Saliency Cut in Stereo Images

Jianteng Peng, Jianbing Shen*, Yunde Jia
Beijing Laboratory of Intelligent Information Technology, School of Computer Science
Beijing Institute of Technology, Beijing 100081, P. R. China

Xuelong Li
Center for OPTical IMagery Analysis and Learning (OPTIMAL)
State Key Laboratory of Transient Optics and Photonics
Xi'an Institute of Optics and Precision Mechanics
Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China

Abstract

In this paper, we propose a novel saliency-aware stereo images segmentation approach using the high-order energy items, which utilizes the disparity map and statistical information of stereo images to enrich the high-order potentials. To the best of our knowledge, our approach is first one to formulate the automatic stereo cut as the high-order energy optimization problems, which simultaneously segments the foreground objects in left and right images using the proposed high-order energy function. The relationships of stereo correspondence by disparity maps are further employed to enhance the connections between the left and right images. Experimental results demonstrate that the proposed approach can effectively improve the saliency-aware segmentation performance of stereo images.

1. Introduction

With the recent increase in 3D contents such as 3D movies, which has ignited the rapid development of 3D displays and depth cameras. This has arisen the necessity in 3D image segmentation and creation approaches. However, it is challenging to directly apply 2D image segmentation approaches [16, 11] to stereoscopic images, because the additional information such as disparity maps in stereoscopic images introduces additional constraints for improving the performance of stereo segmentation. Since the existing 2D image segmentation methods do not take these constraints into account, simple extensions of existing methods usually fail to produce good results for stereo images.

The representative and popular image segmentation methods are the graph-based approaches which treat the segmentation problem as a minimum cut or maximum flow problem through a graph partitioning structure, such as the normalized cuts methods [16, 7] and graph cuts framework [11, 3, 1]. However, these approaches can not meet the requirements of automatically segmenting the salient objects as a binary labeling optimization problem from stereo images. The most related method to our work is the interactive stereo object selection approach using lower-order graph cut method in [13]. But our approach is different to their method in two aspects: one is our automatic saliency-aware foreground objects segmentation, the other is the recent high-order energy optimization [10, 8, 6] is developed to improve the segmentation performance.

The topic of salient foreground objects detection and segmentation is very active in recent years, which has been used in numerous computer vision application scenarios such as object detection [18]. Although many saliency-aware segmentation approaches for 2D images have been developed, including saliency-driven total variation segmentation [5], conditional random field model [14] and saliency filters [12], as well as saliency detection and segmentation via low rank matrix [15], the solution of saliency-aware stereo cut is still seldom mentioned and discussed. Furthermore, the existing single-image saliency segmentation approaches still suffer from some limitations to be applied to stereo segmentation. The most limiting one is that saliency segmentation results are obtained by thresholding the saliency maps with an adaptive threshold [12] or by employing the graph cut as a post-processing step [14, 4]. The segmentation performance of these approaches mostly de-

*Corresponding author: Jianbing Shen (shenjianbing@bit.edu.cn). This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805, the Key Program of NSFC-Guangdong Union Foundation under Grant U1035004, the National Natural Science Foundation of China under Grant 61272359 and Grant 61125106, the Program for New Century Excellent Talents in University (NCET-11-0789), and the Shaanxi Key Innovation Team of Science and Technology (2012KCT-04).

pends on the performance of image saliency techniques. In order to address the above problems, we propose a novel saliency-aware stereo cut algorithm using our high-order energy optimization for stereo image pairs.

Our method considers the stereo segmentation problems as the labels distribution problem and utilizes the high-order energy minimization method to solve it. The task of saliency-aware stereo images cut is to build one st-graph to simultaneously segment the left and right images. The original low-order graph-cut method only uses the data term and smooth term to measure the labels of every pixel. In contrast, we add the high-order term to construct the high-order energy function to optimize the segmentation results. Our high-order term includes more prior information of the stereo images, such as the set of corresponding pixels, which will improve the performance of stereo segmentation, especially when the salient object has the similar color and texture with the background. Our high-order energy optimization approach has solved the problem of saliency-aware stereo segmentation where the foreground objects are simultaneously segmented out from both the left and right images. The segmentation result is obtained by minimizing our high-order energy function, which is based on the corresponding pixels and their neighboring pixels from the disparity maps and saliency maps.

