Road detection is an important research topic in computer vision, which can be widely used in robot navigation, automatic driving, object detection in road scene and so on [1]. At present, the methods of road detection are mainly divided into two categories: based on low-level feature analysis of image and based on the understanding of the image geometry.

The methods based on low-level feature analysis of image are usually performed in units of pixels or blocks in the image. By analyzing their features of color [1–4], gradient (edge) [6,7], texture [5], or a combination of multiple features [2,5], pixels or blocks are divided into road region or non-road region. Most algorithms currently use a variety of feature information to process images. But for each feature, its limitations are inevitable and will affect the stability of algorithms. The color information is easily disturbed by the illumination environment, so the images are usually transformed to HIS, HSV, normalized RGB or the dark channel [8] space to weaken the influence of shadow. Moreover, some algorithms project the image into the log-log domain space to extract the gray image [9,25]. These algorithms are dependent on the effect of image transformation, and they are not always effective for complex...

---

**References**

1. [Cite Reference 1 here]
2. [Cite Reference 2 here]
3. [Cite Reference 3 here]
4. [Cite Reference 4 here]
5. [Cite Reference 5 here]
6. [Cite Reference 6 here]
7. [Cite Reference 7 here]
8. [Cite Reference 8 here]
9. [Cite Reference 9 here]
10. [Cite Reference 10 here]
illumination environment. The gradient information analysis is mainly used for edge detection\cite{3,10-14}, such as Ref \cite{12} using the watershed algorithm to detect the edge of the road, or fitting the boundary of road by gradient image. These methods are usually effective for structured roads with sign-lines, but the algorithms are not stable when the road boundary is not obvious or the road surface is cluttered. The texture information is usually used to obtain high-level features (such as vanishing points), or determine the road region by classifiers with others features \cite{5}, and the texture direction can be also used for image segmentation \cite{15}. Texture can overcome the effects of the illumination environment to some extent. But the calculation speed of processing texture information is slow, and the algorithms are not stable when the textures in the road surface are insufficient (such as gravel road, muddy road).

The methods based on the understanding of the image geometry usually extract some high-level features which can describe the geometry of the road like the vanishing point \cite{8,11,16,21} and the horizon line \cite{11,18,19,20}, and then estimate the range of the road region. These abstract features are extracted through the analysis of the low-level information and their accuracies are susceptible to the details of the image. In particular, the vanishing point detection is very time-consuming. Besides, the priori road structure model is needed and used to estimate the shape of the road region in the image. The model is regarded as a reference for road extraction like Ref \cite{19,23,24}. These methods cannot be flexible to adapt to the actual situation of the road image, and they are not applicable for dealing with unstructured road images in the wild environment. In addition, some algorithms based on machine learning or deep learning use low-level features and abstract high-level features.

In particular, there are few studies on the road detection algorithms for shadow-affected images, and these algorithms are only applicable for the simple shadow scenes, e.g., Ref \cite{9}. In most cases, shadows cannot effectively be overcome based on the gray image extracted by the minimum entropy method. Ref \cite{25} extracted the illumination invariance image with the method proposed in Ref \cite{26}, which can better weaken the shadow effect. But the algorithm for complex road images is not stable enough. Therefore, the road detection algorithm should not be only applied to ordinary images, but also should be robust for processing the special road images.

As mentioned above, both methods based on low-level feature analysis and methods based on understanding of the image geometry cannot achieve desirable performance in complex road images, especially shadow affected images. To solve those problems, a novel road region detection method based on quadratic estimation model is proposed, whose detection accuracy and computation time are significantly improved.

As shown in Fig. 1, it depicts an overview of our method. Firstly, the illumination invariant images are processed, and color analysis is performed based on the a priori triangular region as a road region sample. Then the probability maps of road region are generated to more significantly improve. The probability maps are used to estimate the range of the road region. These abstract features are extracted through the analysis of the low-level information and their accuracies are susceptible to the details of the image. In particular, the vanishing point detection is very time-consuming. Besides, the priori road structure model is needed and used to estimate the shape of the road region in the image. The model is regarded as a reference for road extraction like Ref \cite{19,23,24}. These methods cannot be flexible to adapt to the actual situation of the road image, and they are not applicable for dealing with unstructured road images in the wild environment. In addition, some algorithms based on machine learning or deep learning use low-level features and abstract high-level features.

As shown in Fig. 1, it depicts an overview of our method. Firstly, the illumination invariant images are processed, and color analysis is performed based on the a priori triangular region as a road region sample. Then the probability maps of road region are generated to more significant probability maps, which will be calculated by analyzing the probability histogram. Next, the road region is estimated for the first time as a known road model with the optimized probability map. Then, the gradient image is extracted from the illumination invariant image and the combined probability map. Finally, the road region can be obtained by analyzing the gradient image and the estimated road region.

The contributions of our proposed method are as follows:

1. In order to calculate more accurate distribution angle and eliminate shadow effects, the linear regression method is proposed to replace the minimum entropy method to obtain the distribution angle of pixels in log-log space and extract the gray image (in Section 2.2). The distribution angle calculated by our method is closer to the actual distribution shape of data. At the same time, it can reduce the shadow influence greatly by combining the illumination invariant image.

