StereoSnakes: Contour Based Consistent Object Extraction For Stereo Images

Ran Ju Tongwei Ren Gangshan Wu
State Key Laboratory for Novel Software Technology
Nanjing University, China
juran@smail.nju.edu.cn, {rentw,gswu}@nju.edu.cn

Abstract

Consistent object extraction plays an essential role for stereo image editing with the population of stereoscopic 3D media. Most previous methods perform segmentation on entire images for both views using dense stereo correspondence constraints. We find that for such kind of methods the computation is highly redundant since the two views are near-duplicate. Besides, the consistency may be violated due to the imperfectness of current stereo matching algorithms. In this paper, we propose a contour based method which searches for consistent object contours instead of regions. It integrates both stereo correspondence and object boundary constraints into an energy minimization framework. The proposed method has several advantages compared to previous works. First, the searching space is restricted in object boundaries thus the efficiency significantly improved. Second, the discriminative power of object contours results in a more consistent segmentation. Furthermore, the proposed method can effortlessly extend existing single-image segmentation methods to work in stereo scenarios. The experiment on the Adobe benchmark shows superior extraction accuracy and significant improvement of efficiency of our method to state-of-the-art. We also demonstrate in a few applications how our method can be used as a basic tool for stereo image editing.

1. Introduction

The enthusiasm on stereoscopic media has been lit up by the population of 3D movies and TV programs in recent years. With the increasing amount of stereoscopic data, tools to handle such a kind of media turns to be an urgent demand to support television broadcasting, film industry and even daily life photo editing. Among the tasks for stereo media handling, one of the most essential tools is to consistently extract objects from stereo image pairs, as shown in Fig. 1.

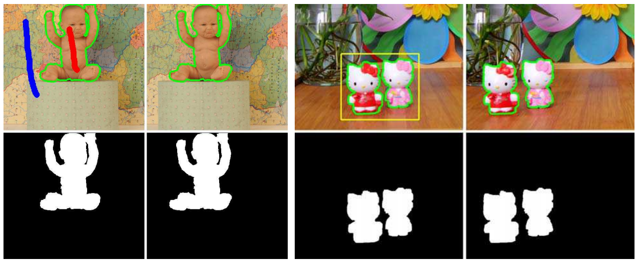


Figure 1. Consistent object extraction for stereo images. First row: the interactions on stereo images and the resultant object contours. Second row: object masks generated from contours. Our method can effortlessly make single-image segmentation methods work for stereo scenarios with only one view interaction, e.g. graph cuts [6] (left) and GrabCut [27] (right).

Extracting objects from stereo images with single-image tools usually causes inconsistency, because they need to be applied view by view, let alone the doubled workload. However, inter-view consistency is of great importance to supply good 3D experience. To overcome the problems, some previous methods apply stereo correspondence constraints to single image models and perform segmentation jointly or successively [21, 24]. We find that the computational cost for these models are quite high as they perform at least double operations. Since the two views of a stereo image pair are naturally near-duplicate, we believe that once the contour of one view is given, the extraction for the other view could be obtained with very little cost. Besides, previous methods suffer from the inaccuracy of current stereo matching algorithms, which may also lead to inconsistent extractions.

In this paper, we propose a contour based method to overcome the above limitations. First, after investigating the results of previous methods, it is obvious to conclude that most of the ground-truth regions can be covered by a rough segmentation, while errors usually occur on boundaries. Consequently, matching contours from one view to another seems to be a smarter choice other than matching entire images. Since the search space for contours is much smaller than pixels from entire images, the cost for consistent extraction is significantly reduced.

The second improvement is the relaxation of stereo correspondence constraint. Previous models apply implicit or explicit dense correspondence between views, which may suffer from the imperfectness of current stereo matching algorithms. In contrast, the matching of two contours are much easier and more accurate than dense stereo matching, since contours are usually full of gradient and appearance variance around. The problem of accurate contour correspondence is to miss occluded regions. To solve this problem, we add an object boundary term in our optimization function to pull the contours towards real boundaries, so that the occluded regions can be recalled.

Besides, some previous works [24, 23] integrate stereo correspondence constraints and single-image segmentation model into a unified framework, i.e. a single cost function in optimization. As a result, it is difficult for them to select different single view methods. In contrast, our method works as an independent module which does not care about specific methods for single view extraction. Any current single-image methods can effortlessly adopt our method to handle stereo tasks. We show an example in Fig. 1, where graph cuts [6] and GrabCut [27] are respectively combined with our method to conduct consistent object extraction.

We evaluate the proposed method on the Adobe open dataset [24], which includes 31 stereo image pairs and corresponding ground truth. Compared to state-of-the-art, our method shows superior extraction accuracy, and significantly improves efficiency. We further combine our method with some current single-image segmentation tools, like Magnetic Lasso [22], GrabCut [27] and [6, 11, 12, 20]. We show that our method is competent to make them handle stereo images. At last, we give some applications of our method to show its usefulness in stereo image editing.

The contribution of this paper can be briefly stated as follows. First, we propose a novel consistent object extraction method tailored for stereo images, which shows superior performance to state-of-the-art due to the exploration of object contour properties. Second, we show that our method can serve well as an independent module to combine with single-image methods, and consequently makes them work for stereo scenarios effortlessly.

2. Related Works

Single image segmentation. There are generally two categories of models for single image segmentation: boundary based and region based [32]. The former one, represented by snakes [17] and intelligent scissors [22], extracts an object by tracing its contour using single image properties. Our method differs from them in objective and searching space: we are searching for an optimal contour corresponding to a given one under stereo correspondence constraint. Another category of model is based on region, which considers both region statistics and inter-regional similarities, like graph cuts [6], geodesic distance [1], random walks [11] and [25, 10] etc. This category of methods simplifies the user interactions a lot, but are more computationally expensive.

