Underwater Object Segmentation Algorithm Based on Depth Information

Xiaowen Zhu, Zheyu Hu, Dawei Tu* and Xu Zhang

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China.

* Corresponding author e-mail: tdw@shu.edu.cn

Abstract: In the process of developing the ocean, the underwater robot's accurate positioning of the object is the key to the success of the underwater mission. Due to the interference of environment and other factors, the traditional GrabCut algorithm cannot accurately locate the target from the image collected by the visual sensor. This paper proposes an image segmentation algorithm based on depth Information, which is improved on the GrabCut algorithm and fuses depth information. First, the foreground area of the original image is extracted by using the depth information of the image to become a new image to be segmented. Then, two interactive operations are conducted on the new image. Finally, GrabCut algorithm is used to obtain the segmentation result of the target. Compared with GrabCut algorithm, the algorithm in this paper is more effective, which can improve segmentation accuracy and target positioning effect.

1. Introduction

In order to develop marine resources, advanced underwater robots have been developed to replace human beings in the exploration and salvage of resources. Whether the underwater robot can accurately locate the target is the foundation of successful underwater operation, and image segmentation has become an important technology for the underwater robot to achieve accurate target positioning [1,2,3]. Among many methods of image segmentation, GrabCut algorithm proposed by Rother et al. has been widely applied in image segmentation [4], and many scholars have conducted relevant research on it [5,6,7]. In recent years, a variety of 2D/3D fusion visual sensors, such as Microsoft Kinect v2, Intel Realsense D435 and other devices, are characterized by the color 2D image, the corresponding pixel superimposed depth information, which provides great convenience for underwater object segmentation and positioning.

In view of the poor segmentation effect under complex background, this paper proposes a method combining GrabCut and depth information. The noisy background and foreground areas in the original image are separated by depth information. In the foreground area, the pixels which have a low similarity compared to the background around the object are abandoned, GrabCut algorithm is used only in the area of interest, which can effectively improve the segmentation accuracy.
2. GrabCut Algorithm
The GrabCut algorithm uses the Gaussian mixture model to model the foreground and background color of the image. After the user specifies the bounding rectangle of the object to be segmented, the GMM parameter of the foreground / background is updated, and the background in the bounding rectangle is removed by the graph cut optimization. The image pixels are represented by an array \( Z = \{ z_1, z_2, \cdots, z_n \} \), \( \alpha = \{ \alpha_1, \alpha_2, \cdots, \alpha_n \} \) (where \( \alpha_n \in [0,1] \), \( \alpha_n = 0 \) (background), \( \alpha_n = 1 \) (target)) is the array of pixels corresponding to the opacity. The vector \( K = \{ k_1, k_2, \cdots, k_n \} \) (\( k_n \in \{1,2,\cdots,K\} \), typically \( K=5 \)), is the Gaussian component of the Gaussian mixture model corresponding to the background or foreground of pixels. The foreground extraction problem can be transformed into the problem of optimal energy function.

For color images on RGB space, the Gibbs energy function used by GrabCut is:

\[
E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z)
\]  (1)

The data term \( U \) is defined as:

\[
U(\alpha, k, \theta, z) = \sum_n D(\alpha_n, k_n, \theta, z_n)
\]  (2)

where \( D(\alpha_n, k_n, \theta, z_n) = -\log P(z_n|\alpha_n, k_n, \theta) + \log \pi(\alpha_n, k_n) \) is a Gaussian probability distribution, and \( \pi(\alpha_n, k_n) \) are mixture weighting coefficients, so that:

\[
D(\alpha_n, k_n, \theta, z_n) = \log \pi(\alpha_n, k_n) + \left( \frac{1}{2} \right) \log \det \Sigma(\alpha, k) + \frac{1}{2} [z_n - \mu(\alpha, k)]^T \Sigma(\alpha, k)^{-1} [z_n - \mu(\alpha, k)]
\]  (3)

Therefore, the parameters of the model are now:

\[
\theta = \{ \pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k), \alpha = 0,1,k = 1 \cdots K \}
\]  (4)

Where \( \mu(\alpha, k) \) is the weight of the total number of samples in the Gaussian probability distribution; \( \Sigma(\alpha, k) \) is the mean value of Gaussian model. \( \Sigma(\alpha, k) \) is covariance.