2. We proposed a more stable method to merge multiple probability maps (in Section 3.2) and a modified method for single probability map based on the relationship analysis between the regions of the image (in Section 3.3). Besides, a reset strategy for probability map is proposed based on the histogram analysis (in Section 4). Aimed at the histogram smoothing problem, this new adaptive method can be used to select the smoothing filter size. This method is based on the main frequency analysis at DCT domain.

3. For road detection, we proposed a quadratic estimation method to detect the road region (in Section 5). Different from the Ref \cite{19,23,24} methods using priori road models, our road model is estimated by color analysis. The estimated road model is more accord with the road region.

![Fig. 1. The overview of our method. A. the original image, B. the illumination invariance images, C. the probability maps, D. the merged probability map, E. the significance probability map, F. the estimated road region for the first time, G. the gradient image, H. the balanced gradient image, I. the result of road detection.](image-url)
2. Illumination invariant image extraction

For unstructured road images, the shadow on the road will seriously affect the complete extraction of the road region. In this paper, we use the method of Ref [26] to obtain the illumination invariant image through the single RGB original image, and further process each channel of the illumination invariant image to obtain the information needed for the subsequent processing.

In [26], the researchers found that the RGB-values of the shadow parts and the non-shadow parts in the image under natural illumination environment have a linear relationship in the Log-RGB domain. Thus, the ratio of the RGB-values of the light is assumed, and the color component of the object is separated from the illumination component in the illumination environment so as to obtain an illumination invariant image. The result of this method is shown in the Fig. 2(a).

To a certain extent, this method can weaken the influence of illumination and shadow. Since the elimination of illumination component is determined by these coefficients [25,26], this set of fixed coefficients is not sufficient to accurately describe the color components of the illumination for different images, and some shadow effects are still not completely removed.

After the illumination invariant image $I_{inv}$ is processed, three gray difference images will be obtained from the RGB-channels information of $I_{inv}$.

2.1. Gray image obtaining through the description of the log-log coordinate system

In the Ref [9], the illumination model in the environment of Planckian lighting, Lambertian surfaces, and a narrowband camera is established. According to the theoretical derivation and model simplification, the RGB image is projected to $\log(R/G)\log(B/G)$ coordinate system. Then, the method of Ref [9] obtains the optimal projection direction angle according to the minimum entropy theory. The coordinate points are projected into rectilinear direction of the angle, and the corresponding gray value is determined by the projection position. Finally, the RGB-channels image is compressed into a single-channel gray-scale image [27,28].

In fact, the minimizing entropy method is essentially to find an angle $\alpha$ for the distribution of pixels in $\log(R/G)\log(B/G)$ space to make the gray-scale image have the minimal information entropy, while the proposed method is not intended to eliminate the effects of illumination. In this paper, the linear regression method of statistics is used to calculate the direction angle $\alpha$ for $\log(R/G)\log(B/G)$ space. It makes the line with angle $\alpha$ describe the distribution of the pixels in $\log(R/G)\log(B/G)$ space better. Thus, the gradation of generated gray-scale image is highlighted and illumination effects are weakened. Compared with the minimum entropy method, our method is more stable, the calculation speed is faster, and the direction angle is more accurate.

For RGB image $I$, the RGB values of pixel $p_i \in I$, $i = 1...N$ is $(R(p_i), G(p_i), B(p_i))$. For $I$, take $X = (x_1, x_2, ..., x_N)$, where $x_i = \log\left(\frac{R(p_i)}{G(p_i)}\right) - \log\left(\frac{G(p_i)}{B(p_i)}\right)$, and $Y = (y_1, y_2, ..., y_N)$, where $y_i = \log\left(\frac{B(p_i)}{G(p_i)}\right) - \log\left(\frac{G(p_i)}{R(p_i)}\right)$, then the variances $\text{var}(X)$ and $\text{var}(Y)$ of $X$ and $Y$, and the covariance $\text{var}(XY)$ are calculated.

Even $X$ and $Y$ may have a correlation, the linear regression angle $\alpha_x = \arctan(\text{var}(XY)/\text{var}(X))$ for $X$ is likely to be different from the angle $\alpha_y = \arctan(\text{var}(Y)/\text{var}(XY))$ for $Y$. In order to make the obtained angle more stable, let $\alpha = (\alpha_x + \alpha_y)/2$. Where $\alpha$ is the required direction angle. Then the gray-scale image $G_{inv}$ can be gotten. For pixel $p_i \in G_{inv}$, $i = 1...N$, $G_{inv}(p_i) = x_i \cos(\alpha) + y_i \sin(\alpha)$.
As shown in Fig. 2(b), the linear regression method is easier to distinguish road region and other regions than the minimizing entropy method. Besides, Fig. 2(c) shows that the direction angle of the linear regression method is closer to the distribution of pixels than minimizing entropy method. Obviously, the linear regression method has a better effect.

