}$$

where Tg is the factor controlling the weight of $(1 - Dg_k)$ in computing Pg_k ($Tg = 0.01$ in this work).

Clustering-based background seed selection aims at selecting background seeds outside the road regions. Although clustering regions have many kinds of textures, background seeds are only selected from the category with a larger proportion in the result of the background region clustering, which is less likely to belong to the road pattern. Besides, Equations (10) and (11) penalize the superpixels close to road regions. In these ways, clustering-based background seed selection is able to work well.

For on-board road images, according to the fact that the sky is likely to appear in the region above the vanishing line (the blue area in Figure 6b) and that the features of this region are quite different from that of the road and background clustering region, we need to select two background seeds in the region above the vanishing line. Let S_{sky1} and S_{sky2} be two superpixel-level background seeds, which are selected by Equation (12).

$$\begin{cases} sky1 = \arg\min_{k} \sqrt{(x_k - 1)^2 + (y_k - 1)^2} \\ sky2 = \arg\min_{k} \sqrt{(x_k - x_{\max})^2 + (y_k - 1)^2} \end{cases} \tag{12}$$

where $(1, 1)$ and $(x_{\max}, 1)$ are the coordinates in the top left corner and top right corner of the road images. The selection of S_{sky1} and S_{sky2} is illustrated in Figure 10, and some qualitative results of seed selection at the superpixel level are illustrated in Figure 11.
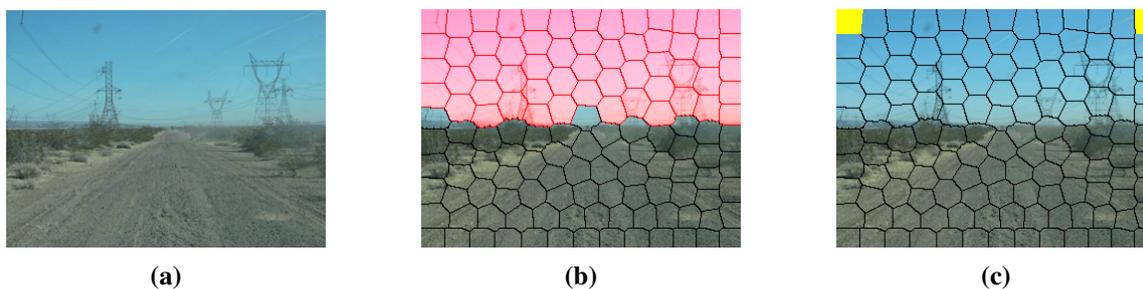

(**a**)   (**b**)   (**c**)

Figure 10. The selection of superpixel-level background seeds above a vanishing line: (**a**) test road image; (**b**) superpixels above the vanishing line (colored in red); (**c**) the result of the selection (colored in yellow).

Figure 11. The results of superpixel-level seed selection (road seeds are colored in red; background seeds are colored in green).

3.2. Segmentation Using the GrowCut Framework

After seed points are obtained, the seed superpixels iteratively try to occupy their neighbors according to GrowCut [14] at the superpixel level until convergence. For the superpixel S_i, let Sl_i be its label (road or background), where $Sl_i \in \{-1(background), 0(undetermined), 1(road)\}$. $\overrightarrow{F_i}$ denotes the feature vector of the superpixel S_i and θ_i denotes the strength ($\theta_i \in [0,1]$), which stands for the ability to attack or defend. Thus, the state of the superpixel S_i can be defined by a triplet $(Sl_i, \theta_i, \overrightarrow{F_i})$. Initially, the label of each superpixel is set as:

$$Sl_i = \begin{cases} 1 & S_i \in Sr \\ -1 & S_i \in Sg \\ 0 & S_i \in \overline{Sr \cup Sg} \end{cases} \tag{13}$$

where Sr and Sg denote the sets of road and background seeds, respectively. the initial value of strength is set as:

$$\theta_i = \begin{cases} 1 & S_i \in Sr \cup Sg \\ 0 & S_i \in \overline{Sr \cup Sg} \end{cases} \tag{14}$$

Moreover, we need to define the distance between two feature vectors. Álvarez *et al.* had demonstrated that intraclass variability caused by lighting conditions is a major challenge for road detection [10]. As a field robot may work under different lighting conditions (e.g., different illuminations, weather and shadow conditions), in this work, illuminant invariance features [10] and RGB color features are used to describe each superpixel. Illuminant invariance features are able to make the road detection more robust in shadowy and varying illumination conditions. For each pixel in a color image, illuminant invariance feature I_i can be obtained by:

$$\begin{cases} I_i = e^{\beta} \\ \beta = \log(\dfrac{R_i}{G_i + 1}) \cdot \cos(\theta) + \log(\dfrac{B_i}{G_i + 1}) \cdot \sin(\theta) \end{cases} \tag{15}$$

where (R_i, G_i, B_i) denotes the RGB color feature of the i-th pixel. The illuminant invariance feature of the i-th superpixel is defined as follows:

$$FI_i = \frac{\sum\limits_{k \in S_i}^{n} I_k}{n_i} \tag{16}$$

where n_i denotes the number of pixels in superpixel S_i. Similarly, the RGB color feature of the i-th superpixel can be obtained by:

$$FR_i = \frac{\sum\limits_{k \in S_i}^{n} R_k}{n_i}, FG_i = \frac{\sum\limits_{k \in S_i}^{n} G_k}{n_i}, FB_i = \frac{\sum\limits_{k \in S_i}^{n} B_k}{n_i} \tag{17}$$

The distance between feature vectors i and j is denoted by $Dm_{i,j}$ and takes the form:

$$Dm_{i,j} = \frac{|FI_i - FI_j| + Km \cdot \sqrt{(FR_i - FR_j)^2 + (FG_i - FG_j)^2 + (FB_i - FB_j)^2}}{1 + Km} \tag{18}$$

where Km is the factor controlling the weight of the RGB color feature in computing $Dm_{i,j}$ ($Km = 0.2$ in this work).