2. Our approach

2.1. The generalized graph-cut optimization

The original graph-cut optimization method [3, 11] adopts two energy terms, which includes the data term E_{data} and the smooth term E_{smooth} to construct the energy function for the image segmentation problem. As defined before, we represent the left and right images from the stereo image pair as I_l and I_r, the pixels of objects which we want to segment out should be labeled 1, and the rest pixels are set with the label 0. The original graph-cut based optimization method for single image segmentation is naturally extended for the task of stereo images segmentation by adding the extra high-order energy item E_{high}, which is based on the following energy function:

$$E = E_{data} + \lambda_s E_{smooth} + \lambda_c E_{high} \tag{1}$$

The specific form of E_{data} will be discussed in Section 2.3, and the classic form of E_{smooth} is defined as follows:

$$E_{smooth} = \sum_{(p,q)\in\mathcal{N}} exp\{-\frac{||\mathbf{f}_p - \mathbf{f}_q||^2}{2\sigma^2}\} V(L_p, L_q)$$

$$V(L_p, L_q) = \begin{cases} 0, & \text{if } L_p = L_q \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

where $\mathcal{N}$ is the set of all the neighborhood pixels, $\mathbf{f}_p$ is the feature vector of pixel p and σ is the variance of all the neighborhood pixels.

2.2. High-order energy term

As mentioned before, our high-order energy term is employed to improve the stereo segmentation performance. High-order term usually includes more variables than the low-order one, therefore, more statistics information is contained and exploited in the definition of high-order energy term. In our saliency-aware stereo image segmentation method, the proposed high-order energy term will involve more prior information of stereo images, such as the corresponding pixels and the disparity map. Our high-order term considers both the corresponding pixels and their neighboring pixels in stereo image pair. The proposed high-order term encourages the variables of both the corresponding pixels and their neighboring pixels to get the same value.

As a result, neighborhoods of the corresponding pixels are added in our high-order energy term and they are encouraged to take the same labels as well. Our high-order energy term is defined according to all the corresponding pixels and their neighboring pixels. The 4-connected neighbors are taken into consideration in our experiments. Let $N_1(p)$, $N_2(p)$, $N_3(p)$, $N_4(p)$ represent the up, down, left, right neighbors of a certain pixel p, respectively. Then our high-order energy item is defined as:

$$E_{high} = \sum_{(p,q)\in\mathcal{C}} \psi(p,q) \cdot \sum_{i=1}^{4} F[p, q, N_i(p), N_i(q)] \tag{3}$$

$$\psi(p,q) = exp\{-\frac{\|\bar{\mathbf{f}}_p - \bar{\mathbf{f}}_q\|^2}{2\sigma^2}\} \tag{4}$$

$$F[p, q, N_i(p), N_i(q)] = |L_p - L_q| + |L_{N_i(p)} - L_{N_i(q)}| \tag{5}$$

where p and q are the corresponding pixels from the left and right images, respectively, and $\mathcal{C}$ is the pair set of the correspondent pixels. $\psi(p,q)$ measures the similarity of p and q, and $\bar{\mathbf{f}}_p$ is the average features of square patch centered in pixel p. $F[.]$ is a penalty function, which measures the difference of labels of p and q, and their neighboring pixels. Since all the variables L_p, L_q, $L_{N(p)}$ and $L_{N(q)}$ in $F[.]$ are the binary labels, the value of $F[.]$ can only be 0, 1, or 2. For example, when $L_p = 1$, $L_q = 0$, $L_{N(p)} = 1$, $L_{N(q)} = 1$, then $F[p, q, N(p), N(q)] = 1$.