In this paper, the linear regression method is used to obtain the angle $\alpha$ of the $\log(R/G) - \log(B/G)$ space for the image $I_{inv}$ (As Fig. 2(d) shown, the result of process $I_{inv}$ is better than processing the original image $I$) and get the gray-scale image $G_{inv}$.

2.2. Gray difference images obtaining

In order to describe the degree of color difference between the road region and other regions in the image, the illumination invariant image $I_{inv}$ and gray-scale image $G_{inv}$ are used to get three gray-scale difference images. According to Ref [8] and Ref [17], the road region is often approximate a triangular region in the road images. Thus, we defined the priori triangular region $\Omega_{tri}$ as shown in Fig. 3(a). It represents the range of the first road estimation model, and it is used to calculate the average value of the road region in gray-scale images.

The specific calculation method is as follows:

1. In order to filter out the influence of noise points in the image on the extraction of gray difference images, the gray-scale image $G_{inv}$ is smoothed by Gaussian convolution filter, and the obtained image is denoted as $\tilde{G}_{inv}$. The average value in $\Omega_{tri}$ of $\tilde{G}_{inv}$ is denoted as $\tilde{G}_{inv}$. And then the gray difference image $G_{inv}$ is obtained. For pixel $p_i \in G_{inv}, i = 1...N$, $G_{inv}(p_i) = |G_{inv}(p_i) - \tilde{G}_{inv}|$.

2. Three channels image of the illumination invariant image $I_{inv}$ can be processed with the method described above step. The average gray values $(\bar{R}_{inv}, \bar{G}_{inv}, \bar{B}_{inv})$ for the three channels of $I_{inv}$ are used to get the final difference image $C_{inv}$. For pixel $p_i \in C_{inv}, i = 1...N$, $C_{inv}(p_i) = \sum_f |\tilde{u}_{Hf}(p_i) - \bar{H}_f|/\max(|\tilde{u}_{Hf}(p_i) - \bar{H}_f|)$, where $H \in \{R, G, B\}$, and $\tilde{u}_{Hf}$ is the smoothed images of $I_{inv}$ after Gaussian filtering.

3. We project the image $I_{inv}$ into HSV space and smooth the chromaticity image $H_{inv}$ by Gaussian filter to get the image $H_{inv}$. The average value $\bar{H}_{inv}$ is calculated in $\Omega_{tri}$ of $H_{inv}$, and then the gray difference image $H_{inv}$ can be calculated (see Eq. (1)) for pixel $p_i \in H_{inv}, i = 1...N$.

$$H_{inv}(p_i) = \left\{ \begin{array}{ll} H_{inv}(p_i) - \bar{H}_{inv} & |H_{inv}(p_i) - \bar{H}_{inv}| \leq 180 \\ 360 - |H_{inv}(p_i) - \bar{H}_{inv}| & |H_{inv}(p_i) - \bar{H}_{inv}| > 180 \end{array} \right.$$

(1)

The Gaussian convolution filter $F_2$ used in this section is defined as follows: For each point $s$ in the filter, its weight is $F_2(s) = \frac{1}{K}\exp((-|d_x(s)^2 + d_y(s)^2|)/2d^2)$, where $d_x(s)$ and $d_y(s)$ are respectively the distances from the current point to the center of the filter in the horizontal and vertical directions, $d$ is the size of filter, $K = \sum s \exp((-|d_x(s)^2 + d_y(s)^2|)/2d^2)$ is the normalized coefficient. The gray difference images $G_{inv}, C_{inv}$ and $H_{inv}$ are shown as Fig. 3(b).
3. The probability maps calculation and processing

3.1. Road region probability maps calculation

For the gray difference image $D$, the value of one pixel represents the difference between this pixel and the standard pixel samples in a description of a particular feature. The lower the value of the pixel, the smaller the difference between them, and the more similar they are.

The probability map $P$ can be generated using the difference image $D$. For the pixel $p_i \in P, i = 1...N$, $P(p_i) = 1 - D(p_i)/\max(D(p_i) | p_i \in D)$ is its probability value. The histogram equalization of $P$ can increase the probability contrast of pixels in road or non-road regions of the probability map. In this step, we take the relatively large (and relatively small) probability values into 1 (and 0), and the probability values of the other pixels are adjusted according to the proportion.

Above method is used to process the image $G_{inv}, C_{inv}, H_{inv}$ to obtain the corresponding probability maps $P_G, P_C, P_H$ (shown in Fig. 4(b)), respectively.

3.2. Probability maps merging

Three probability maps $P_G, P_C, P_H$ have shown the possible location and scope of the road region to a certain extent. But for a certain pixel, its probability values on three maps are usually not equal. In many cases, some probability maps do not describe the target region very well. For many pixels, not every probability value of their probability map is worthy to reference. Thus, we use a method that can help us to derive a more reliable probability distribution by analyzing the description of multiple probability maps.

