Automatic object detection and segmentation from underwater images via saliency-based region merging

Yafei Zhu, Lin Chang, Jialun Dai, Haiyong Zheng*, Bing Zheng
College of Information Science and Engineering
Ocean University of China
Qingdao 266100, China
*Corresponding author: zhenghaiyong@ouc.edu.cn

***Abstract*—Underwater object detection and segmentation has been attracting a lot of interest, and recently various systems have been designed. In this paper, we introduce a novel technique to automatically detect and segment objects from underwater images via saliency-based region merging. The method is composed of three main steps. Firstly, a salient object detection model is used to detect the position of salient objects in underwater image. Secondly, background prior is applied to determine the approximate background location. Thirdly, the region merging based interactive image segmentation method is improved by adding the determined object and background location information as the user inputs so that the algorithm becomes automatic. The experimental results show that it's efficient to segment objects from the underwater image by the proposed method.**

I. Introduction

Recent years have witnessed rapidly increasing interest in underwater object detection and segmentation. It is motivated by the importance of underwater object detection and segmentation in applications such as in maintenance, repair of undersea structures, marine sciences, and homeland security.

Several underwater object detection and segmentation systems have been designed successfully. Kabatek et al. [1] did the research on underwater objects detection for electro-optical imagery data. Williams et al. presented a method for underwater target detection using context in [2]. Shen et al. [3] realized a successful system of moving object detection. An underwater acoustic image segmentation is proposed based on deformable template [4]. In [5], an efficient and fast underwater image segmentation method using thresholding with class 3 fuzzy C-means clustering and CLAHE enhancement method was proposed.

These methods have their own advantages and disadvantages, and none of them works equally well for all kinds of images. It is difficult to detect and segment objects in underwater environment without the details of the objects. Also, in the underwater object detection task, the decayed color and the haze effect would significantly decrease the contrast between the object and the background. Many traditional object detection methods are distorted and can hardly be taken for precise object detection.

After a long period of evolution, research on saliency detection has developed greatly. In salient object detection task, we only focus on the most salient objects in images. Though the underwater image is blurry and low contrast, we can still detect the position of relatively salient objects using salient object detection method. The leading method of all the benchmark salient object detection and segmentation methods in [6] is a discriminative regional feature integration (DRFI) approach [7]. Instead of purely relying on the cues extracted only from the input image, DRFI resorts to human annotations to automatically discover feature integration rules, thus is more effective in detecting underwater objects than existing models.

The background prior, which assumes that the area in the narrow border of the image belongs to the background, performs very robust in many detection tasks. So we use background prior to determine the background location information in underwater images.

Recently, semi-automatic segmentation techniques that allow solving moderate and hard segmentation tasks via user inputs are becoming more and more popular. Since object information can be obtained by salient object detection method and background information can be extracted by background prior, the semi-automatic segmentation method can then be improved to be automatic using the determined object and background location information instead of user inputs.

The framework of our proposed method is shown in Fig. 1. Firstly, the dark channel prior [8] is applied to enhance the image by removing the "haze" that makes the underwater image unclear. Then, for detecting salient objects in the enhanced image with complicated background, we choose DRFI [7] model to generate the saliency map containing "saliency values" per pixel. And the obtained saliency map is further processed to determine the location information of object and background. Finally, the maximal similarity based region merging (MSRM) method [9] is improved to be automatic by adding the determined object and background location information as the user inputs.