Stereo image segmentation. In [24] Price et al. proposed a framework to simultaneously segment both views by integrating dense stereo correspondence term into the graph cuts model. The model is adopted by [16] and [23]. In [33] it has been shown that sparse correspondence can achieve comparable result. These methods, however, are computationally expensive and closed in framework, thus difficult to select different single image models. Lo et al. developed a stereoscopic 3D copy&paste system [21], which aims at extracting an object from one stereo image and composite to another one. For consistent segmentation, they employ a slight variant of Snapcut [2], which will be introduced later.

Contour matching and tracking. Some early works perform stereo matching/tracking on contours [17, 7]. Different from them, we are searching for real object boundaries instead of accurate contours, which will be discussed in Sect. 3.3. Besides, we exploit more powerful stereo cues instead of single image properties. Contours have also been used for object tracking [8, 36]. Our method differs from them in two aspects. First, contour tracking aims at tracing the motion of objects, which allows a certain error margin. In contrast, our method serves for stereo image editing, which has a much higher requirement on consistency. Second, the motion in contour tracking usually appearers to be a 2D flow, while for stereo image segmentation, epipolar geometry [37] is a more strict constraint to restrict the parallax between views.

Co-segmentation. Co-segmentation [28, 3] targets at picking out the same object from a collection of images. Rother et al. [28] model the problem as a MRF minimization that integrates smoothness and histogram matching term which forces similar foreground appearances. The research handles the extraction task on a collection of loosely related images, while for our scenario, stereo images are tightly related between views, and thus adaptable to more strict constraints on both boundary and appearance.

Video segmentation. Another research topic related to our work is video segmentation. The interframe consistency for video object extraction is similar to the case of stereo images segmentation. Snapcut [2] and Livecut [26] are two famous interactive video segmentation works. Users interactively segment the first frame and the result is used to guide the segmentation on consecutive frames. These works, like object tracking, mainly focus on temporal motion consistency. In contrast, for stereo images we care about spatial consistency caused by parallax.

3. Consistent Object Extraction

We give an overview of our method in Fig. 2. First, we extract the object-of-interest on either view using single-

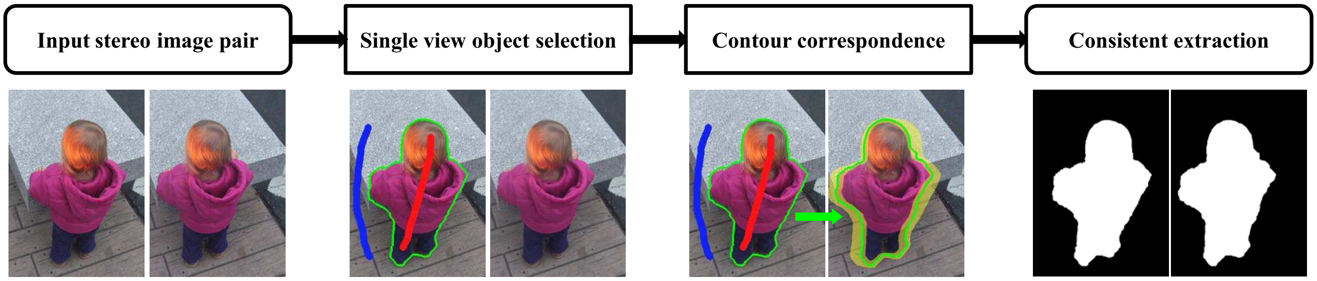


Figure 2. An overview of the proposed method. Given a pair of stereo image, our method allows user to interactively segment on one view using any current single-image methods. The consistent object extraction on the other view is then generated by the proposed method. Our method works on boundaries and supplies object masks as output.

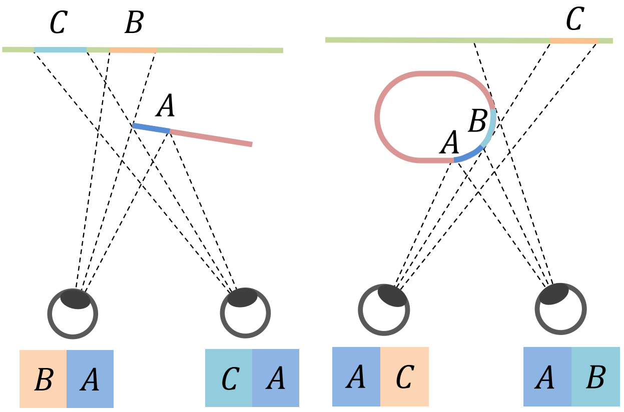
image methods. There are many good choices for single image segmentation, as introduced in Sect. 2. In this section we take graph cuts [6] for example, and in Sect. 4 we will give more results that our method combined with current single-image methods. Next, we transform the extraction result into contours and search for the corresponding contours in the other view. We integrate stereo correspondence and object boundary constraints into an energy minimization function to find the optimal contours. At last, we recover the extraction masks from the contours using [14].