According to the GrowCut [14] framework, each superpixel is regarded as a cell, and cells with labels 1 and -1 proliferate from seeds. In the process of proliferation, they attack and occupy their differently-labeled neighbors. Let S_i and S_j be two neighbor cells with different labels $((Sl_i \neq Sl_j) \wedge (Sl_i \neq 0) \wedge (Sl_j \neq 0))$. Let S_i be the attacking cell and S_j the attacked cell; then, S_i occupies S_j if:

$$g(Dm_{i,j}) \cdot \theta_i > \theta_j \tag{19}$$

where $g(x)$ is a monotonically-decreasing function:

$$g(x) = 1 - \frac{x}{\max \|Dm_{i,j}\|_2} \tag{20}$$

If S_i occupies S_j, then the label and strength of S_j will change according to the state of S_i:

$$\begin{cases} Sl_j = Sl_i \\ \theta_j = g(Dm_{i,j}) \cdot \theta_i \end{cases} \tag{21}$$

Each superpixel in the given road image tries to occupy its differently-labeled neighbors iteratively until all of the superpixels converge to a stable configuration. As shown in Figure 12, the road segment is obtained after convergence.
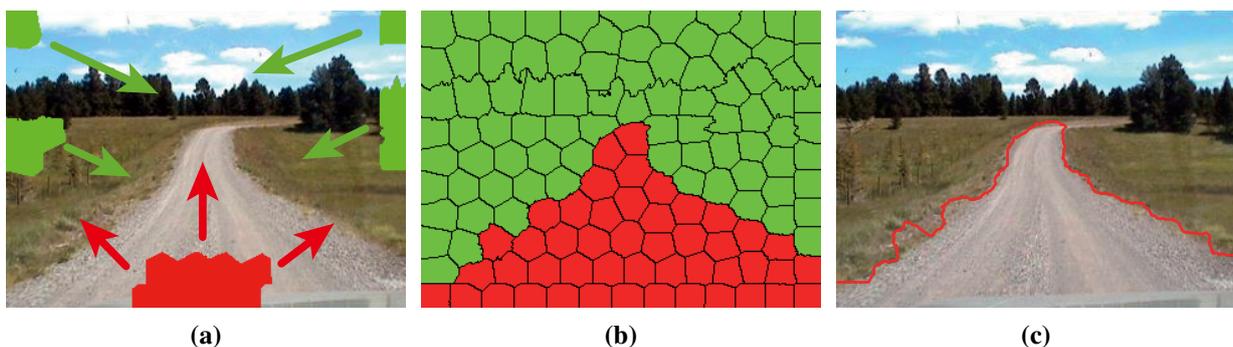

Figure 12. Illustration of segmentation using GrowCut: (**a**) the proliferation of seeds; (**b**) the convergence of the iteration; (**c**) the contour of the obtained road segments.

4. Refinement Using High-Level Information

Due to the complexity of the road scene, there still may exist some isolated superpixels or unsmooth edges in the road segments obtained by GrowCut (see Figure 13). The initial road segments still need to be refined by some high-level information. Olga Veksler [26] used the information called the "star shape prior" for image segmentation. Similarly, He *et al*. [5] proposed to use the information of road shape priors for the road segmentation by encoding the information into a graph-cut framework.
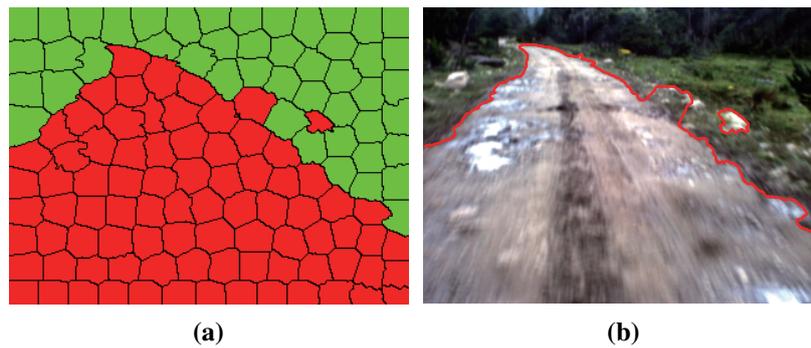
Figure 13. Problems with initial road segments obtained by GrowCut: (**a**) the convergence of the iteration; (**b**) the contour of road segments.

Inspired by these previous works, we implement the conditional random field (CRF) [15] at the pixel level to integrate some high-level information (e.g., shape prior, road vanishing point) to achieve a more robust detection.

In this work, the high-level information integrated in the refinement is listed as follows:

- There exist no isolated area in the road segments nor background segments;
- In on-board road images, road segments are shrinking from bottom to top;
- The direction of the road is relevant to the position of the road vanishing point.