One good property of $F[.]$ is that it can be transformed into low-order energy term easily. However, the energy term may not perform well when we fixed the coefficients of penalty function $F[.]$. Then $F[.]$ should be adaptive to the image contents. There are two pixel-pairs in the high-order clique, which includes the pair of corresponding pixels (p, q) and the pair of their neighboring pixels $(N(p), N(q))$. We define three situations with different values of f_0, f_1, and f_2, which represent no pixel-pair, only one pixel-pair, and two pixel-pairs with different labels, respectively. If these three variables f_0, f_1, and f_2 can be

determined adaptively according to different contents of local regions in image, which will result in a better segmentation performance. For example, the smooth image region such as the backgrounds of sky and grass land belongs to the aforementioned third situation of two pixel-pairs with different labels. The penalty should be set with large values at these regions so as to make the two pixel-pairs have the same labels. However, we should use small penalty values in the edge regions to make pixel-pairs have the same labels. Since data term and smooth term will play the predominant role for achieving the segmentation performance in these boundary regions, which is further analyzed and illustrated in Fig.1.

According to the aforementioned analysis, a new and more flexible form of high-order penalty $F[.]$ is then designed as follows:

$$F[p, q, N_i(p), N_i(q)] = \mathcal{F}(u) \tag{6}$$

$$u = |L_p - L_q| + |L_{N_i(p)} - L_{N_i(q)}| \tag{7}$$

$$\mathcal{F}(u) = \begin{cases} f_0, & \text{if } u = 0 \\ f_1, & \text{if } u = 1 \\ f_2, & \text{if } u = 2 \end{cases} \tag{8}$$

where f_0, f_1 and f_2 are the three parameters that control the penalty values, which represent no pixel-pair, only one pixel-pair, and two pixel-pairs with different labels, respectively. Therefore, the values of f_0, f_1 and f_2 satisfy the relationship $f_0 < f_1 < f_2$. There is no penalty when $f_0 = 0$ and the labels of the corresponding pixels and their neighborhoods are exactly the same. When $f_2 = 2f_1$, Equ(6) is equal to Equ(5) by multiplying a factor, which can be solved by low-order graph-cut technique. Since the value of $F[.]$ is in the range of $[0, 2]$, we then set $f_2 = 2$ in our experiments to make the value of $F[.]$ to be in the same range. f_1 will be changed adaptively according to the image contents of the corresponding pixels in stereo images pair.

It is difficult to directly reduce the order of Equ(6) because $\mathcal{F}(.)$ is not linear. Fortunately, we can use the minimum selection reduction method [11, 8, 6] to solve this problem well. We begin by defining the labels of corresponding pixel-pair in left and right images as x_0 and y_0, and the labels of their related neighbors as x_1 and y_1, where $x_0, y_0, x_1, y_1 \in \{0, 1\}$. Then Equ(6) is rewritten as:

$$F[x_0, y_0, x_1, y_1] = \mathcal{F}(|x_0 - y_0| + |x_1 - y_1|) \tag{9}$$

By introducing the new auxiliary binary variables w_i ($i = \{0, 1, 2, 3, 4\}$), Equ(9) can be reduced.

Lemma. In the energy minimization condition, $\mathcal{F}(|x_0 - y_0| + |x_1 - y_1|)$ can be reduced by following equations:
Define $a = f_2 - 2f_1$

if $a > 0$

$$\begin{aligned} &\mathcal{F}(|x_0 - y_0| + |x_1 - y_1|) = \\ &\quad (f_1 + 6a)(x_0 + x_1 + y_0 + y_1) + \\ &\quad (-2)a(w_1 + w_2 + w_3 + w_4) + (-12a)w_0 + \\ &\quad (-3a)(x_0x_1 + x_0y_1 + x_1y_0 + y_0y_1) + \\ &\quad (-4a - 2f_1)(x_0y_0 + x_1y_1) + \\ &\quad (-2a)[w_1(x_0 + x_1 + y_0) + w_2(x_0 + x_1 + y_1) + \\ &\quad w_3(x_0 + y_0 + y_1) + w_4(x_0 + x_1 + y_0 + y_1)] + \\ &\quad 4aw_0(x_0 + x_1 + y_0 + y_1) \end{aligned} \tag{10}$$