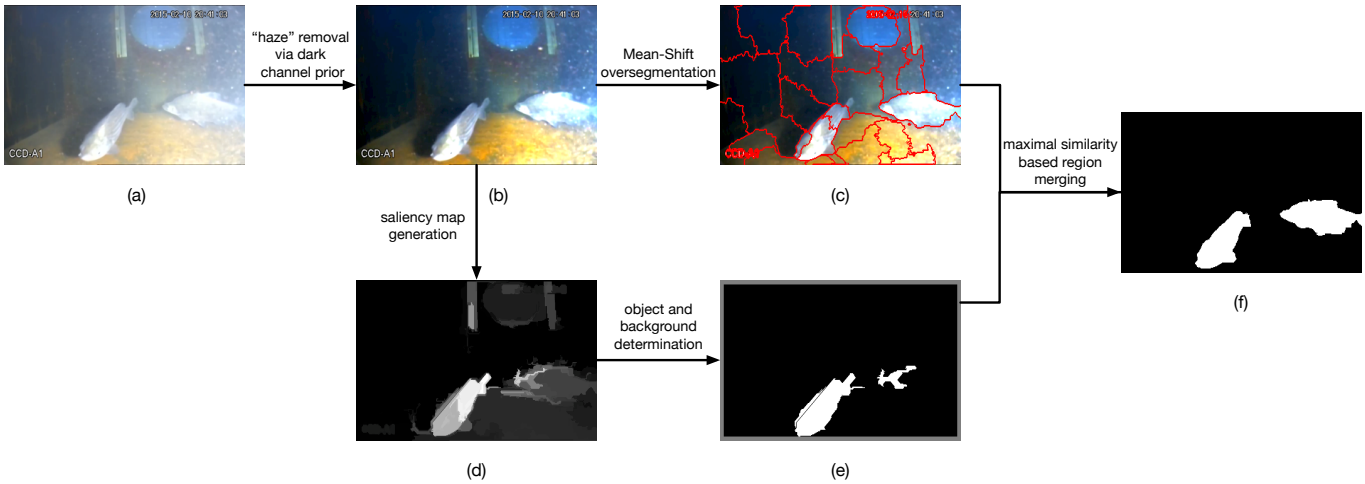


Fig. 1: The framework of our proposed method for object detection and segmentation from underwater images.

II. Method

A. Dark channel prior algorithm

The dark channel prior was proposed to remove haze from a single input image [8]. It is based on the following observation on haze-free outdoor images: in most of the non-sky patches, at least one color channel has very low intensity at some pixels. In other words, the minimum intensity in such a patch should has a very low value. Formally, for an image $\mathbf{J}$, its dark channel is defined as

$$J^{dark}(x) = \min_{c\in\{r,g,b\}} (\min_{y\in\Omega(x)} (J^c(y))) \tag{1}$$

where J^c is a color channel of $\mathbf{J}$ and $\Omega(x)$ is a local patch centered at x. Except for the sky region, the intensity of J^{dark} is low and tends to be zero, if $\mathbf{J}$ is a haze-free outdoor image. And the above statistical observation or knowledge is called the *dark channel prior*.

In computer vision and computer graphics, the model widely used to describe the formation of a haze image is as follows:

$$\mathbf{I}(x) = \mathbf{J}(x)t(x) + \mathbf{A}(1 - t(x)) \tag{2}$$

where $\mathbf{I}$ is the observed intensity, $\mathbf{J}$ is the scene radiance, $\mathbf{A}$ is the global atmospheric light, and t is the medium transmission describing the portion of the light that is not scattered and reaches the camera. The goal of haze removal is to recover $\mathbf{J}$, $\mathbf{A}$, and t from $\mathbf{I}$.

Using the haze imaging Equation (2) and the dark channel prior together, the final scene radiance $\mathbf{J}(x)$ can be derived as:

$$\mathbf{J}(x) = \frac{\mathbf{I}(x) - \mathbf{A}}{max(t(x), t_0)} + \mathbf{A} \tag{3}$$

A typical value of t_0 is 0.1.

The dark channel prior is effective for a variety of hazy images while it maybe invalid when the scene objects are inherently similar to the atmospheric light and no shadow is cast on them.

The underwater images are similar with the haze images as they are all degraded by medium. Besides, they doesn't conform the failure condition. Therefore, dark channel prior algorithm can be used to remove the haze in underwater images. Fig. 1(b) shows an example of the result processed by the dark channel prior.