To build the CRF model at the pixel level, let l_i denote the label (road or background) of the pixel i and l denote the set of all label assignments. The CRF energy to minimize can be written as:

$$E(l) = \sum_{i \in V} \phi_i(\hat{l}_i, l_i) + \lambda \sum_{d_{ij} \in D} \psi_{ij}(l_i, l_j) + \sum_{d_{ij} \in D} \varphi(l_i, l_j) + w \sum_{i \in V} \eta_i \tag{22}$$

where $\phi_i(\hat{l}_i, l_i)$ is the unary term enforcing the label l_i (road or background) to take a value close to the label $\hat{l}_i$ obtained by GrowCut:

$$\phi_i(\hat{l}_i, l_i) = \begin{cases} 1 & \hat{l}_i \neq l_i \\ 0 & \hat{l}_i = l_i \end{cases} \tag{23}$$

In Equation (22), ψ_{ij} is the pairwise term penalizing the different assignments for neighboring pixels i and j. Let $\vec{C_i}$ be the RGB color feature vector of the i-th pixel; ψ_{ij} takes the form:

$$\psi_i(\hat{l}_i, l_i) = \begin{cases} e^{-\beta \|\vec{C_i} - \vec{C_j}\|_2} & l_i \neq l_j \\ 0 & l_i = l_j \end{cases} \tag{24}$$

Inspired by [5], a second-order term φ_{ij} is employed to incorporate the road shape prior into the CRF framework. To describe the road shape prior, let h be the middle line of the road segment obtained by GrowCut (see Figure 14a). Lp and Rp denote the left and right part of h, respectively. i is an arbitrary pixel in the road segment. $j_1 \ldots j_8$ are neighbors of i (see Figure 14b). From the high-level information, the road shape implies that:

$$\begin{cases} \forall i, (i \in Lp) \wedge (j_3 \in Road) \Rightarrow i \in Road \\ \forall i, (j_7 \in Lp) \wedge (i \in Road) \Rightarrow j_7 \in Road \\ \forall i, (i \in Rp) \wedge (j_7 \in Road) \Rightarrow i \in Road \\ \forall i, (j_3 \in Rp) \wedge (i \in Road) \Rightarrow j_3 \in Road \end{cases} \tag{25}$$
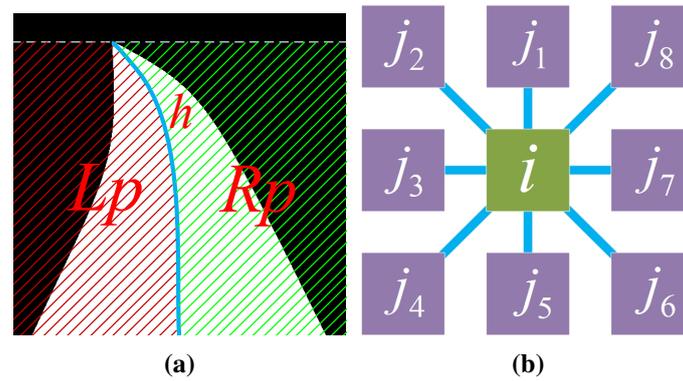
Figure 14. Illustration of the road shape prior: (**a**) a road segment with the middle line h; Lp denotes the left part of h; Rp denotes the right part of h; (**b**) a standard eight-neighborhood system with a center pixel i and its neighbors $j_1 \ldots j_8$.

From Equation (25), the second-order term φ_{ij} can be defined as:

$$
\varphi_{ij}(l_i, l_j) = \begin{cases} \infty & if \; \{(l_i = 0 \; and \; l_j = 1 \; and \; i \in Lp \; and \; j = j_3) \\ & or \; (l_i = 1 \; and \; l_j = 0 \; and \; j \in Lp \; and \; j = j_7) \\ & or \; (l_i = 0 \; and \; l_j = 1 \; and \; i \in Rp \; and \; j = j_7) \\ & or \; (l_i = 1 \; and \; l_j = 0 \; and \; j \in Rp \; and \; j = j_3)\} \\ 0 & otherwise \end{cases} \tag{26}
$$

In the CRF energy Equation (22), the last term η_i is used to incorporate road vanishing point information. For the calculation of η_i, we first need to obtain a probability map according to the position of the road vanishing point. Let D_p and M respectively denote the length and the center of the projection of the road segment onto the last line of the road image. V denotes the road vanishing point. The road image is divided into several regions illustrated in Figure 15, where the length of BM and CM is one fourth of that of D_p, and the length of AM and DM is three fourths of that of D_p. We define the probability of the road pattern to be presented at each pixel in the region VBC is 100%. In VAB and VCD, pixels closer to the center line VM have a higher probability of being road pixels. Let (x_v, y_v) be the coordinates of the road vanishing point; ImH and ImW denote the height and width of the road image, respectively; (x_m, ImW) denotes the coordinates of M. As shown in Figure 15, $f_{L1}(y)$, $f_{L2}(y)$, $f_{R1}(y)$ and $f_{R2}(y)$ are respectively the linear functions of VB, VA, VC and VD. The linear functions takes the form:

$$
\begin{cases} f_{L1}(y) = (y - y_v) \cdot \dfrac{x_v - (x_m - 0.5D_p)}{y_v - ImH} + x_v \\[2mm] f_{L2}(y) = (y - y_v) \cdot \dfrac{x_v - (x_m - 0.75D_p)}{y_v - ImH} + x_v \\[2mm] f_{R1}(y) = (y - y_v) \cdot \dfrac{x_v - (x_m + 0.5D_p)}{y_v - ImH} + x_v \\[2mm] f_{R2}(y) = (y - y_v) \cdot \dfrac{x_v - (x_m + 0.75D_p)}{y_v - ImH} + x_v \end{cases} \tag{27}
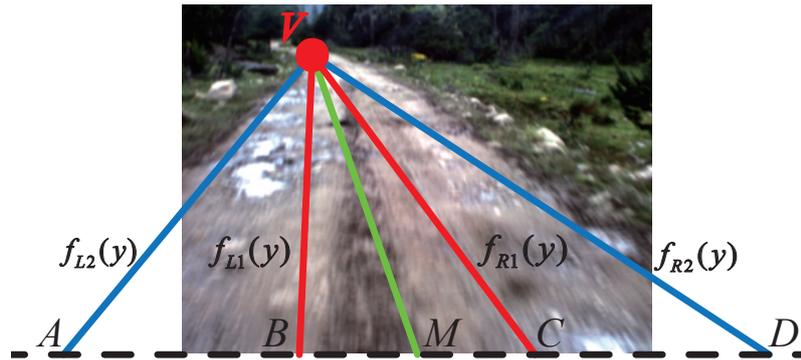$$

Figure 15. Image division for probability map calculation.

Let (x_i, y_i) be the coordinates of the i-th pixel in the image; the last term η_i of the CRF energy stands for the probability obtained according to the position of the road vanishing point and can be obtained by Equation (28). The probability map is illustrated in Figure 16.

$$
\eta_i = \begin{cases}
0 & x_i < f_{L2}(y_i) \\
\dfrac{x_i - f_{L2}(y_i)}{f_{L1}(y_i) - f_{L2}(y_i)} & f_{L2}(y_i) \leq x_i < f_{L1}(y_i) \\
1 & f_{L1}(y_i) \leq x_i \leq f_{R1}(y_i) \\
\dfrac{x_i - f_{R2}(y_i)}{f_{R1}(y_i) - f_{R2}(y_i)} & f_{R1}(y_i) < x_i \leq f_{R2}(y_i) \\
0 & x_i > f_{R2}(y_i)
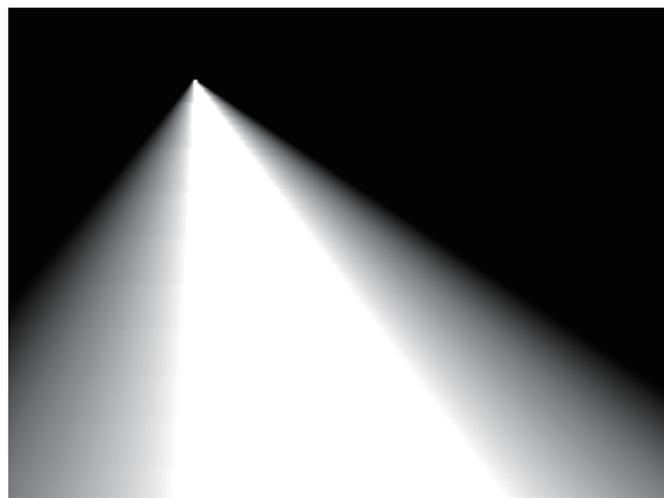\end{cases}
\tag{28}
$$

Figure 16. Probability map obtained according to the position of the road vanishing point (the brighter pixel stands for the higher probability).

The CRF energy $E(l)$ in Equation (22) is minimized using Graph Cut [27]. The result of refinement is illustrated in Figure 17, which manifests that the refinement using high-level information is effective in road detection.

(a)   (b)

Figure 17. Comparison between the results of GrowCut and refinement: (**a**) the result of GrowCut; (**b**) the result of refinement.

5. Results and Discussion

5.1. Common Performance

To validate the proposed road detection model, we evaluate the performance of the proposed approach on the OffRoadScene database [28], which consists of 770 unstructured road images, which are captured by driving on several challenging unstructured roads. The database contains various types of road scenes with different texture, shadows, illuminations and weather conditions. All of the images are resized to 320×240 for testing. The ground truth of the database is manually labeled. As most of the unstructured roads have no clear borders, the database uses ambiguous regions to describe the road borders; as shown in Figure 18, the ambiguous regions are unconcerned with respect to the experimental results.
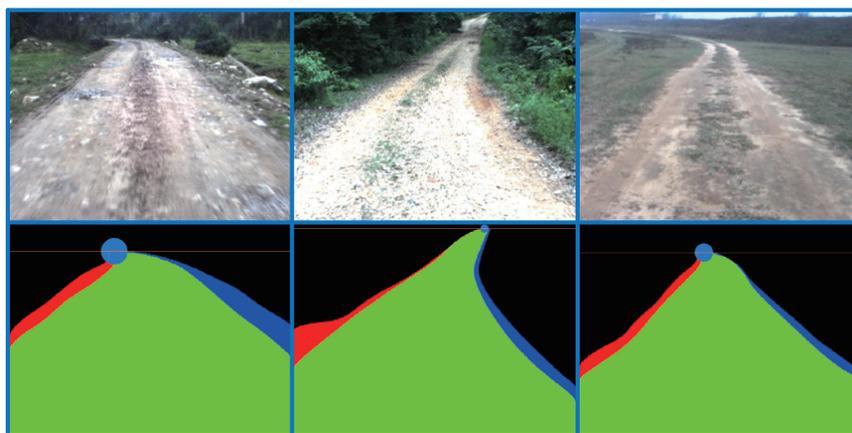

Figure 18. Images with ground truths in the database. The first rows are original images; the second rows are the corresponding ground truths; the road region is marked with the green color; ambiguous regions of the left and right edges are marked with red and blue colors, respectively.