3.1. StereoSnakes model

Suppose the extraction for one view is obtained and recorded in a mask K, where each entry $K(p_i)$ could be 1 or 0, indicating "Object" or "Background" respectively. We then extract the contours of the object using [31]. The borders are recorded as a set of vectors: $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_Z\}$. Each vector $\mathcal{C}_m$ indicates a closed curve, which encodes the location of boundary pixels clockwise: $\mathcal{C}_m = \{p_1, p_2, ..., p_{L_m}\}$. For each contour $\mathcal{C}_m$, we formulate the contour correspondence as an energy minimization problem in the disparity space:

$$E(d) = \sum_{p_i \in \mathcal{C}_m} C_S(p_i, p_i - d(p_i)) + \lambda_o C_O(p_i, p_i - d(p_i)) + \lambda_s N(p_i, p_{i-1}) \tag{1}$$

in the left part, d stands for the required disparities and $E(d)$ is the objective energy score. In the right of the equation, the first term denotes the matching cost between corresponding pixels p_i and $p_i - d(p_i)$ from two views. $d(p_i)$ is the disparity of p_i, which is restricted in the horizontal direction owing to the epipolar geometry [37]. The second term stands for the object boundary cost, which pulls contours towards real object boundaries. The last term is a smoothness function between adjacent pixels p_i and p_{i-1}. λ_o and λ_s are two weighting parameters to adjust the power of object boundary and smoothness term.



(a) Occlusion (b) Self-occlusion

Figure 3. Occlusion and self-occlusion illustration. (a) Due to binocular disparity, inconsistent local imaging between views occurs when there exists occlusion. (b) Part of object region missing caused by self-occlusion.

3.2. Stereo correspondence cost

The first term of Eq 1 indicates the matching cost between p_i and its corresponding pixel $p_j = p_i - d(p_i)$ in the other view. We measure the pixel-wise matching cost using the absolute difference between their colors:

$$C_{AD}(p_i, p_j) = \sum_{h=\{R,G,B\}} |c_h(p_i) - c_h(p_j)| \tag{2}$$

Considering of robustness, we aggregate the matching cost in a local surrounding window:

$$C_S(p_i, p_j) = \sum_{p_x \in \Phi(p_i) \wedge K(p_x)=1} C_{AD}(p_x, p_y) \tag{3}$$

where $p_y = p_x - d(p_i)$ and $\Phi(p_i)$ is a local window centered at p_i with size $w \times w$. $K(p_x) = 1$ indicates that we only aggregate the costs for object pixels. This is designed to overcome the inconsistency of local appearance due to binocular disparity. We illustrate the problem in Fig. 3 (a). A contour pixel with surrounding areas B and A in the left view appears to be encircled with C and A in the right

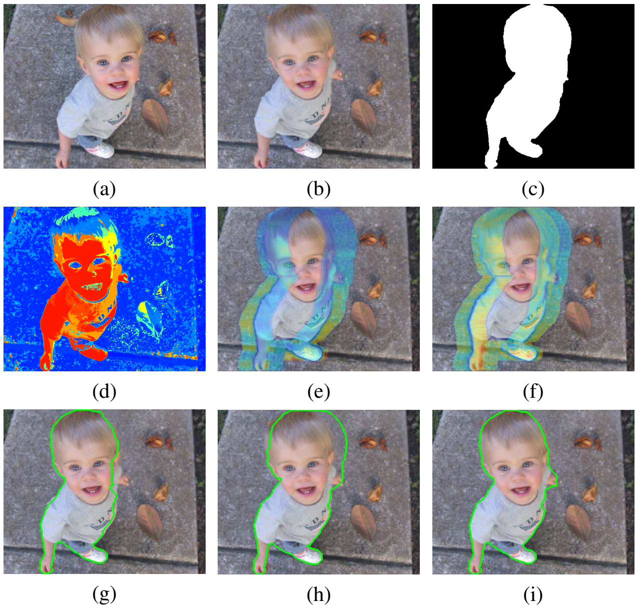


Figure 4. Illustration of the cost terms in our model. (a) Left image. (b) Right image. (c) Left mask. (d) Object likelihood of the right view calculated using histogram of the left. Warmer color (red) indicates higher object probability, vice versa. (e) Stereo correspondence cost C_S. Cooler color indicates lower cost, vice versa. (f) Object boundary cost C_O. (g) Optimal contour obtained using only C_S. (h) Optimal contour obtained using only C_O. (i) Contour jointly optimized using C_S and C_O.

view, which may lead to mismatch. However, our stereo correspondence term only aggregates the costs in area A and thus preserves consistency well.

Minimization of the stereo correspondence term forces the objective contour to have similar local appearances to the known one. Usually, the local areas around object contours are rich in gradient and appearance changes, since one side of the contour belongs to object and the other falls into background. This property makes it much easier to locate object contours than densely match entire images even using very simple matching cost functions, which is a major advantage of our method to previous works. The aggregation cost can be efficiently computed within a complexity of $O(1)$ using the integral image technique [35]. One could also choose advanced matching cost functions and aggregation methods [29, 13, 34, 15] to handle specific cases like photometric distortions, specular reflectance and other sophisticated problems.

3.3. Object boundary cost

Sometimes accurate contour correspondence will miss object regions due to self-occlusion as shown in Fig. 3 (b). We give an example in Fig. 4. The left hand of the boy is occluded in the left view but visible in the right. Consequently, an accurate contour correspondence will miss the invisible part as shown in Fig. 4 (g).

To solve this problem, we employ an object boundary term to pull the contours towards real object borders. We first use the extraction result in the left view to model foreground and background color distributions. We find that color histogram is sufficient for this task while being efficient. Suppose the histograms for object and background are denoted as H_O and H_B respectively, we measure the object boundary cost as:

$$C_O(p_i, p_j) = \sum_{p_x \in \Phi(p_i)} |Pr(O|p_y) - K(p_x)| \tag{4}$$

$$Pr(O|p_y) = \frac{H_O(c(p_y))}{H_O(c(p_y)) + H_B(c(p_y))} \tag{5}$$

where $p_y = p_x - d(p_i)$ and $c(p_y)$ is the RGB color of point p_y. $Pr(O|p_y)$ is the posterior probability of pixel p_y to be object. We show the object probability of the right view in Fig. 4 (d). C_O gets minimum score when and only when both the object and background part of $\Phi(p_i)$ matches the other view, that's where the real object boundary is.