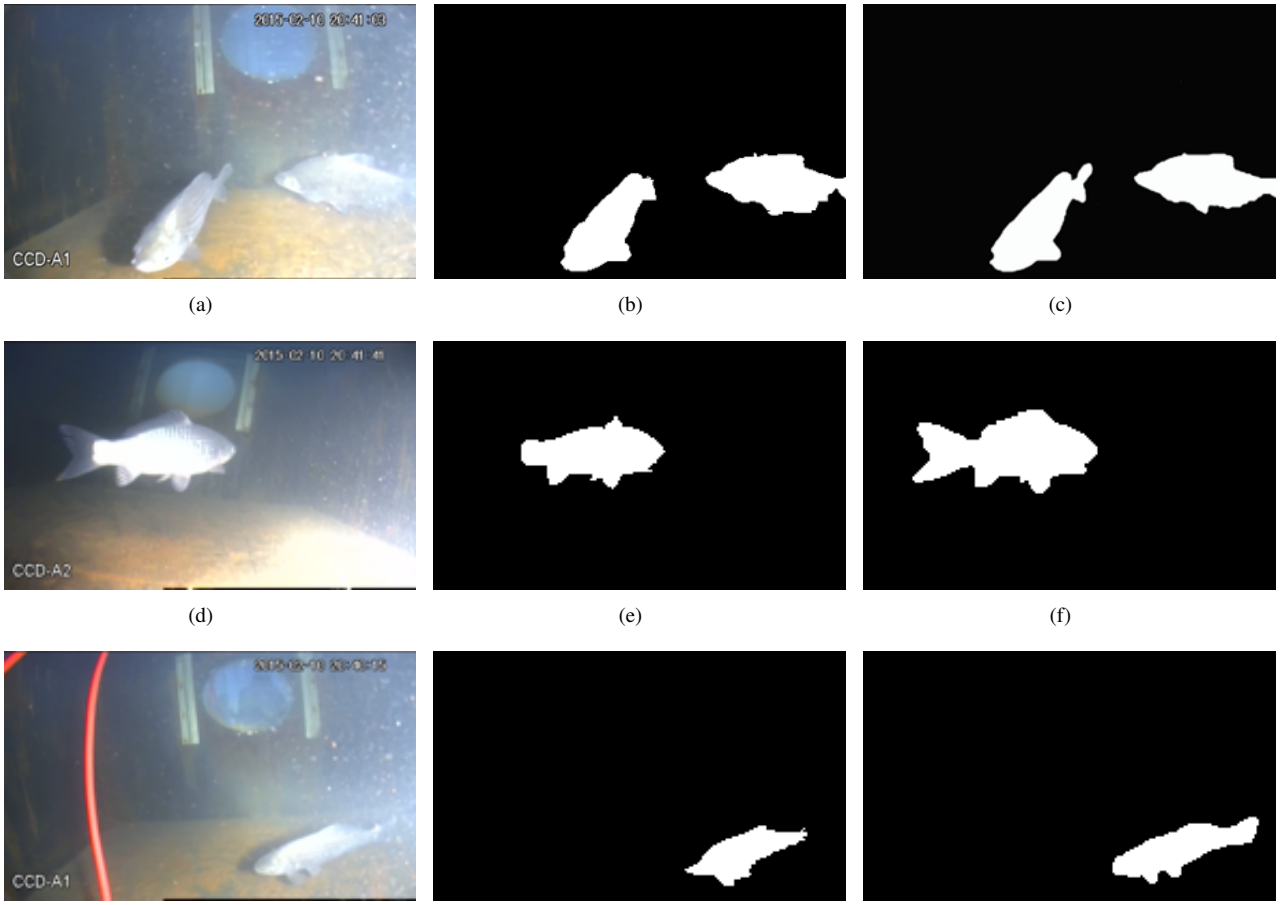
B. Discriminative Regional Feature Integration (DRFI) algorithm

Salient object detection method can be used to approximately obtain the location of the foreground, in terms of saliency map, from an image where the foreground would draw the attentions of humans at the first sight of an image. In practice, salient object detection methods are commonly used as a first step of many applications including object recognition, object segmentation, object tracking and so on. Recently, a lot of research efforts have been made for saliency computation. A comprehensive survey of salient object detection can be found from [10].

Jiang et al. [7] address the salient object detection problem using a discriminative regional feature integration (DRFI) approach. It integrates the regional contrast, regional property and regional backgroundness descriptors together to form the master saliency map, thus can deal with the challenging cases such as detecting salient objects from low-quality underwater images. Fig. 1(d) shows an example of saliency map generated by DRFI.

C. Object and background determination

For each pixel in the saliency map, the higher of the value, the more salient of it in the original image. In other words, normally, for an image, pixels which have higher values in the corresponding saliency map belong to objects. Therefore, pixels whose values in the saliency map are greater than a certain threshold can be taken as object pixels. The Otsu method [11] is adopted to compute the threshold from



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 2: Qualitative comparison: (a)(d)(g) Original image. (b)(e)(h) Our segmentation result. (c)(f)(i) Human labeled map.

the saliency map for the determination of object location information.

The background prior assumes that the image boundary is mostly background. For a 192×116 image, we define pixels in the 6-pixel wide narrow border region of the image as background pixels. To verify such a definition, we made a simple survey on 100 images and found that 98% of pixels in the border area belongs to background.

In Fig. 1(e), we use marks to indicate the object and background location information. The white parts are the object markers while the gray parts in the image boundary are the background markers, the remaining black parts are non-marked.

D. Mean-Shift oversegmentation

An initial segmentation is required to partition the image into homogeneous regions for merging. Any existing low level segmentation methods, such as mean shift [12], [13], watershed [14], SLIC [15] and Turbopixel [16], can be used for this step. In this paper, we choose to use the mean shift method for initial segmentation because it has less over segmentation and can well preserve the object boundaries. Fig. 1(c) shows an example of the mean shift initial segmentation.

E. Object and background marking

For each region in the oversegmented result (Fig. 1(c)), we regard it as background region (denoted by M_B) if there is at least one background pixel in it, then, a region is regarded as object region (denoted by M_O) if the number of object pixels in it accounts for more than 10% of the region area. If there is neither object nor background pixel in one region, we call it non-marker region, denoted by N. Now the whole image is divided into three parts: M_O, M_B, and N.