As shown in Table 2, quantitative evaluations are provided using four types of pixel-wise measures: precision (P), accuracy (A), false positive rate (FPR) and recall (R). In Table 2, TP and TN denote the number of road pixels correctly detected and the background pixels correctly detected, respectively.

FP and FN denote the number of background pixels incorrectly marked and the road pixels incorrectly identified, respectively.

Table 2. Performance metric.

Pixel-Wise Measure	Definition
Precision	$P = \frac{TP}{TP+FP}$
Accuracy	$A = \frac{TP+TN}{TP+FN+FP+TN}$
False Positive Rate	$FPR = \frac{FP}{TN+FP}$
Recall	$R = \frac{TP}{TP+FN}$

In the experiment, we compare our proposed method to two state-of-the-art road detection algorithms: Kong *et al.*'s method [9] and Álvarez *et al.*'s method [10]. Quantitative results are shown in Table 3.

Table 3. Quantitative results on the OffRoadScene database.

Method	P(%)	A(%)	FPR(%)	R(%)
Kong *et al.*'s method [9]	98.1278	88.0466	2.1710	80.7174
Álvarez *et al.*'s method [10]	96.7897	84.9126	3.7357	76.4284
Our proposed method	**98.4708**	**97.0281**	**2.1243**	**96.5481**

5.2. Scale Sensitivity

In order to validate the sensitivity of algorithms to image scale, test images from the OffRoadScene database are resized to different scales. Let $m \times n$ be the size of images in the database; s denotes the scale factor, which resizes images to $(m \cdot s) \times (n \cdot s)$ $(s \in [0,1])$. A Gaussian filter is applied to smooth the image before resizing to a smaller scale. In the experiment, as shown in Figure 19, we resize images to 10 scales by setting s from 0.1 to 1 using a step-size of 0.1.

The results of the scale sensitivity evaluation are shown in Figure 20, which demonstrates that the precision, accuracy and recall of the proposed method are highest and and false positive rate is lowest among the three methods. Moreover, the precision, accuracy, recall and false positive rate of the proposed method are highly stable compared to the other two methods. This means that our proposed method is less sensitive to image scale.

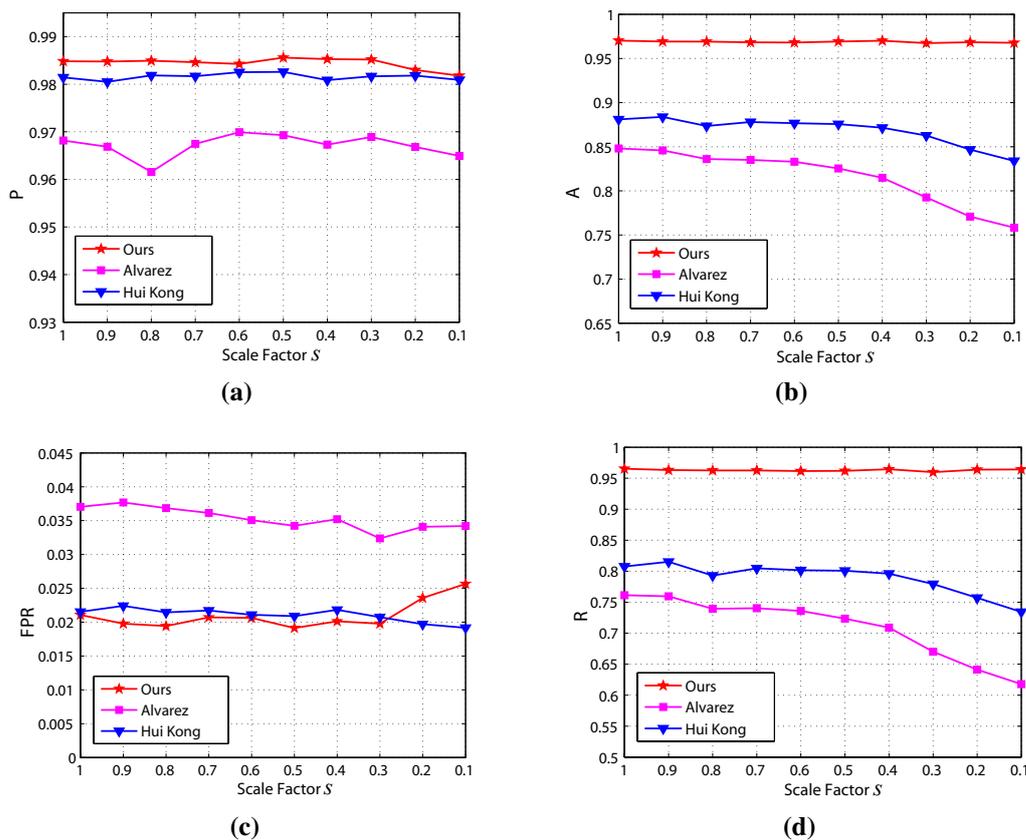**Figure 19.** Illustration of the scale sensitivity evaluation.