if $a < 0$

$$\begin{aligned} &\mathcal{F}(|x_0 - y_0| + |x_1 - y_1|) = \\ &\quad f_1(x_0 + x_1 + y_0 + y_1) + \\ &\quad 4a(w_1 + w_2 + w_3 + w_4) + 12aw_0 + \\ &\quad 5a(x_0x_1 + x_0y_1 + x_1y_0 + y_0y_1) + \\ &\quad (4a - 2f_1)(x_0y_0 + x_1y_1) + \\ &\quad (-2a)[w_1(x_0 + x_1 + y_0) + w_2(x_0 + x_1 + y_1) + \\ &\quad w_3(x_0 + y_0 + y_1) + w_4(x_0 + x_1 + y_0 + y_1)] + \\ &\quad (-8a)w_0(x_0 + x_1 + y_0 + y_1) \end{aligned} \tag{11}$$

It is clear that when $a = 0$, $\mathcal{F}(.)$ becomes linear and can be simplified to $f_1(x_0 + x_1 + y_0 + y_1) - 2f_1(x_0y_0 + x_1y_1)$. If $a \neq 0$, the original high-order term can also be reduced by this **Lemma**. Then a st-graph is built and the final solution is computed by the min-cut/max-flow algorithm. For each corresponding pixel-pair p and q, there is four high-order energy terms. We use the method in [19] to get the disparity map, and then find the set of corresponding pixels $\mathcal{C}$ in Equ(3).

Instead of fixing the three parameters ($f_0 = 0$ $f_1 = 1$ and $f_2 = 2$), we set f_1 free and let the local region to decide its value in our high-order term. We adopted such strategy for obtaining the adaptive penalty function with the following reason. As shown in Fig.1, p_1 and q_1 are in the sky region with low gradients, but p_2 and q_2 are on the boundaries of sun umbrella. From the labels of these pixels, we can see that $F[p_1, q_1, N(p_1), N(q_1)] = f_1$ and $F[p_2, q_2, N(p_2), N(q_2)] = f_1$. According to the label distribution of $p_1, q_1, N(p_1), N(q_1)$, and f_1 in this situation should be given a large value. However, the labels of $p_2, q_2, N(p_2)$, and $N(q_2)$ are the same with the ones of ground truth. As a result, f_1 should have a small value to reduce the role of high-order penalty. Then, the values of f_1 for every high-order clique should be adaptive to their local patches. Generally, the variance of pixels in the 3×3 local patches can determine whether the patch is smooth or contains edges, then f_1 is computed by the variance var of local patches in our experiments. The higher var is, the more possible one edge is there, the smaller f_1 is. The smaller var is, the smoother the patch is, and the larger f_1