---

Fig. 4. Generation and processing of the probability maps. (a) The method of calculating the auxiliary probability maps $P_G$ (shown as the first row) and $P_H$ (shown as the second row). (b) The calculating process for generating and merging the probability maps.
For a set of probability maps \([P_1, P_2, \ldots, P_k]\), a pixel \(p\) in them has the probability values as \((P_1(p), P_2(p), \ldots, P_k(p))\). Sort these values from low to high as \((P_1'(p), P_2'(p), \ldots, P_k'(p))\), and we can take the median value \(P_{\text{med}}'(p)\), the highest value \(P_k'(p)\), and the averaging value \(P(p)\) for them. For pixel \(p\), some appropriate probability values can be chosen to represent its true probability value \(P(p)\) according to its probability values distribution by (2).

If \(P_{\text{med}}'(p)\) is large enough (decided by \(t_1\)), it means that there are at least half of the probability maps to judge the pixel \(p\) whether have a great probability value. If \(P(p)\) is low (decided by \(t_2\) and the \(P_k'(p)\) is not large enough (decided by \(t_3\), the \(p\) is less likely to belong to the target region. Other cases are relatively complex, so we use \(P(p)\) as its final probability value.

In Eq. (2), \(t_1, t_2, t_3 \in (0, 1)\). \(P(p)\) is more dependent on the total probability values. When \(t_1\) and \(t_2\) are lower and \(t_3\) is higher, then the merged probability map is more likely to be affected by the irrational probability values.

\[
P(p) = \begin{cases} 
\frac{1}{k-mid+1} \sum_{i=mid}^{k} P_i'(p) & P_{\text{med}}'(p) > t_1 \\
\frac{1}{k-mid} \sum_{i=1}^{mid} P_i'(p) & P(p) < t_2 \& P_k'(p) < t_3 \\
P(p) & \text{in other case for } p 
\end{cases}
\]

Above method is used to process the probability maps \(P_1, P_2, P_3\), and get a more reliable probability map \(P\). According to a lot of experiments, in this paper, \(t_1 = 0.9\), \(t_2 = 0.6\), \(t_3 = 0.9\).

3.3. Probability map correction

For a probability map \(P\), there are often some pixels or blocks surrounded by a region with a higher(lower) probability value, but its own probability value is low(high). We believe that the enclosed pixels or blocks are the parts of the region which surrounds it, and their probability values are needed to be corrected based on the outside region.

For probability map \(P\), we can take two auxiliary probability maps \(P_{t_1}\) and \(P_{t_2}\). In \(P_{t_1}\) or \(P_{t_2}\), the pixels or blocks surrounded by a region with a higher (or lower) probability values can be changed to a higher (or lower) probability values. These processing methods are shown as Fig. 4(a).

Pixel \(p\) has three probability values: \((p), P_{t_1}(p)\) and \(P_{t_2}(p)\), then the probability value \(P(p)\) for \(p\) can be corrected by Eq. (3).

\[
P(p) = \begin{cases} 
P_{t_1}(p) & P(p) - P_{t_1}(p) < \varepsilon \& P_{t_2}(p) - P(p) \geq \varepsilon \\
P_{t_2}(p) & P(p) - P_{t_1}(p) \geq \varepsilon \& P_{t_2}(p) - P(p) < \varepsilon \\
P(p) & \text{in other case for } p 
\end{cases}
\]

The corrected probability map \(P\) exhibits the characteristics of blocking in the image range. The blocks with the higher probability value have a significant difference from those blocks with lower probability value. Within each block, the change of probability value is smoother. And the calculation process for the probability map \(P\) is shown in Fig. 4(b).

4. Probability map reset

Although the road region in the probability map \(P\) has been highlighted to a certain extent, the probability value differences between road region and non-road region are not often obvious enough, so that it is not easy to divide them well with a constant probability threshold. The probability map \(P\) needs to be resettled.

In this paper, the probability map \(P\) has been transformed into ordinary gray-scale image to process. In order to eliminate the high-frequency fluctuation in the histogram and keep its macro-low frequency fluctuation tendency, the histogram needs to be smoothed and then the trough value is found in the histogram as the segmentation threshold. And the probability values of the pixels in new probability map \(P'\) can be reset according to the threshold segmentation result of the histogram to increase the difference degree of probability between the regions with high probability and low probability.

4.1. Histogram smoothing based on frequency analysis

For different histograms, it is difficult to determine the fixed filter size. Therefore, a self-adapting method is presented to confirm the size of filter by analyzing the main frequency \(\omega\) of the spectrum obtained by taking the DCT transform for the histogram (To some extent, \(\omega\) can describe the fluctuant trend in whole view of the histogram). The cycle of this main frequency \(\omega\) can be calculated to \(d_{\text{filter}}\), which is the size of filter.

The specific processing is as follows:

1. The frequency histogram \(\text{Hist}_f\) is obtained by taking the DCT transform for gray histogram \(\text{Hist}_g\). Then, \(\text{Hist}_f\) is divided into several frequency groups. Next, the average frequency \(\bar{s}(i)\) and the maximum frequency \(s_{\text{max}}(i)\) for each frequency group \(D_f(i)\) are calculated.

2. The candidate frequency groups: \(D_f(i); i \in \{i | s_{\text{max}}(i) \leq 0.2 s_{\text{max}} \text{ and } i > 3 \}\) and \(\bar{s}(i) \leq \bar{s}(i-1) \text{ and } \bar{s}(i) \leq \bar{s}(i+1)\) are chosen. Where \(s_{\text{max}}\) is the maximum frequency of \(\text{Hist}_f\).