Figure 20. Results of the scale sensitivity evaluation: (**a**) precision; (**b**) accuracy; (**c**) false positive rate; (**d**) recall.

5.3. Noise Sensitivity

There may exist noise in the images captured by the on-board camera in a real scene, which is a challenge of vision-based road detection. In this work, Gaussian noise is added to images to evaluate the noise sensitivities of the algorithms:

$$\hat{I}(i,j,k) = I(i,j,k) + n(i,j,k) \tag{29}$$

where $\hat{I}(i,j,k)$ denotes that the noise with value $n(i,j,k)$ is added to the k-th channel of the pixel located in row i and column j of the image. $n(i,j,k)$ takes the form:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \tag{30}$$

where σ denotes variance. As shown in Figure 21, we add noise with variance σ ranging from 0.05 to 0.4 to validate the algorithms.

The results of the noise sensitivity evaluation are shown in Figure 22, which shows that the precision and accuracy of the proposed method are highest and the false positive rate is lowest among the three methods in most noise conditions. Besides, the precision, accuracy, recall and false positive rate of the proposed method are highly robust compared to the other two methods, which implies that the proposed method is less sensitive to noise.

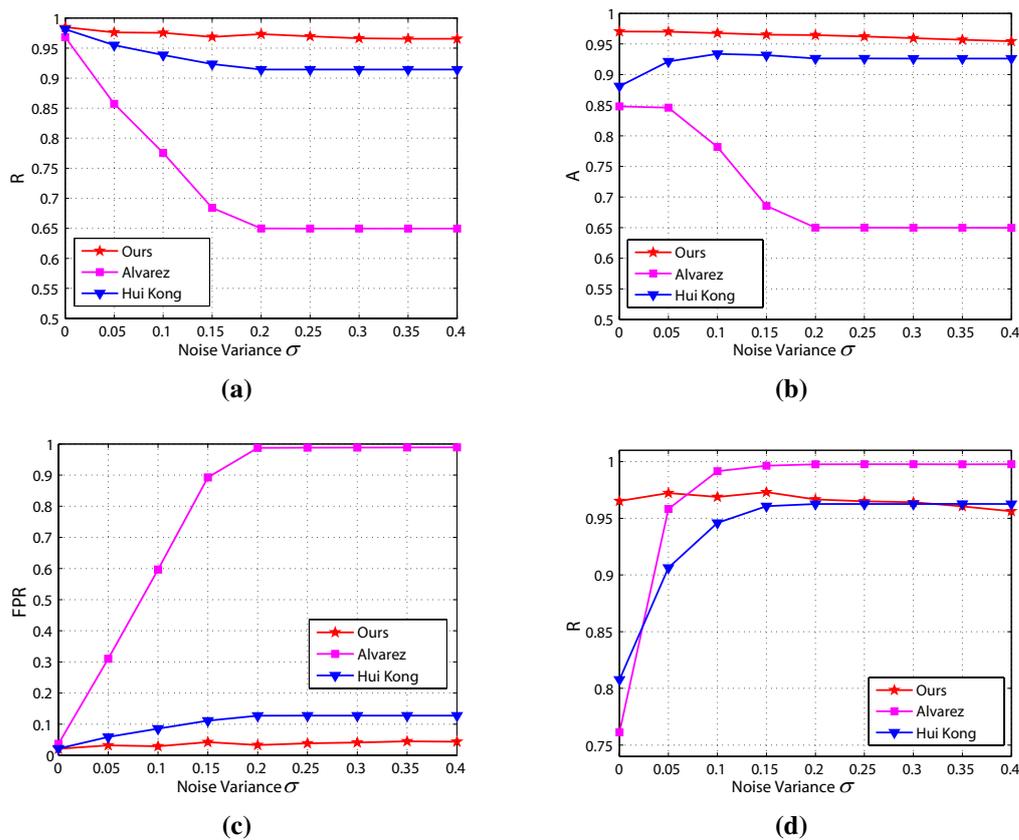Figure 21. Illustration of the noise sensitivity evaluation.

Figure 22. Results of the noise sensitivity evaluation: (**a**) precision; (**b**) accuracy; (**c**) false positive rate; (**d**) recall.