We show C_O and C_S in possible searching space in Fig. 4 (e) and (f). It can be found that C_S favors accurate corresponding contours and thus tend to miss self-occluded object regions. In contrast, C_O has a more powerful discrimination to strong real object boundaries, but is prone to be puzzled by weak borders. The results in Fig. 4 (g) and (h) shows the defects of single cost terms respectively. Fortunately, a combination of the two terms complement each other and thus produces superior results, as shown in Fig. 4 (i).

3.4. Optimization using dynamic programming

Typically our objective function (Eq. 1) can be optimized using graph cuts [5, 18] or primal-dual methods [19]. However, in this task we find it can be solved more efficiently using dynamic programming in the disparity space. We first write the smoothness function as:

$$N(p_i, p_{i-1}) = \begin{cases} C_p, & \text{if} \quad |d(p_i) - d(p_{i-1})| \le \tau_d \\ \infty, & \text{otherwise} \end{cases} \tag{6}$$

where p_i and p_{i-1} are adjacent pixels in a contour $\mathcal{C}$. It should be noted that $\mathcal{C}$ is a closed curve and thus the last pixel p_L is adjacent to the first one p_0. C_p is a penalty term for discontinuous disparities: $C_p(p_i, p_{i-1}) = |d(p_i) - d(p_{i-1})|$. τ_d sets an upper bound of acceptable discontinuities. We should mention that the above smoothness function is used to restrict the change of the objective contour to that in the known view, while in previous works [17, 8, 36] they are mainly used to control the shapes of the contours.

Then we can give the state transition function:

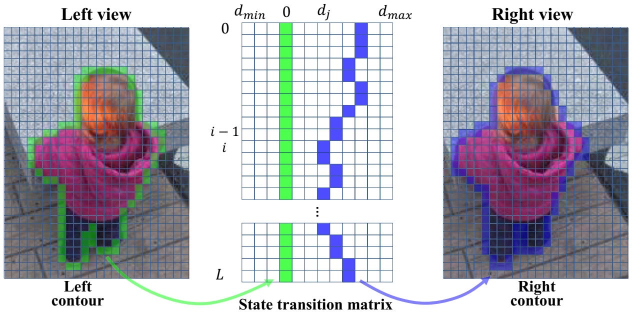


Figure 5. Illustration of our contour correspondence algorithm. Contours are extracted from the segmented view, then mapped to the disparity space. The corresponding contour path is then obtained by dynamic programming as described in Algorithm. 1. At last the contour is mapped back to the image space and the object mask is recovered.

$$E_{stf}(i,j) = \begin{cases} E(p_i, d_j), & i = 1 \\ E(p_i, d_j) + \min\limits_{t \in [j-\tau_d, j+\tau_d]} E_{stf}(i-1, t), & otherwise \end{cases} \tag{7}$$

where $E_{stf}(i,j)$ is the state energy. $E(p_i, d_j)$ is the correspondence energy for a single pixel p_i given a hypothesized disparity d_j:

$$\begin{aligned} E(p_i, d_j) = {} & C_S(p_i, p_i - d_j) \\ & + \lambda_o C_O(p_i, p_i - d_j) + \lambda_s N(p_i, p_{i-1}) \end{aligned} \tag{8}$$

Now we can get the state transition matrix according to Eq. 7. After the computation, we traceback the path with the lowest state energy to find the optimal contour. The traceback function is written as:

$$d(i) = \begin{cases} \text{index}\,(\min\limits_{t \in [1,D]} E_{stf}(i,t)), & i = L \\ \text{index}\,(\min\limits_{t \in [d(i+1)-\tau_d, d(i+1)+\tau_d]} E_{stf}(i,t)), & otherwise \end{cases} \tag{9}$$

where D is the possible disparity range and L is the contour length. $\text{index}(\min *)$ returns the index of the minimum value.

We given an example in Fig. 5 for illustration. Suppose the object is selected in the left view and the extracted contour is shown in green in the leftmost. The state transition matrix M, as shown in the middle, is of L rows and D columns, where D ranges from d_{min} to d_{max}. The left contour corresponds to the zero-disparity column, as shown in green. The state energy of each entry (i,j) is calculated as $E_{stf}(i,j)$ according to Eq. 7. After the calculation of the entire state transition matrix, we traceback along the minimum energy path, as shown in blue. Each node (i,j) in the path can be mapped to a pixel $p_i - d_j$ in the right view, as shown in the rightmost. Obviously, the complexity for our contour correspondence algorithm is $O(LD)$.

Algorithm 1 Contour Correspondence

```
Input:  $\mathcal{C} = \{p_1, p_2, ..., p_L\}$
Output: $\mathcal{C}' = \{p_1', p_2', ..., p_L'\}$
// State transition matrix calculation
for each cell $(i,j)$ in $M_{L \times D}$ do
    $M_{i,j} = E(p_i, d_j)$
end for
for $i = 2$ to $L$ do
    for $j = 1$ to $D$ do
        $M_{i,j} = \min(M_{i-1,j-1} + \lambda_s, M_{i-1,j},$
                $M_{i-1,j+1} + \lambda_s) + M_{i,j}$
    end for
end for
// Minimum energy path traceback
$d_L = \text{index}(\min(M_{L,1}, M_{L,2}, ..., M_{L,D}))$
$p_L' = p_L - d_L$
for $i = L-1$ to 1 do
    $d_i = \text{index}(\min(M_{i,d_{i+1}-1}, M_{i,d_{i+1}}, M_{i,d_{i+1}+1}))$
    $p_i' = p_i - d_i$
end for
```

Implementation details. In our implementation, we set τ_d as 1 since generally the contour disparities vary smoothly. The possible disparity range $[d_{min}, d_{max}]$ could be assigned experimentally, or roughly estimated using sparse feature point matching [4]. We give the complete process of our contour correspondence method in Algorithm. 1. The corresponding contour $\mathcal{C}'$ may be discontinuous due to disparity changes. So we link every two discontinuous pixels using a line to get a closed contour. At last, the object mask is recovered from the contour using [14].