3. The set of candidate frequencies \(\omega_{\text{max}}(i)\) can be obtained, where \(\omega_{\text{min}}(i)\) is the minimum intensity for candidate frequency group \(D_f(i)\). The minimum frequency \(\omega \in \{\omega_{\text{min}}(i)\}\) is regarded as the main frequency of \(\text{Hist}_f\). Then the half size of filter
d₀ = \left[ \min(20, \max(N_{\text{hist}}/\omega)), 5 \right], \text{ where } N_{\text{hist}} \text{ is the series of } \text{Hist}_P. \text{ And } d_{\text{filter}} = 2d₀ + 1 \text{ is the size of filter.}

(4) A one-dimensional Gaussian filter \( F \) is constructed, and \( \text{Hist}_P \) is smoothed to take \( \text{Hist}_P \) with filter \( F \). Here, \( F(x) = \frac{1}{K} \exp\left(-\frac{(x - d₀ - 1)^2}{2d₀^2}\right) \), \( K = \sum_{x=1}^{d_{\text{filter}}} \exp\left(-\frac{(x - d₀ - 1)^2}{2d₀^2}\right) \).

4.2. Probability map saliency optimization method

The new probability map \( P' \) can be calculated as follows:

(1) The probability image \( P' \) can be transformed to gray-scale image \( I_P \). The larger probability values and the smaller probability values are drawn into 255 and 0, and other pixels are evaluated the gray-scale values from 0 to 255 according to their original values, evenly. And the gray-scale image \( I_P \) can be taken. Then we take \( I_P(p) = 255 - I_P(p) \) for each pixel \( p \in I_P \). For those pixels in \( I_P \) with lower values, they are more likely to belong to the road region in the feature of color.

(2) The histogram \( \text{Hist}_P \) of \( I_P \) can be calculated as a sample sequence. As shown in Fig. 5, the sequence has obvious fluctuation (the low frequency fluctuation) as a whole. But there is also the high frequency fluctuation in the local intervals. Smoothing \( \text{Hist}_P \) with the Gaussian filter \( F \) obtained by the method of Sec. 4.1 can remove the high frequency interference at the detail and preserve the frequency of the histogram.

(3) For \( \text{Hist}_P \), it defines the local neighborhood width by \( d_{\text{filter}} \), and find the maximum frequency gray levels \( \{h_{\text{peak}}\} \) as the peaks of \( \text{Hist}_P \). The lowest frequency gray level between two adjacent peaks is the valley \( h_{\text{val}}(i) \), which is the split threshold \( T_i \) of each local region of \( \text{Hist}_P \).

(4) For \( \text{Hist}_P, \{T_i\}_{i=1,2,\ldots,N_P} \), the two adjacent thresholds i.e., \( T_i \) and \( T_{i+1} \), can be used to determine a gray interval \( \rho_i, \tilde{h}_i \) (i) is the average gray value of \( \rho_i \), and \( N_P(i) \) is the total number of pixels of \( \rho_i \).

(5) The transfer coefficient \( \varphi \) can be calculated. For two adjacent gray intervals \( \rho_i \) and \( \rho_{i+1} \), the transfer coefficient \( \varphi(i, i + 1) = \exp(-\eta(i, i + 1)), i = 2,3,\ldots,N_P, \) where \( N_P = N_P - 1 \) is the total number of gray intervals, and \( \eta(i, i + 1) = \eta_\varphi(i, i + 1) \times \eta_\varphi(i, i + 1) \). \( \eta_\varphi(i, i + 1) = \frac{\tilde{h}_i(i) - \tilde{h}_\varphi(i)}{d_{\text{filter}}} \) is the correlated coefficient of gray-scale value, and \( \eta_\varphi(i, i + 1) = \min(N_P(i)/N_P(i + 1), N_P(i + 1)/N_P(i)) \) is the correlated coefficient of the pixel number. It is mean that if two adjacent gray intervals have similar gray values, or the number of pixels for one interval is less than the other one, obviously, they should have more similar probability values, and their transfer coefficient for probability will be large.

(6) The new probability values for segmentation thresholds \( \{T_i\}_{i=1,2,\ldots,N_P} \) can be calculated by (4).

\[
\begin{align*}
P'(T_i) &= P'(T_{i-1}) = 1, \\
P'(T_{i+1}) &= P'(T_i) \times \varphi(i, i) = 2,3,\ldots,N_P
\end{align*}
\]

(7) The new probability values of other gray levels can be calculated. For the gray level \( h_j \in \rho_i \), its significance probability is

\[
P'(h_j) = \frac{N(T_i, h_j)}{N(T_i)} \left[ P'(T_{i+1}) - P'(T_i) \right] + P'(T_i), \text{ where } N(T_i, h_j) \text{ is the number of pixels with gray values between } [T_i, h_j].
\]

Thus, the new probability map \( P' \) (shown in Fig. 5) can be obtained, which is more significant by the above-mentioned gray-scale adjustment with the probability map \( P \). Compared with the probability map \( P, P' \) is more prominent in the high probability region of \( P \), and further weakens the probability value of the low probability region in \( P \).