The smoothness term \( V \) can be obtained by using Euclidean distance in colour space:

\[
V(\alpha, z) = \gamma \sum_{(m,n) \in C} |\alpha_n \neq \alpha_m| \exp(-\beta \|z_m - z_n\|^2)
\]  (5)

Where \( C \) is the set of pairs of neighboring pixels; \( \gamma \) is an adaptive \( \lambda \) parameter; \( \beta \) is a constant term.

If a pixel is given a mark that is closer to the real situation, the corresponding data item punishment will be smaller, so that the total energy function will be reduced, and the algorithm will keep iterating and converging to the optimal segmentation. The GrabCut image segmentation problem becomes the minimum energy segmentation problem. As shown in fig. 1, there are two pictures. One is the original picture, and other one is its segmentation result by adopting GrabCut.

![Figure 1. Original picture and its segmentation result](image)

3. An Improved GrabCut Algorithm
In an image with a complex background, even if the user selects the target area accurately, sometimes the similarity between the background inside the rectangular box and the pixel outside the rectangular box is low. Therefore, it is possible to incorrectly define the background pixel in the rectangle as the foreground pixel, or the foreground pixel in the rectangle as the background pixel. This results in a decrease in the accuracy of the segmentation results. In view of the above two phenomena, this paper proposes to improve GrabCut and use the depth information of the image to separate the background from the foreground, so as to reduce the possibility of misjudgment.
3.1. Implementation Method
The new algorithm sets the appropriate threshold according to the gray value of the depth map, and divides it into foreground and background regions. The color image corresponding to the foreground region is saved as a new image to be segmented. In the new image, two interactions are adopted to reduce the number of interfering pixels, so as to improve the segmentation speed and accuracy of GrabCut algorithm. In the Grabcut algorithm, all pixels are divided into foreground pixels and background pixels, so the existence of interfering pixels reduces the accuracy of segmentation. Two interactions are used to divide all image pixels into three categories. In addition to the foreground and background pixels, there is also a kind of abandoned pixel. The pixels outside the large rectangular box are defined as abandoned pixels, pixels between two rectangles are the initial background pixels (alpha = 0), and pixels inside rectangle two are initial foreground pixels (alpha = 1). After classifying all pixels, we should begin initialization process. We don't calculate the abandoned pixels, pixels between two rectangles will be calculated.

According to the above analysis, the improved GrabCut algorithm’s process is as follows:
1) Input the original color image and depth image, and divide them into foreground area and background area including the target according to the gray value of depth image;
2) The color image corresponding to the foreground area is saved as a new image to be segmented;
3) Drag the rectangle on the new image to select the target area. Drag another rectangle around the target area;
4) Classify all pixels to three categories: the initial background pixels, the initial foreground pixels and abandoned pixels;
5) The Gaussian components of the two GMMs are initialized according to the RGB values of the initial background pixel and the initial foreground pixel of the rectangle;
6) Calculate the probability that each pixel in the target area belongs to two GMM Gaussian components, then the pixel belongs to the component with the highest probability;
7) Study and optimize GMM parameters;
8) Partition estimation until convergence;
9) Based on the GMM parameters of the previous step, the maximum flow / minimum cut is performed in pixel units;

4. Results And Evaluation
For ease of assessment and comparison, the images in this experiment were obtained from the NJUDS2000 data set, which included nearly 2000 stereo image, depth maps and groundtruth masks. This paper will test the algorithm from two aspects of segmentation effect and accuracy, and compare it with the traditional GrabCut algorithm

4.1. Algorithm for Instance
Figure 2 is part of the renderings involved in this algorithm experiment.

![Figure 2. Specific process of experimental example](image_url)
As shown in Figure 2, (a) is the original RBG image and (b) is the corresponding depth map. Firstly, the depth map is grayscale, and the threshold value is divided into foreground mask and background mask. Then the obtained binary image was used to extract the foreground region of the original image to obtain (d). Then, on the basis of (d), select the target region and the interested region to obtain the binarization graph of the target (e); Finally, target (f) is segmented in the original image according to target binarization graph (e).