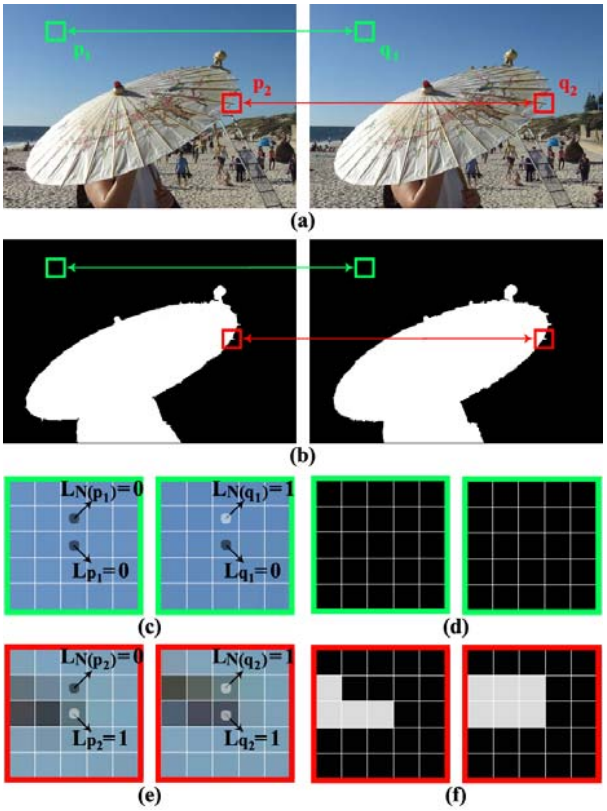

Figure 1. An example explaining the reason that f_1 adaptively takes different values at different patches. (a) Input stereoscopic images with two pairs(p_1,q_1) (p_2,q_2) of corresponding pixels; (b) the ground truth segmentation of foreground objects; (c) and (e) are the zoomed local patches in (a); (d) and (f) are the zoomed ones in (b).

should be. In summary, the values of f_1 should still be in the range $[0, 2]$ where f_1 is adaptively computed for all the high-order cliques.

2.3. Automatic saliency-aware stereo cut

As shown in Fig.2, our automatic saliency-aware stereo cut by our high-order energy term has the ability to segment the foreground objects from the input stereo image pairs. Based on the pre-computed saliency maps (Fig.2, middle row), our approach simultaneously segments out the same foreground objects in both the left and right images.

The most important part in our data term is the likelihood measurement of one pixel in the image pair belonging to the foreground objects. We use the saliency feature to measure such likelihood of the data term in our energy function. The saliency of both images in the stereo pair (S^l and S^r) can be pre-computed. In our experiments, we have used the contrast-based saliency measurement method [4] to produce the saliency estimations for computational efficiency. Since there are numerous saliency estimation approaches developed by the researchers in the society of computer

Figure 2. An example of foreground objects segmentation in stereo image pair. Top: input stereo image pair; middle: the computed saliency maps; bottom: the segmented foreground objects (signpost).

and human vision, our approach is not limited one particular saliency measure approach and other methods can also be employed [9]. We have the two following observations about the relationships between the image saliency and the foreground objects, which is inspired by the perceptual research [2]. The first is that the more possible one pixel belongs to the foreground, the larger saliency value it gets. The second is that the more similar the color of one pixel is with the color of the foreground objects, the more possible it belongs to the foreground region. According to these two observations, we define the data item as follows:

$$\begin{aligned} E_{data} &= w_s E_s + w_c E_c \\ &= w_s\Big(\sum_{p\in\mathcal{P}^l} U^s(L_p, S_p^l) + \sum_{p\in\mathcal{P}^r} U^s(L_p, S_p^r)\Big) \\ &+ w_c \sum_{p\in\mathcal{P}^l \bigcup \mathcal{P}^r} U_c(L_p, c_p) \end{aligned} \tag{12}$$

where $\mathcal{P}^l$ and $\mathcal{P}^r$ are the pixel sets from the left image and right one. s_p and c_p are the saliency and color features of pixel p. This data term contains two unary terms U^s and U^c where w_s and w_c are their weights. U^s encourages the pixel having the high saliency value to get the label 1. According to the prior of foreground objects having large saliency values, then the specific form of U^s is defined as:

$$U^s(L_p, S_p) = \delta(L_p, 1)(1 - f(s_p)) + \delta(L_p, 0) f(s_p) \tag{13}$$

Figure 3. Visual comparison of our approach to previous single-image saliency segmentation approaches. The first row is the input stereo pairs. The second, the third and the fourth rows are the segmentation results by [14], [4] and our method respectively; The last row is the ground truth masks.

$$\delta(p,q) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{otherwise} \end{cases}$$

where $\delta(.,.)$ is the delta function and $f(.)$ is a non-decreasing linear function as $f(s_p) = s_p$.