5. Road region detection

5.1. Road region estimation

As previously described, our probability maps \( P \) and \( P' \) are obtained by gray-scale analysis and the RGB-channels of the
illumination invariant image $I_{inv}$ processing, which can show the possible road region in the image space after removing the illumination effect. Especially in $F'$, the road region is obviously enhanced.

The Otsu method [29] as the threshold segmentation method is used to process the $F'$ and obtain the road region marked by binary map $S_{Road}$, where those pixels labeled 1 are regarded as road region because they have a larger probability value in $F'$, and their colors are more closer with the road sample region $Ω_{road}$. Similarly, those pixels labeled 0 are regarded as non-road region.

As in usual road images, the road region is located in the middle-lower part of the image. Those connectivity domains belong to road region $Ω_{road}$, where domains labeled 1 in $S_{road}$ and their centers are located in the middle-lower part of the image. In this paper, the middle-lower region is a rectangular region in image space. And the connectivity domains should be large enough. The connectivity domains that don’t conform to these rules in $S_{road}$ are marked as 0.

We found that the road region $Ω_{road}$ which is marked as 1 in $S_{road}$ may be still incomplete or incorrect in some cases. It is because we only consider the color information of the image, and don’t consider the gradient of color in the image space. Therefore, the effective road edge information needs to be extracted in order to further correct the slope of the road region.

5.2. Gradient image extraction and correction

Firstly, the gradient image $E_{inv}$ is obtained from $I_{inv}$, which has smoothed the illumination invariant image $I_{inv}$ with Gaussian filtering. $I_R$, $I_G$, $I_B$ are the three-channels image of $I_{inv}$, and they are filtered by the gradient filter operator $F = [1, -1]$, $F = [1, -1]$ to obtain the gradient images $|E_{Rx}|$, $|E_{Gy}|$. For pixel $p(x, y)$, $E_{H}(p) = (|E_{Rx}(x, y) - E_{Rx}(x + 1, y)| + |E_{Gy}(x, y) - E_{Gy}(x, y + 1)|) / \max_{q \in E_{H}}(q)$. Then the gradient image $E_{inv}$ can be taken by Eq. (5):

$$E_{inv}(p) = \min_{h \in \{R, G, B\}} E_{H}(p)$$

(5)

For unstructured road images, the parts of the non-road region (even some parts inside the road region) are often rich in gradient information. The illumination invariant image $I_{inv}$ preserves the gradient information and the edge of road boundary may be not as obvious as these ineffective edges. Thus, we also need to strengthen the gradient strength of the edge of road boundary. And the probability image $P$ has a characteristic of blocking. Then we can get the gradient image $E'$: for pixel $p$, $E'(p) = \min(E_{inv}(p), E_{p}(p))$.

In order to enhance the differences of gradient strength for different pixels, we balance $E'$ as following method:

Assuming the gray-scale of $E'$ is $[0, T]$, we balance the values of pixels belonging to $[T_0, T]$ in the gray-scale region to $[0.2T, T]$. For the original gray value $h \in [T_0, T]$, the new gray value can be calculated by Eq. (6):

$$h = \frac{\sum_{h=T_0}^{T} N(h) - 0.8T}{\sum_{h=T_0}^{T} N(h)} + 0.2T$$

(6)

Other pixels values are balanced to $[0.02T]$, then the gradient image $E'$ can be obtained. The road boundary in $E'$ is much clearer than $E$.

5.3. Road region detection

In order to use the gradient image $E$ to detect road region, effectively, the information of invalid gradient within road region should be estimated. Then the threshold of road boundary extraction is determined adaptively according to the strength of these invalid gradient.

In $E$, the pixels belonging to $Ω_{road}$ are chosen as the sample of road region and their gradient strength are counted. The threshold of road boundary $T_E$ is the minimum strength of pixels. The binary map $S_{inv}$ can be taken by Eq. (7):

$$S_{inv}(p) = \begin{cases} 1 & E(p) \leq T_E \\ 0 & E(p) > T_E \end{cases}, \text{for each pixel } p$$

(7)

In fact, $S_{inv}$ is a connected domain map, and each connected domain is separated by a more obvious edge. We choose those connection domains with the common pixels of $Ω_{road}$ as the road region $Ω_{road}$, $Ω_{road}$ (or some parts of it) and the sample of road region $Ω_{inv}$ are similar in color, and the edge of $Ω_{road}$ is more obvious.

After correcting the $Ω_{road}$ by morphological operation (expansion, corrosion, etc.), we fill the holes and remove the sky region from $Ω_{road}$, then we can get the more structured road region $Ω_{road}$ as the result we need. The process of road region detection is shown in Fig. 6.