4. Experiments and Analysis

4.1. Experimental settings

We evaluate our method on the Adobe open dataset [24], which is a benchmark designed for testing object extraction methods for stereo images. The dataset includes 31 stereo image pairs and manually labeled ground truth masks. The parameters of our method are set as $\{\lambda_O, \lambda_S, w\} = \{0.6, 30, 13\}$ throughout the experiments. We use 8^3 bins for color histograms, 8 bins per channel. All the experiments are conducted on a machine with a 3.4GHz Intel i7-4770 CPU and 16GB memory.

We first compare our method with state-of-the-art methods related to stereo image segmentation. After that, we show the capability of our method that extends current single-image methods to stereo scenarios. At last, we give

Table 1. Evaluation results on the Adobe open dataset. The first two rows show the number and percentage of mislabeled pixels of different methods. The last row gives the average runtimes of different methods.

Method	ST [24]	CT [9]	CO [3]	SN [2]	Ours
Errors (#)	481	1277	2995	1094	**439**
Errors (%)	0.23	0.61	1.43	0.53	**0.21**
Runtime (s)	0.650	0.413	0.532	0.715	**0.031**

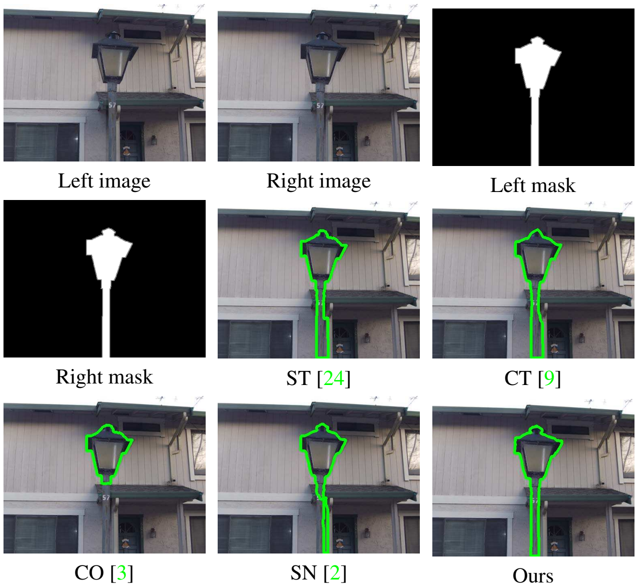

Left image
Right image
Left mask
Right mask
ST [24]
CT [9]
CO [3]
SN [2]
Ours

Figure 6. Comparison results on the stereo image "Lamppost1" from the Adobe open dataset. The results show the extractions on the right view with the left ground truth mask as input.

some applications of our method in stereo image editing.

4.2. Comparison with related methods

We compare our method to four state-of-the-art methods related to stereo image segmentation: stereocut (ST) [24], contour tracking (CT) [9], iCoseg [3] (CO), and Snapcut [2] (SN) [2]. We choose them according to a full coverage of related directions: stereo image segmentation, contour tracking, co-segmentation and video object extraction. Since our major concern is about consistent object extraction, for all the compared methods we input with ground truth masks of one view and compare the consistent extraction results on the other view. We employ the number and percentage of mislabeled pixels as evaluation metrics [24].

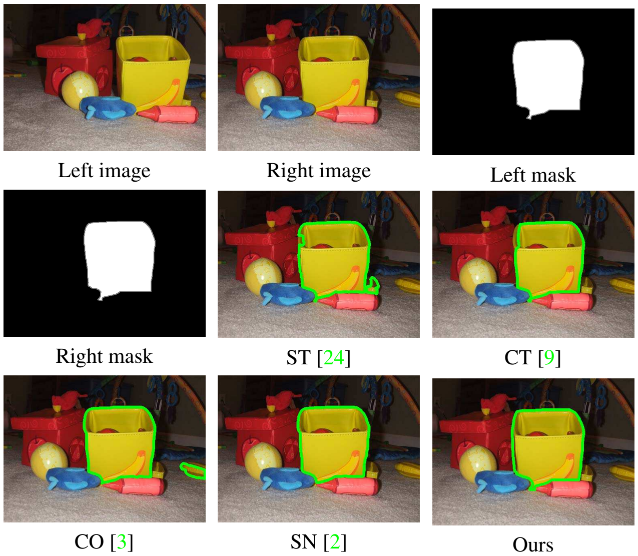

Left image
Right image
Left mask
Right mask
ST [24]
CT [9]
CO [3]
SN [2]
Ours

Figure 7. Comparison results on the stereo image "Toys2" from the Adobe open dataset.

The quantitative evaluation results are shown in Table. 1. The performances of the methods could be divided into three tiers: first the stereo methods (ST and ours), second the contour tracking and video methods (CT and SN), and last the co-segmentation method (CO). This can be explained by the fact that more strict constraints could lead to more accurate extractions. As we know, stereo epipolar constraint is stronger than temporal optical flow in videos, and further tighter than common object correlation in co-segmentation. As a result, the powerful cue makes stereocut and our method perform the best in the benchmark. However, stereocut is time expensive because it performs global optimization on both images with dense matching links between two views. The other methods are also time consuming due to similar reasons. In contrast, our method performs much more efficient because we eliminate the computation redundancy in those regions far from object contours. Besides, we implement a few efficient methods for acceleration, e.g. integral image for cost aggregation, histogram for color modeling, and dynamic programming for optimization. As a result, our method is able to efficiently work at more than 30fps in the Adobe open dataset, whose average image resolution is 416 × 502.