Another unary term U^c encourages that one pixel having the similar features with the foreground objects will get label 1. Therefore, the color distributions of foreground and background pixels using the saliency values should be computed first, which constructs the histograms of foreground and background objects as the weights. As mentioned before, the foreground objects tend to own a high saliency value, then the form of U^c is defined as follows:

$$U^c(L_p, c_p) = -\delta(L_p, 1) \cdot log h_1^c(c_p) - \delta(L_p, 0) \cdot log h_0^c(c_p) \quad (14)$$

where h_1^c and h_0^c are the weighted histograms of foreground and background pixels.

3. Experimental results

In order to evaluate our saliency-aware stereo segmentation method, we build a new data set of 100 stereo pairs, which contains stereo images with a variety of scene contents and their ground truth. The stereo image pairs in our data set are all downloaded from Flickr. Three users are asked to indicate the most salient regions with a rectangle in both left and right images, then we use the most consistent 100 images with the above labeling rectangles by three users. The final ground truth of salient objects are manually segmented by a user using Photoshop, which takes about five minutes on average to segment a pair of stereo images. Our source code and supplementary materials will be publicly available online at *http://cs.bit.edu.cn/shenjianbing/hoe.html* .

We first compare our saliency cut method with the recent state-of-the-art methods of single-image saliency segmentation in [14] and [4] so as to demonstrate the advantage of the proposed method. As shown in Fig. 3, our approach achieves the best segmentation performance than the other single-image saliency segmentation methods [14, 4]. The segmented regions by our approach (Fig. 3, fourth row) have the most accurate foreground region with the ground truth masks (Fig. 3, last row). Though the other two algorithms can segment out the most part of the foreground objects from stereo images, however, the background pixels are more or less split into the salient foreground objects incorrectly, such as the guitar object (Fig. 3, second row) and child in the toy cars (Fig. 3, third row). Our approach is essentially different from the other approaches for single-image saliency segmentation. Our method simultaneously segments the left and right images by considering the disparity maps and statistical information of stereo images in our high-order energy optimization, while the existing single-image saliency segmentation approaches just

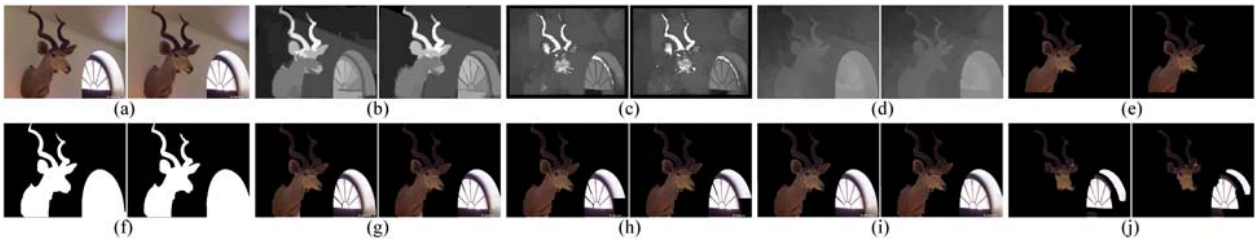

Figure 4. Illustration of our algorithm with different saliency maps: (a) the input image pair; (b), (c), and (d) the saliency maps from the methods in [4], [14] and [9], respectively; (f) the ground truth masks; (g), (h), and (i) the segmentation results by our approach via different saliency maps in (b), (c) and (d); (e), (j) the single-image saliency segmentation results by [4] and [14].

segment the left and right images separately and does not utilize the corresponding depth information between left and right images. Therefore, our method successfully cut out the correct foreground objects with the high saliency values.

Our algorithm is insensitive to saliency maps, which is illustrated in Fig. 4. We perform the comparison experiments by using the same high-order energy function but using different saliency detection approaches [14, 4, 9]. Our algorithm can still obtain the same good segmentation results (Fig. 4(g)-(i)) with different saliency maps (Fig. 4(b)-(d)). Our segmentation performance is also better than the result by these single-image cut approaches (Fig. 4(e),(j)), since these approaches are essentially based on the conventional graph-cut algorithm. In contrast, our approach designs the high-order energy item, which helps to optimize the segmentation results. Fig. 5 gives the comparison results to demonstrate the advantage of our high-order energy. The results in Fig. 5(c) and (d) have the same data term and smooth term. Our full approach using the high-order item achieves the better segmentation result (Fig. 5(d)) than the one without it (Fig. 5(c)). Our high-order energy encourages the corresponding pixels by the disparity maps to have the same labels after optimization. Therefore, the segmentation accuracy will be further improved by finding the corresponding pixels in the left and right stereo images according to our high-order energy.