6. Experimental results and analysis

6.1. The sample library

The sample library² (published online), which is implemented in this paper, contains 409 road images. It is built based on the

² https://pan.baidu.com/s/1tjC4tKHCYcfabgmw1HX4w
dataset in Ref [30]. The experimental image samples in different environments and shooting conditions are derived from the Internet. Among them, there are 43 structured road images and 366 unstructured road images. Many of them have shadow influence. In this paper, all images are regulated to the same size 240 × 320 to build the standard sample library.

6.2. Results comparison

As we know that the methods based on deep learning or machine learning need a lot of training and testing data. While the traditional methods do not need dataset to learn the detection model. In order to evaluate the algorithms fairly, considering the operating platform factor, we compared five road extraction methods related to our methods, e.g. Wang method [14], Kong method [17], Ding method [22], Duan method [25] and Li method [31]. Road region estimation algorithms of Kong method [17] and Ding method [22] are based on vanishing point location and boundaries. Wang method [14] and Duan method [25] estimate road region mainly based on image segmentation and vanishing point information. Li method [31] estimate road region based on road region probability map and vanishing point location. Among them, Kong method [17], Duan method [25] and Li method [31] can reduce the impact of shadows on road extraction. All programs are running in MATLAB R2013a which is installed with 64bit Windows7 on the Intel(R) Core(TM) i5-4590 CPU 3.30GHz.

6.2.1. Experimental results

In Figs. 7–9, the first column shows the original road images, the second column shows the ground truth maps, the third column shows the results of Wang method, the fourth column shows the results of Kong method, the fifth column shows the results of Ding method, the sixth column shows the results of Duan method, the seventh column shows the results of Li method, and the eighth column shows the results of the proposed method. Figs. 7 and 8 show the results of structured roads and unstructured roads detection using the above methods, respectively.

Qualitative comparisons are shown in Figs. 7–9. Kong method and Ding method use the vanishing point to estimate the triangular road region. These two methods always can’t extract the complete road region. In the case where the vanishing point position is not accurate or the selected road boundaries are not accurate, the road region can’t be extracted accurately, and the triangular road region can also not accurately describe the extent of the unstructured road region. In the most sample images in Figs. 7–9, Wang method can’t extract the road. The results of this method are greatly affected by region segmentation. The segmentation method uses region growth method and has a poor robustness. While, Duan method can extract a more accurate road region, but for some cases,
the road region may be incomplete. This method is over dependent on the position of the vanishing point and the result of the binary segmentation. The binary segmentation method is not robust for complex road image with many types of objects. Besides, Wang method and Duan method can’t detect the correct vanishing point position due to the poor results of lines detection in unstructured road images. However, Li method and our method can detect more accurate road region than the other compared methods. These two methods don’t need to detect lines in the road images, which can make the road region more smooth and completed. The boundaries of the road regions detected by our method are not sharp enough, but our method is more stable to extract the road region correctly for most images. Although the road regions extracted by Li method and our method have smaller differences, our method extracted road region without vanishing point location. The extracted road region of Li method is affected by the vanishing point position, which may cause more errors of the extracted road region. As shown in Fig. 9, Wang method, Ding method, Duan method are not robust for some special road images. These four methods cannot extract the road region, effectively. Kong method, Li method and our method can extract more accurate road region. But our method can extract the road region more robust and accurate than Kong method and Li method for these complex images.

### 6.2.2. Comparison of accuracy

In order to efficiently evaluate the accuracy of the algorithms, four evaluation indicators are used in this paper: precision (PRE), accuracy (ACC), false positive rate (FPR) and recall (TPR). They are calculated as (8).

\[
\begin{align*}
\text{PRE} &= \frac{T_P}{T_P + F_P} \\
\text{ACC} &= \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \\
\text{FPR} &= \frac{F_P}{F_P + T_N} \\
\text{TPR} &= \frac{T_P}{T_P + F_N}
\end{align*}
\]

Where \(T_P\) and \(F_P\) are the total numbers of road region pixels which are determined correctly and incorrectly. \(T_N\) and \(F_N\) are the total numbers of non-road region pixels which are determined correctly and incorrectly.

Table 1 shows the evaluation results of \(\text{PRE}, \text{ACC}, \text{TPR}\) and \(\text{FPR}\) for the standard sample library where \(E\) is the average value, and
SRt is the standard deviation.

For Wang method, the PRE, ACC and TPR are the worst in these methods, and the FPR is only better than Kong method. For Kong method, the PRE is obviously lower than ACC and TPR, and its FPR is relatively high, which indicates that there may be over-estimation problem for the triangle road region by the vanishing point detection. For Ding method and Duan method, the TPR is the highest, which demonstrates that it can extract the region more complete. While the PRE of Li method is obviously lower than our method, which proves that Li method has more overestimation problem than our method. Compared with these six methods, our method gets the best performance on PRE, ACC and FPR, and the TPR is only lower than Li method. Considering the stable of the method, the standard deviation of PRE, ACC, TPR of our method are lower than others, and the FPR is only higher than Duan method, which indicates that our method is more stably. Thus, our method is superior to the other methods overall.

6.2.3. Comparison of calculation time

In order to compare the computational efficiency of the three methods, the average value and the standard deviation of the computation time of all methods are listed in Table 2. Where \( \bar{t} \) is the average calculation time (second per image) for the sample library, and \( SR_t \) is the standard deviation of calculation time for them.