Fig. 6 and 7 show two examples of the results generated by different methods. In Fig. 6, ST and CT both recall extra parts due to occlusion and the high similarity between object and background. CO misses the post because it does not force strong spatial coherency between views. SN erroneously captures a background region while missing the post. In Fig. 7, ST captures a background region in occluded area. CT fails to track the object boundary in the upper left. CO detects an extra object that is similar to the desired extraction. In contrast, our method handles both the two cases well owing to a combination of stereo correspondence and object boundary constraints. We give more results to show the performance of the proposed method in Fig 8.

4.3. Combination with single-image methods

We extend single-image segmentation methods to stereo tasks with our method. Due to space limitation we only select a few typical works for illustration. We first show the effects of our method combined with graph cuts [6], random

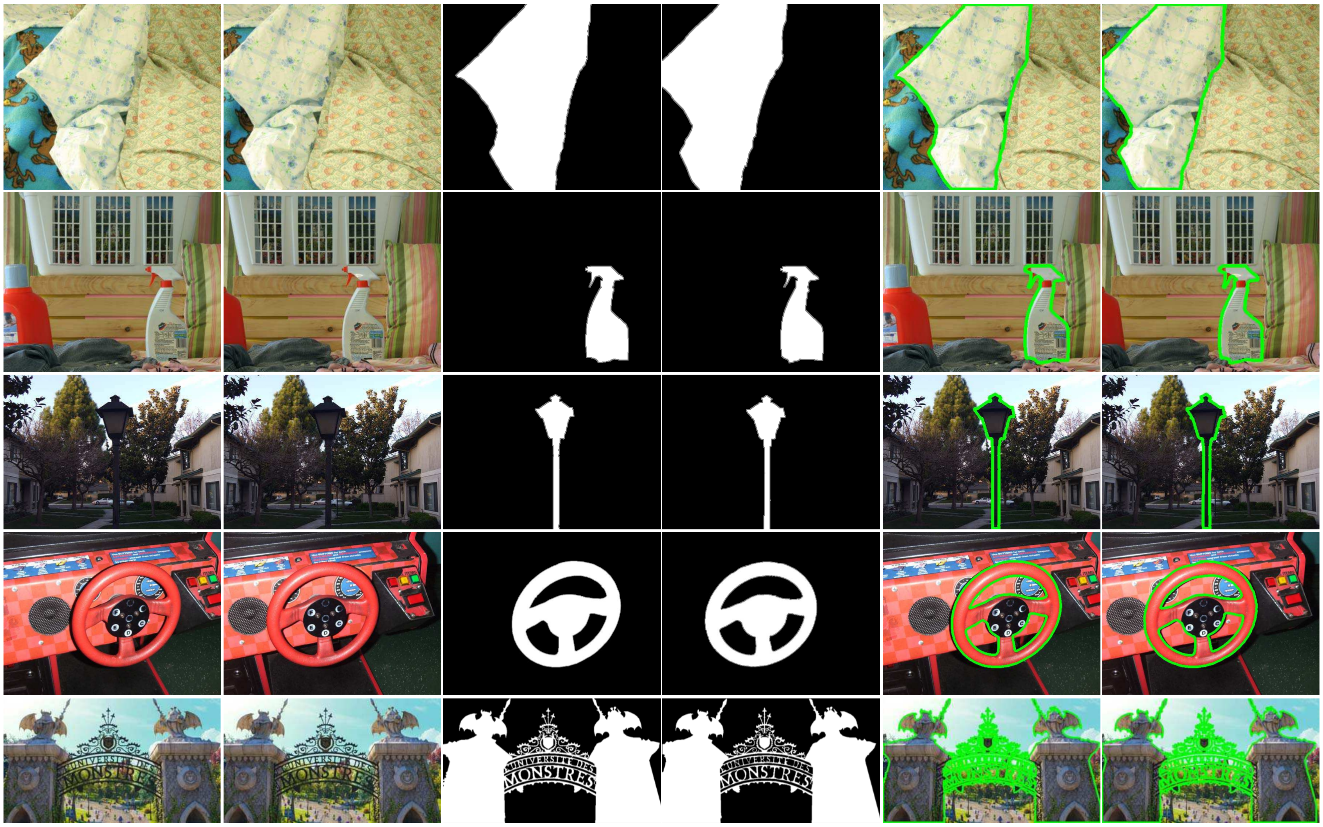
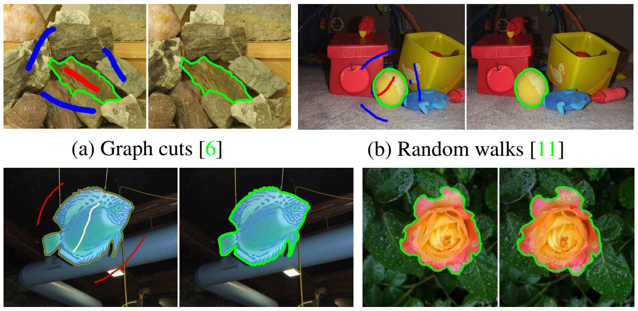


Figure 8. Consistent extractions. Our method is capable of handling occlusions (1st row), small or elongated objects (2nd and 3rd row), indistinct boundaries (3rd row), objects similar to backgrounds (3th and 4th row) and topologically complicated contours (5th row).