5.4. Discussion

As shown in Table 2, Figures 20 and 22, our proposed method performs much better than other methods. Some qualitative results can be seen in Figure 23, where Kong *et al.*'s method [9] is based on the road vanishing point. From the experiments, we can see that the precision of Kong *et al.*'s method is relatively high, and the false positive rate is relatively low. Besides, the method is less sensitive to scale and noise. This is due to the fact that texture-based vanishing point detection mainly depends on the edges of the main lanes instead of the details in road images, and the changes of scale and noise mainly affect the details of road images, while the edges of the main lanes are kept. However, the method may not fit well when the road boundaries are curved (see Figure 23, Rows 1 and 8).
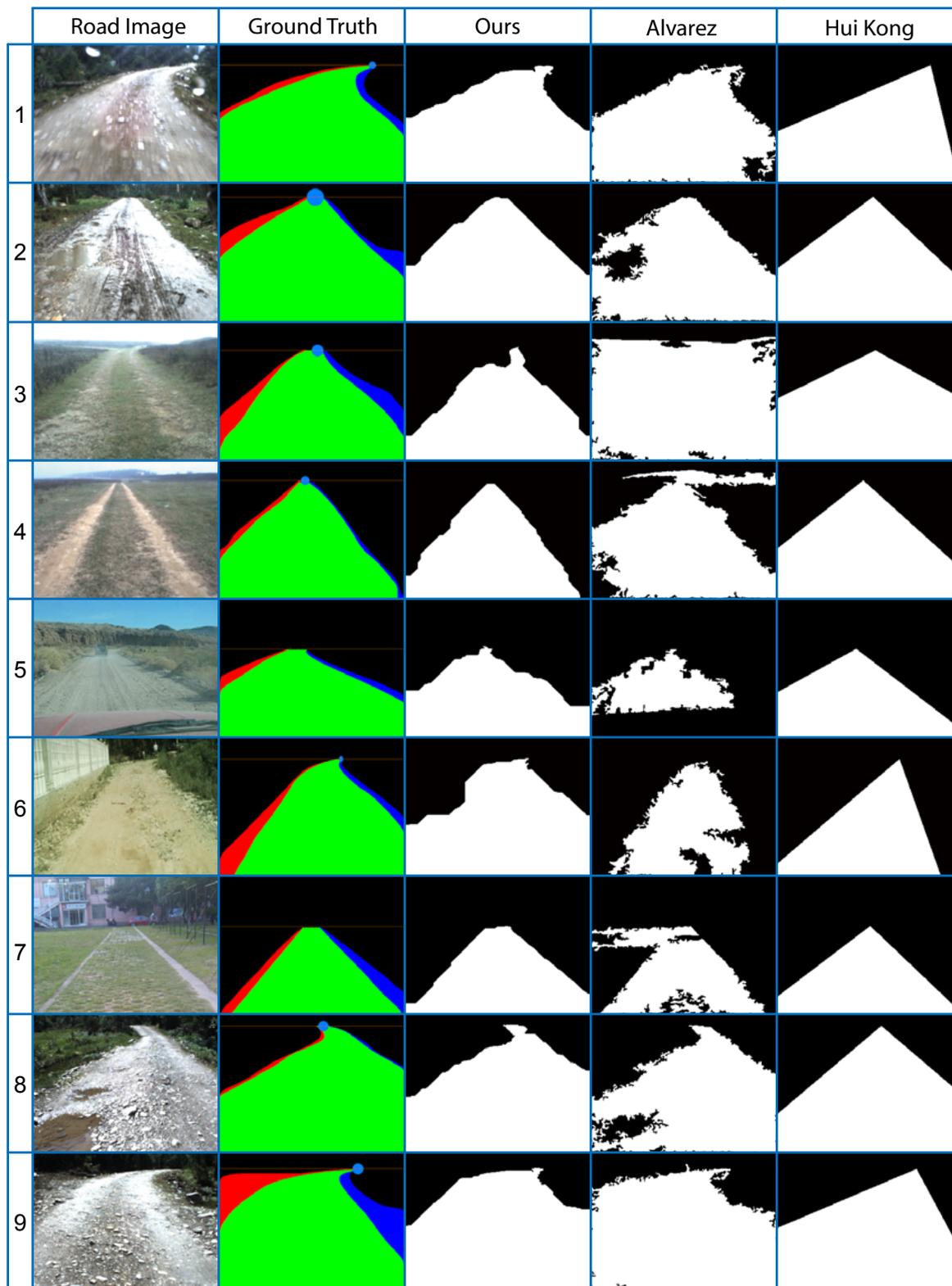
Figure 23. Qualitative comparisons between different methods. Road and background regions in Columns 3–5 are marked with white and black, respectively.

Álvarez *et al.*'s method [10] is based on the illuminant invariance feature, which is robust in shadowy and varying illumination conditions. The method is able to fit the road well when the road boundaries are curved (see Figure 23, Rows 1 and 8). However, it relies on the details of road images, so it has relatively

high sensitivities to scale and noise compared to Kong *et al*.'s method [9] and our proposed method. In addition, Álvarez *et al*.'s method would be suboptimal when the features of the road and background are too similar (see Figure 23, Rows 3, 4 and 7) or some parts of the road are quite different from other road regions (see Figure 23, Rows 2 and 8).

Our proposed method includes both advantages of Kong *et al*.'s [9] and Álvarez *et al*.'s method [10]. The method exploits the road vanishing point for seed selection and high-level information-based refinement, which make it depend less on the details of road images; thus, the method has low sensitivities to scale and noise. Moreover, GrowCut makes it possible to detect road regions when features of the road and background are quite similar. As shown in Figure 23, Rows 4 and 7, the proliferation of seeds will stop near the road boundaries.

6. Conclusions and Future Works

This paper presents a hierarchical road detection approach using a vision sensor. The major contributions of this work include: (1) an MPGA-based method proposed for efficient road vanishing point detection; (2) a road vanishing point and clustering-based strategy presented for unsupervised seed selection; (3) a GrowCut-based approach applied at the superpixel level introduced to obtain the initial road segment; and (4) a high level information-based method proposed to refine the road segment by using the CRF framework. Experimental results of the common performance, scale sensitivity and noise sensitivity evaluation validate the effectiveness and robustness of our proposed approach.

Our future work will focus on applying the proposed method to an SoC (system-on-a-chip) for some robot navigation applications, such as map building, path planning, *etc*.