F. Maximal-similarity based region merging

After mean shift initial segmentation, we have many small regions available. To guide the following region merging process, some descriptor for representing these regions and a rule for merging are needed.

1) Region representation: The initially segmented small regions of the desired object often vary a lot in size and shape, while the colors of different regions from the same object will have high similarity. Therefore, the color histogram is used to represent each region in this paper. We uniformly quantize each color channel into 16 levels and then the histogram of each region is calculated in the feature space of

$16 \times 16 \times 16 = 4096$ bins. Denote by $Hist_R$ the normalized histogram of a region R.

2) Similarity measure: A similarity measure $\rho(R, Q)$ between two region R and Q is defined to accommodate the comparison between various regions.

$$\rho(R, Q) = \sum_{u=1}^{4096} \sqrt{Hist_R^u \cdot Hist_Q^u} \tag{4}$$

3) Maximal similarity based merging rule: Let Q be an adjacent region of R and denote by $\bar{S}_Q = \{S_i^Q\}_{i=1,2,\ldots,q}$ the set of Q's adjacent regions. The similarity between Q and all its adjacent regions, i.e. $\rho(Q, S_i^Q), i = 1, 2, \ldots, q$, are calculated. Obviously, R is a member of $\bar{S}_Q$. If the similarity between R and Q is the maximal one among all the similarities $\rho(Q, S_i^Q)$, we will merge R and Q. The following merging rule is defined:

$$\text{Merge } R \text{ and } Q \quad \text{if } \rho(R, Q) = \max_{i=1,2,\ldots,q} \rho(Q, S_i^Q) \tag{5}$$

4) The merging process: The whole region merging process can be divided into two stages. In the first stage, merging non-marker regions in N with marker background regions in M_B. In the second stage, merging non-marker regions in N adaptively. Fig. 1(f) shows an example of the final segmentation result after region merging.

III. EXPERIMENTAL RESULTS

Fig. 2 shows three examples of the results: (a), (d) and (g) are original images captured from underwater videos, (b), (e) and (h) are the final segmentation results, while (c), (f) and (i) are ground truth labeled by human.

As can be seen, (a) contains two objects, while (d) and (g) each contains one single object. The background in these images are all complex and have low contrast with the objects. From the segmentation results, we can see that our method can deal with the challenging cases where the background is cluttered and segment the right number of objects in all these images. Also, most meaningful object information can be extracted. However, it fails at some details such as the fish tail part in all these images, since the fish tail shares more commons with the underwater background both in color and luminance.

IV. CONCLUSION

In this paper, we address the underwater object detection and segmentation problem using a saliency-based region merging approach. The success of our approach stems from three key factors. One is that we introduce salient object detection method to determine the spatial location of salient objects in blurry and miscolored underwater image. The second one is that we apply background prior to determine the background location information. The last one is that we integrate semi-automatic region merging framework and the object and background location information into a new automatic framework. The experimental results prove that the presented method is valid.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61271406, 61301240, and International Science & Technology Cooperation Program of China under Grant No. 2012DFG22080.

REFERENCES

[1] M. Kabatek, M. R. Azimi-Sadjadi, and J. D. Tucker, "An underwater target detection system for electro-optical imagery data," *OCEANS MTS/IEEE Biloxi*, 2009.

[2] D. P. Williams, G. L. Davies, and A. J. Evans, "Using context to detect underwater objects," *Sensor Signal Processing for Defence*, 2010.

[3] J. Shen, T. Fan, M. Tang, Q. Zhang, Z. Sun, and F. Huang, "A biological hierarchical model based underwater moving object detection," *Computational and Mathematical Methods in Medicine*, vol. 2014, 2014.

[4] Z. Liu, E. Sang, and Z. Liao, "Underwater acoustic image segmentation based on deformable template," *IEEE International Conference on Mechatronics and Automation*, vol. 4, pp. 1802–1806, 2005.

[5] S. Singh, M. Soni, and R. S. Mishra, "Segmentation of underwater objects using clahe enhancement and thresholding with 3-class fuzzy c-means clustering," *International Journal of Computer Applications*, vol. 4, 2014.

[6] B. Ali, C. Ming-Ming, J. Huaizu, and L. Jia, "Salient object detection: A benchmark," *arXiv preprint arXiv:1501.02741*, 2015.

[7] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2083–2090, 2013.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[9] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognition*, vol. 43, no. 2, pp. 445–456, 2010.

[10] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey," *arXiv preprint arXiv:1411.5878*, 2014.

[11] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[12] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[13] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[14] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 583–598, 1991.

[15] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[16] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.