Finally, we quantitatively evaluate the segmentation results using our new stereo data set. We adopt two quality measures: error rate (ER) and boundary recall (BR), to evaluate the performance of segmentation accuracy. ER measures the percentage of pixels from the results that have different labels as the ground truth. The smaller ER is, the better segmentation performance is. ER is defined as follows:

$$ER = \frac{|D|}{|L^l| + |L^r|}; D = \{p|L_p^l \neq GT_p^l\} \cup \{p|L_p^r \neq GT_p^r\} \quad (15)$$

where L and GT are the label distributions of segmentation result and ground truth where the superscripts l and r denote the left and right one of them. D is the set of pixels which has the different labels in L and GT.

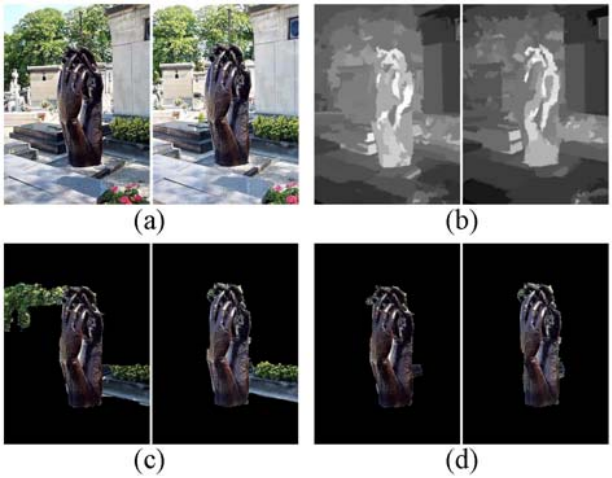

Figure 5. Comparison results with and without our high-order energy item. (a) Input stereo pair; (b) saliency maps; (c) segmentation result by our approach without high-order term; (d) segmentation result via our full approach with high-order term.

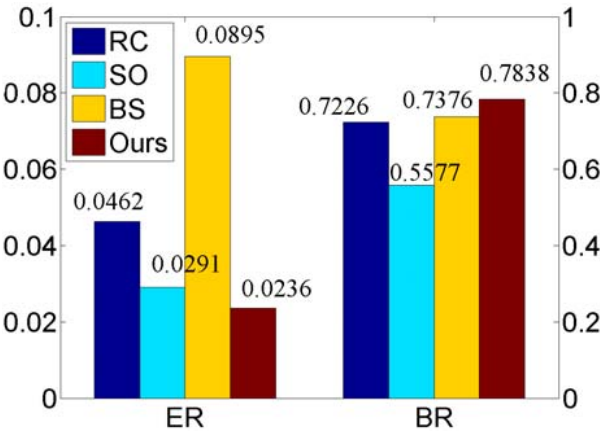

Figure 6. ER and BR evaluation of segmentation results between our method and the single-image saliency-aware segmentation approaches in RC [4], SO [14], and BS [17].

The second measure is the boundary recall, which measures the percentage of boundary from ground truth that is close to the edges from segmentation result. The better seg-

mentation edges are aligned to ground truth, the larger BR will be. We then calculate BR as follows:

$$BR = \frac{\sum_{p\in\delta(GT)} \phi(min_{q\in\delta(L)}||p-q|| < \epsilon)}{|\delta(GT)|} \quad (16)$$

where δ is an operator that gets boundary from label distributions, and $\phi(.)$ is a critical function. If the logical form is true, then the value of ϕ is 1, otherwise it is 0. We have set $\epsilon = 2$ in our experiments.