(a) Graph cuts [6]

(b) Random walks [11]

(c) Geodesic star [12]

(d) Branch-and-mincut [20]

Figure 9. Our method extends current single-image segmentation methods to stereo scenarios.

walks [11], geodesic star [12], and branch-and-mincut [20] in Fig. 9. The objects in the stereo images are first extracted in one view using the above methods. Then our method produces consistent extractions on the other view.

Next we show two examples using different interaction styles from above. In Fig. 10 we show the effects of combining the Magnetic Lasso tool [22] with our method. The Magnetic Lasso is a single-image object selection tool which traces object contours under rough user indications. With our method, the contour paths for both views can be simultaneously and consistently traced as shown in the intermediate results. Note open contours are treated as closed ones by linking the start and end point during the interaction. Another example is give in Fig. 11, where our method is combined with GrabCut [27] that employs bounding box selection as a first operation and scribbling as following refinement editing. The intermediate steps also show consistent extractions between views. Both the two examples show that our method could be of great help to improve user experience.

4.4. Applications

Consistent object extraction contributes a lot in stereo image editing. We give two demonstrations as follows.

Stereo image composition. In advertising and film industries it is popular to extract objects from one photograph and composite to another one. With several basic copy&paste operations and post processing, people can get novel photorealistic pictures. Object extraction plays as an essential step in image composition, which needs to be performed on both views for stereo case. We show an example using our method in Fig. 12. Given the input stereo images, we employ the Magnetic Lasso tool [22] and our method to consistently select the girl in both views. Then we paste the extracted person to another stereo image. Brightness

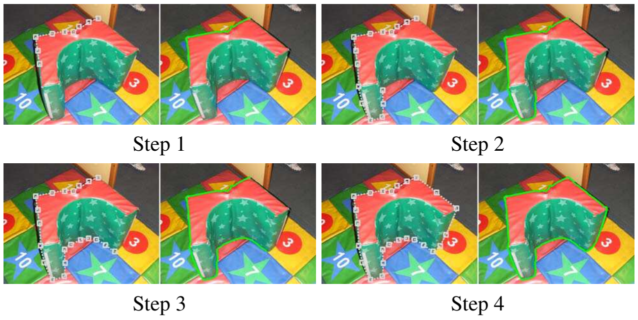


Figure 10. Magnetic Lasso adopts our method to handle stereo images. The intermediate results show consistent contour tracing results for both views.

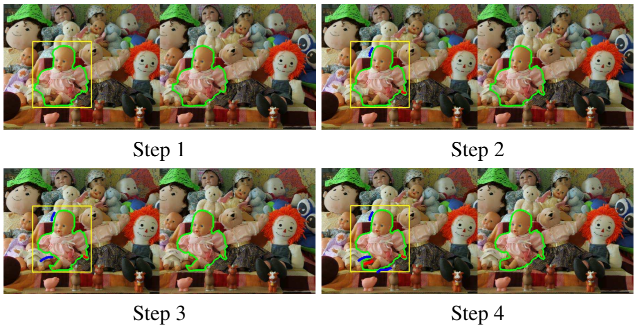


Figure 11. GrabCut extended to stereo scenarios with our method.

adjustment and linear alpha matting are consistently performed on both views. The composed result is shown in Fig. 12 (b) and the anaglyph is given in Fig. 12 (c), which is best viewed with red (left) - cyan (right) glasses.

Stereo image resizing. Content-aware image resizing are usually employed to change the size of an image to fit different displays while preserving important image content well. A simple and effective solution is to extract the important objects from background as a significance map to guide the resizing [30]. For stereo scenarios it is important to keep the content consistency between views during image resizing, where our method can be of service. We show an example of resizing a stereo image to 50% width in Fig. 13. Given the input stereo images, we first segment the important objects on both views as significance maps using the proposed method. Then we resize the background after inpainting and paste the objects onto the background. The results of the left and right views, and the red-cyan anaglyphs both show good 3D experiences owing to the consistent object selection.

5. Conclusions and Future Work

In this paper we have presented a novel consistent object extraction method tailored for stereo images. By utilizing the specific properties of object contour, our method achieves a superior extraction accuracy and significantly improved efficiency to state-of-the-art. Furthermore, the proposed method could effortlessly extend current single-image segmentation methods to work for stereo images. The experiments and extended applications show that our method is powerful in stereo image editing.

In the future, we will investigate more sophisticated stereo image editing tools such as image matting, completion, upsampling and so on. Consistency remains to be the focus problem for these applications, which requires more advanced processing. Besides, we will try to extend our method to stereo video segmentation to support the video editing applications. We believe that our method could be of great help to build the bridge from mono video segmentation to stereo scenarios.

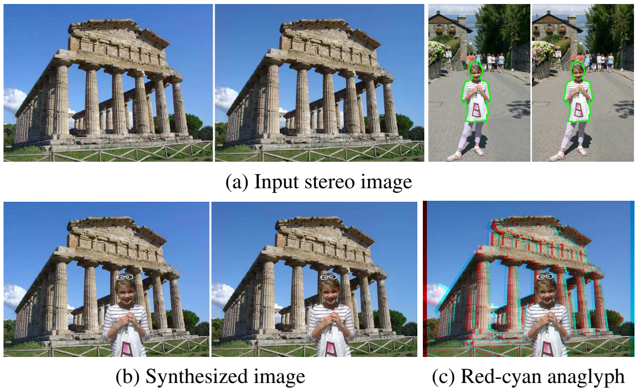


Figure 12. Stereo image composition.