Acknowledgments

Author Contributions

Keyu Lu conceived of the work, realized the algorithms and wrote the manuscript. Jian Li provided many suggestions for the work and made great efforts with respect to the paper revision. Xiangjing An and Hangen He provided their guidance.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Llorca, D.F.; Sánchez, S. Vision-Based Traffic Data Collection Sensor for Automotive Applications. *Sensors* **2010**, *10*, 860–875.
2. Tan, J.; Li, J.; An, X.; He, H. Robust Curb Detection with Fusion of 3D-Lidar and Camera Data. *Sensors* **2014**, *14*, 9046–9073.

3. Xu, W.; Zhuang, Y.; Hu, H.; Zhao, Y. Real-Time Road Detection and Description for Robot Navigation in an Unstructured Campus Environment. In Proceeding of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 27–30 June 2014; pp. 928–933.

4. Rasmussen, C. RoadCompass: Following rural roads with vision + ladar using vanishing point tracking. *Auton Robot* **2008**, *25*, 205–229.

5. He, Z.; Wu, T.; Xiao, Z.; He, H. Robust road detection from a single image using road shape prior. In Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP 2013), Melbourne, Australia, 15–18 September 2013; pp. 2757–2761.

6. Rotaru, C.; Graf, T.; Zhang, J. Color image segmentation in HSI space for automotive applications. *J. Real Time Image Proc.* **2008**, *3*, 311–322.

7. Christopher, R. Combining laser range, color, and texture cues for autonomous road following. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation (ICRA 2002), Washington, DC, USA, 11–15 May 2002; pp. 4320–4325.

8. He, Y.; Wang, H.; Zhang, B. Color-based road detection in urban traffic scenes. *IEEE Trans. Intell. Transp. Syst.* **2004**, *5*, 309–318.

9. Kong, H.; Audibert, J.Y.; Ponce, J. General road detection from a single image. *IEEE Trans. Image Proc.* **2010**, *19*, 2211–2220.

10. Álvarez, J.M.; Ĺopez, A.M. Road detection based on illuminant invariance. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 184–193.

11. Álvarez, J.M.; Ĺopez, A.M.; Gevers, T.; Lumbreras, F. Combining Priors, Appearance, and Context for Road Detection. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1168–1178.

12. Sotelo, M.A.; Rodriguez, F.J.; Magdalena, L. VIRTUOUS: Vision-Based Road Transportation for Unmanned Operation on Urban-Like Scenarios. *IEEE Trans. Intell. Transp. Syst.* **2004**, *5*, 69–83.

13. Vezhnevets, V.; Konouchine, V. "GrowCut"—Interactive multi-label n-d image segmentation by cellular automata. In Proceedings of the Fifteenth International Conference (GraphiCon'2005), Novosibirsk Akademgorodok, Russia, 20–24 June 2005; pp. 150–156.

14. Neumann, J.V.; Burks, A.W. *Theory of Self-Reproducing Automata*; University of Illinois Press: Champaign, IL, USA, 1966.

15. He, X.; Zemel, R.; Carreira-Perpinan, M. Multiscale conditional random fields for image labeling. In Proceedings of the 2004 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington, DC, USA, 27 June–2 July 2004; pp. 1953–1959.

16. Rasmussen, C. Following rural roads with vision + ladar using vanishing point tracking. *Auton Robot* **2008**, *25*, 205–229.

17. Moghadam, P.; Starzyk, J.A.; Wijesoma, W.S. Fast vanishing-point detection in unstructured environments. *IEEE Trans. Image Proc.* **2012**, *21*, 425–430.

18. Si-qi, H.; Hui, C. Studies on Three Modern Optimization Algorithms. *Value Eng.* **2014**, *27*, 301–302.

19. Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.

20. Zhang, X.; Zhang, Y.; Maybank, S.J.; Liang, J. A multi-modal moving object detection method based on GrowCut segmentation. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing, Orlando, FL, USA, 9–12 December 2014; pp. 1–6.

21. Zhang, C.; Xiao, X.; Li, X. White Blood Cell Segmentation by Color-Space-Based K-Means Clustering. *Sensors* **2014**, *14*, 16128–16147.

22. Veksler, O.; Boykov, Y.; Mehrani, P. Superpixels and Supervoxels in an Energy Optimization Framework. In Proceedings of the 11th European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 211–224.

23. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282.

24. Nguyen, T.V.; Tran, T.; Vo, P.; Le, B. Efficient Image Segmentation Incorporating Photometric and Geometric Information. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 6–18 March 2011; pp. 529–533.

25. Ara., N.; Gary, B. Detection of Drivable Corridors for Off-road Autonomous Navigation. In Proceedings of the IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA, 8–11 October 2006; pp. 3025–3028.

26. Veksler, O. Star shape prior for graph-cut image segmentation. In Proceedings of the 10th European Conference on Computer Vision (ECCV 2008), Marseille, France, 12–18 October 2008; pp. 454–467.

27. Szummer, M.; Kohli, P.; Hoiem, D. Learning CRFs Using Graph Cuts. In Proceedings of the 10th European Conference on Computer Vision (ECCV 2008), Marseille, France, 12–18 October 2008; pp. 582–595.

28. Shang, E.; Zhao, H.; Li, J.; An, X.; Wu, T. Centre for Intelligent Machines. Available online: http://www.cim.mcgill.ca/lijian/roaddatabase.htm (accessed on 15 April 2014).