4.2. Comparison of Segmentation Effect
As shown in Figure 3, (a) to (d) correspond to the original image, depth map, GrabCut and the algorithm in this paper.

![Figure 3. Comparison of segmentation effects between the two algorithms](image)

In general, as a result of GrabCut algorithm using color segmentation and didn't add any other information, so when the background is noisy, or close to target the foreground color, background segmentation effect is lacking, and the algorithm of this paper due to the combination of depth image, use the depth information to eliminate noise in the background or foreground color similar area, excluding some interference, so relatively ideal segmentation results.

4.3. Accuracy of Segmentation
In order to further compare the effects of the two algorithms, the segmentation effects of the two methods are measured from the segmentation accuracy. The segmentation accuracy P is:

\[
P = \frac{N_B + N_F}{N} \times 100\%
\]  

Where \(N_B\) is the correct background pixel, \(N_F\) is the correct foreground pixel, and \(N\) is the total number of pixels. The following table shows the accuracy of the corresponding images of the two algorithms.
Table 1. The segmentation accuracy of two algorithms (%)

| Image  | GrabCut | Depth-based GrabCut |
|--------|---------|---------------------|
| Person 1 | 96.30   | 97.14               |
| Car 1   | 96.39   | 97.78               |
| Person 2 | 90.63   | 97.54               |
| Car 2   | 96.74   | 97.51               |
| Person 3 | 94.10   | 96.36               |

As can be seen from the above table, the accuracy of the algorithm in this paper is significantly higher than that of the traditional GrabCut algorithm. The algorithm in this paper is better than the original GrabCut algorithm whether in the scene with large difference in front background color or in the scene with noisy background or small difference in front background color.

5. Conclusion

In view of the low definition of underwater images and the difficulty of target segmentation and positioning, this paper uses depth information to separate the foreground area from the noisy background and uses two interactive methods to segment the target. The experiment shows that the algorithm in this paper can effectively improve the segmentation accuracy of GrabCut method.

Acknowledgment

This research was partially supported by National Key R&D Program of China (Grant no.2016YFC0302401) and National Natural Science Foundation of China (Grant no.61673252).

References

[1] Tongwei Ren, Yan Liu, Gangshan Wu. Image retargeting based on global energy optimization[P]. Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on,2009.
[2] Xiangyang Xu, Wenjing Geng, Ran Ju, Yang Yang, Tongwei Ren, Gangshan Wu. OBSIR: Object-based stereo image retrieval[P]. Multimedia and Expo (ICME), 2014 IEEE International Conference on, 2014.
[3] Cheng M M, Zhang G X, Mitra N J, et al. [IEEE 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Colorado Springs, CO, USA (2011.06.20-2011.06.25)] CVPR 2011 - Global contrast based salient region detection[C]// CVPR 2011. IEEE, 2011:409-416.
[4] Rother C. GrabCut : Interactive foreground extraction using iterated graph cuts[J]. Proceedings of SIGGRAPH '04, 2004, 23.
[5] Hua S, Shi P. GrabCut color image segmentation based on region of interest[C]// 2014 7th International Congress on Image and Signal Processing (CISP). IEEE, 2014.
[6] Diebold J, Demmel N, et al. Interactive Multi-label Segmentation of RGB-D Images[J]. 2015.
[7] Chen, D., Chen, B., Mamic, G., Fookes, C., Sridharan, S.. Improved GrabCut Segmentation via GMM Optimisation[P]. Computing: Techniques and Applications, 2008. DICTA '08.Digital Image,2008