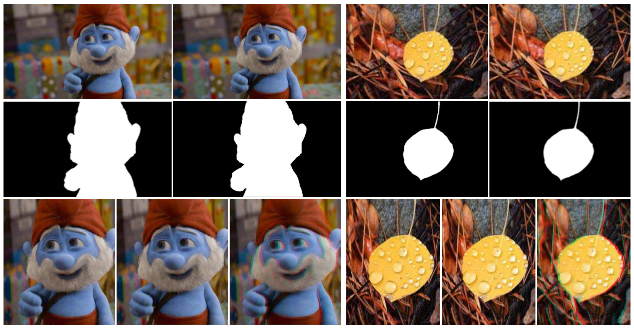


Figure 13. Content-aware stereo image resizing. First row: left and right images. Second row: left and right object masks. Third row: left and right results, red-cyan anaglyph.

Acknowledgements. This work is supported by the National Science Foundation of China under Grant No.61321491 and No.61202320, Research Project of Excellent State Key Laboratory (No.61223003), Natural Science Foundation of Jiangsu Province (No.BK2012304) and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

[1] X. Bai and G. Sapiro. Geodesic matting: A framework for fast interactive image and video segmentation and matting. *IJCV*, 82(2):113–132, 2009. 2

[2] X. Bai, J. Wang, D. Simons, and G. Sapiro. Video snapcut: robust video object cutout using localized classifiers. *ACM Trans. Graphics*, 28(3):70, 2009. 2, 6

[3] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *Proc. CVPR*, pages 3169–3176. IEEE, 2010. 2, 6

[4] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Proc. ECCV*, pages 404–417. Springer, 2006. 5

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23(11):1222–1239, 2001. 4

[6] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Proc. ICCV*, volume 1, pages 105–112. IEEE, 2001. 1, 2, 3, 6, 7

[7] T.-J. Cham and R. Cipolla. Stereo coupled active contours. In *Proc. CVPR*, pages 1094–1099. IEEE, 1997. 2

[8] Y. Chen, T. Huang, and Y. Rui. Parametric contour tracking using unscented kalman filter. In *Proc. ICIP*, volume 3, pages 613–616. IEEE, 2002. 2, 4

[9] C.-Y. Chung and H. H. Chen. Video object extraction via mrf-based contour tracking. *IEEE Trans. CSVT*, 20(1):149–155, 2010. 6

[10] C. Couprie, L. Grady, L. Najman, and H. Talbot. Power watershed: A unifying graph-based optimization framework. *IEEE Trans. PAMI*, 33(7):1384–1399, 2011. 2

[11] L. Grady. Random walks for image segmentation. *IEEE Trans. PAMI*, 28(11):1768–1783, 2006. 2, 7

[12] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *Proc. CVPR*, pages 3129–3136. IEEE, 2010. 2, 7

[13] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. CVPR*, pages 1–8. IEEE, 2007. 4

[14] K. Hormann and A. Agathos. The point in polygon problem for arbitrary polygons. *Comput. Geom.*, 20(3):131–144, 2001. 3, 5

[15] A. Hosni, M. Bleyer, and M. Gelautz. Secrets of adaptive support weight techniques for local stereo matching. *CVIU*, 117(6):620–632, 2013. 4

[16] R. Ju, X. Xu, Y. Yang, and G. Wu. Stereo grabcut: Interactive and consistent object extraction for stereo images. In *Advances in Multimedia Information Processing–PCM 2013*, pages 418–429. Springer, 2013. 2

[17] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988. 2, 4

[18] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE Trans. PAMI*, 26(2):147–159, 2004. 4

[19] N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. PAMI*, 29(8):1436–1453, 2007. 4

[20] V. Lempitsky, A. Blake, and C. Rother. Image segmentation by branch-and-mincut. In *Proc. ECCV*, pages 15–29. Springer, 2008. 2, 7

[21] W.-Y. Lo, J. van Baar, C. Knaus, M. Zwicker, and M. Gross. Stereoscopic 3d copy & paste. In *ACM Trans. Graphics*, volume 29, page 147. ACM, 2010. 1, 2

[22] E. N. Mortensen and W. A. Barrett. Intelligent scissors for image composition. In *ACM International Conference on Computer Graphics and Interactive Techniques*, pages 191–198. ACM, 1995. 2, 7

[23] J. Peng, J. Shen, Y. Jia, and X. Li. Saliency cut in stereo images. In *Proc. ICCVW*, pages 22–28. IEEE, 2013. 2

[24] B. L. Price and S. Cohen. Stereocut: Consistent interactive object selection in stereo image pairs. In *Proc. ICCV*, pages 1148–1155. IEEE, 2011. 1, 2, 5, 6

[25] B. L. Price, B. Morse, and S. Cohen. Geodesic graph cut for interactive image segmentation. In *Proc. CVPR*, pages 3161–3168. IEEE, 2010. 2

[26] B. L. Price, B. S. Morse, and S. Cohen. Livecut: Learning-based interactive video segmentation by evaluation of multiple propagated cues. In *Proc. ICCV*, pages 779–786. IEEE, 2009. 2

[27] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graphics*, 23(3):309–314, 2004. 1, 2, 7

[28] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In *Proc. CVPR*, volume 1, pages 993–1000. IEEE, 2006. 2

[29] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42, 2002. 4

[30] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch. Automatic image retargeting. In *Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia*, pages 59–68. ACM, 2005. 8

[31] S. Suzuki. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985. 3

[32] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010. 2

[33] H. E. Tasli and A. A. Alatan. User assisted disparity remapping for stereo images. *Signal Processing: Image Communication*, 28(10):1374–1389, 2013. 2

[34] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. In *Proc. CVPR*, pages 1–8. IEEE, 2008. 4

[35] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, volume 1, pages 511–518. IEEE, 2001. 4

[36] A. Yilmaz, X. Li, and M. Shah. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Trans. PAMI*, 26(11):1531–1536, 2004. 2, 4

[37] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *IJCV*, 27(2):161–195, 1998. 2, 3