The ER and BR statistics between our saliency-aware method and previous single-image saliency segmentation approaches with 100 stereo images are shown as the histograms in Fig.6. Our method outperforms the other three single-image methods for segmenting the stereo images. The segmentation method in [4] is a refinement process via grabcut on their saliency map. A trimap is updated by dilating and eroding the current segmentation result at each iteration. This process will lead to details loss of salient object regions, which will result in a large value of ER. Both the segmentation results by [14] and [17] rely on the performance of saliency maps, and the boundary of salient objects will become not accurate when their methods can not produce the good saliency maps for the complicated scenes. Therefore, their results will have a high value of ER. Our high-order energy item is developed by the disparity maps and statistical information of corresponding pixels in the stereo images, which efficiently improves the performance of saliency-aware stereo segmentation.

4. Conclusion

This paper presents a novel saliency-aware stereo segmentation approach using the high-order energy terms, which simultaneously cuts out the foreground objects from stereo image pairs. Our method takes full advantage of the disparity map and statistical information of stereo images to enrich the high-order potentials. With the guidance of saliency maps, not only data term and smooth term but also the high-order term is added in our high-order energy function. This energy term contains more prior information of stereo image pairs, which helps to improve the segmentation performance by encouraging the corresponding pixels to obtain the same labels in different stereo images. Our high-order energy term contains more variables and more statistics information of stereo image pairs where the penalty function is adaptively determined by the color and texture information of local patches. The experimental results have shown that our saliency-aware cut approach achieves the high quality segmentation results for stereo image pairs.

References

[1] Y. Boykov and G. Funka-Lea. Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[2] N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search: an information theoretic approach. *Journal of Vision*, 9(3):1–24, 2009.

[3] A. B. C. Rother and V. Kolmogorov. Grabcut-interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23.

[4] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu. Global contrast based salient region detection. *In Proceedings of IEEE CVPR*, pages 409–416, 2011.

[5] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof. Saliency driven total variation segmentation. *In Proceedings of IEEE ICCV*, pages 817–824, 2009.

[6] A. C. Gallagher, D. Batra, and D. Parikh. Inference for order reduction in markov random fields. *In Proceedings of IEEE CVPR*, pages 1857–1864, 2011.

[7] B. Ghanem and N. Ahuja. Dinkelbach ncut: an efficient framework for solving normalized cuts problems with priors and convex constraints. *International Journal of Computer Vision*, 89(1):40–55, 2010.

[8] I. Hiroshi. Higher-order clique reduction in binary graph cut. *In Proceedings of IEEE CVPR*, pages 2993–3000, 2009.

[9] H. C. K. Chang, T. Liu and S. Lai. Fusing generic objectness and visual saliency for salient object detection. *In Proceedings of IEEE ICCV*, pages 914–921, 2011.

[10] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009.

[11] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[12] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: contrast based filtering for salient region detection. *In Proceedings of IEEE CVPR*, pages 733–740, 2012.

[13] B. L. Price and S. Cohen. Stereocut: consistent interactive object selection in stereo image pairs. *In Proceedings of IEEE ICCV*, pages 1148–1155, 2011.

[14] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä. Segmenting salient objects from images and videos. *In Proceedings of ECCV*, pages 366–379, 2010.

[15] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. *In Proceedings of IEEE CVPR*, pages 853–860, 2012.

[16] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[17] Y. Xie, H. Lu, and M. H. Yang. Bayesian saliency via low and mid level cues. *IEEE Trans. on Image Processing*, 22(5):1689–1698, 2013.

[18] L. Yang, N. Zheng, J. Yang, M. Chen, and H. Cheng. A biased sampling strategy for object categorization. *In Proceedings of IEEE ICCV*, pages 1141–1148, 2009.

[19] Q. Yang. A non-local cost aggregation method for stereo matching. *In Proceedings of IEEE CVPR*, pages 1402–1409, 2012.