Obviously, our method is much faster than the two other methods in calculating speed aspect. And according to the \( SR_t \) of different methods, the running time of our method is more stable. It means that our method is more easily to be controlled for the time consumption in practical applications.

6.3. The algorithm analysis

Our method is based on the quadratic estimation to extract road region. The priori triangular road region is regarded as the first road region model to take the road color samples for illumination invariant image. By analyzing the probability maps in color, the estimated road region is obtained as the second road region model. After that, the effective road boundary is extracted by gradient detection of the illumination invariant image and the probability map, as well as the gradient analysis of the second road region model. Then the road region is finally obtained. As shown in Fig. 10, there are some unsuccessful examples of our method. Although our method can adapt to different types of images, it may lead to problems in some special images, such as over estimation or leak estimation because the edges of the road boundaries are not be detected which caused by the inconspicuous edge gradient, or the color differences in the road region are too large (e.g. the shadow or other illumination influence are not eliminate, effectively).

Overall, our method has the following problems:

(1) The performance of the proposed method is affected by the effect of illumination invariant image. For some road images with complex shadows, the illumination invariant images can’t eliminate the shadow effects on the road region completely.

(2) Although the road regions are always unsmoothness, for the road images with large differences in road region, the combination and optimization of the probability map may make the difference within the road region more obvious.

(3) The extraction of the effective road boundary in the gradient image depends not only on the trimming effect of itself, but also on

\[ \begin{align*}
\text{Table 1} \\
\text{The evaluation results of the methods. (The bold are the best results).} \\
\begin{array}{|c|c|c|c|}
\hline
\text{Method} & \text{PRE} & \text{ACC} & \text{FPR} & \text{TPR} \\
\hline
\text{Wang method [14]} & 0.4786 \pm 0.4107 & 0.7981 \pm 0.1337 & 0.0582 \pm 0.0908 & 0.4189 \pm 0.4183 \\
\text{Kong method [17]} & 0.6866 \pm 0.0636 & 0.8473 \pm 0.0205 & 0.1539 \pm 0.0290 & 0.8855 \pm 0.0468 \\
\text{Ding method [22]} & 0.8523 \pm 0.1830 & 0.8243 \pm 0.1034 & 0.0347 \pm 0.0516 & 0.5040 \pm 0.2706 \\
\text{Duan method [25]} & 0.8426 \pm 0.0945 & 0.9186 \pm 0.0135 & 0.0171 \pm 0.0011 & 0.7728 \pm 0.0995 \\
\text{Li method [31]} & 0.9138 \pm 0.0157 & 0.9496 \pm 0.0029 & 0.0344 \pm 0.0028 & 0.9202 \pm 0.0120 \\
\text{Our method} & 0.9607 \pm 0.0074 & 0.9574 \pm 0.0013 & 0.0161 \pm 0.0022 & 0.8889 \pm 0.0062 \\
\hline
\end{array}
\end{align*} \]

\[ \text{Table 2} \\
\text{The computation time of the methods (second per image). (The bolds are the best results).} \\
\begin{array}{|c|c|}
\hline
\text{Method} & \bar{t} \text{ (second per image)} \\
\hline
\text{Wang method [14]} & 0.7404 \\
\text{Kong method [17]} & 58.9353 \\
\text{Ding method [22]} & 1.9975 \\
\text{Duan method [25]} & 0.8892 \\
\text{Li method [31]} & 0.2453 \\
\text{Our method} & 0.1553 \\
\hline
\end{array} \]
the second road region model. In certain cases, it still can’t avoid the interference for the obvious texture inside the road. Or road extraction failure due to road borders can’t be effectively enhanced.

7. Conclusion

In this paper, a road detection method is presented which based on illumination invariant image and quadratic estimation. The algorithm obtains the first estimated road region by color analysis with a priori triangular road model, and the final road region is detected by gradient information analysis with the first estimated road region. The experimental results validate that our method is more stable, and the computational efficiency is improved obviously. However, in special cases, our method cannot eliminate the influence of complex environment, and the accuracy of road extraction needs to be improved. In the further, for the elimination of the shadow or road region illumination influence, the intrinsic image extraction algorithm can be further optimized to combine different illumination region features in multiple color transformation spaces to better remove shadow and illumination effects. The distinction between the road region and the non-road region can be judged by using different features to construct voting functions of the super-pixels. Or the road region and the non-road region can be classified by a machine learning method, which learns classification model based on the fusion features of different regions. If there are enough complex labeled road images with shadows and other illumination effects, the deep learning method can be used to learn the road region extraction model. For the extraction and optimization of road boundaries, more boundary constraints can be introduced and optimization algorithms such as level-sets method can be integrated to make the extracted road region more complete and accurate.

Conflict of interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted. This research does not involve human participants or animals.

Funding

This study was funded by the National Natural Science Foundation of China (grant number 61401455).

References


Fig. 10. Failure cases for our method. (a) the original road images, (b) the ground truth, (c) the reset probability maps, (d) the gradient images, (e) our method results.


