Depth-Aware Image Seam Carving

Jianbing Shen, *Senior Member, IEEE,* Dapeng Wang, and Xuelong Li, *Fellow, IEEE*

***Abstract*—Image seam carving algorithm should preserve important and salient objects as much as possible when changing the image size, while not removing the secondary objects in the scene. However, it is still difficult to determine the important and salient objects that avoid the distortion of these objects after resizing the input image. In this paper, we develop a novel depth-aware single image seam carving approach by taking advantage of the modern depth cameras such as the Kinect sensor, which captures the RGB color image and its corresponding depth map simultaneously. By considering both the depth information and the just noticeable difference (JND) model, we develop an efficient JND-based significant computation approach using the multiscale graph cut based energy optimization. Our method achieves the better seam carving performance by cutting the near objects less seams while removing distant objects more seams. To the best of our knowledge, our algorithm is the first work to use the true depth map captured by Kinect depth camera for single image seam carving. The experimental results demonstrate that the proposed approach produces better seam carving results than previous content-aware seam carving methods.**

***Index Terms*—Energy optimization, image retargeting, Kinect depth camera, saliency, seam carving.**

I. Introduction

WITH THE rapid development of modern mobile devices and multimedia technologies in recent years, it is important to view the high quality images with different resolutions and aspect ratios on these devices. In order to make better use of the limited display space of mobile devices, we often need to resize the images with changing their aspect ratios. The simple resizing method is to scale and crop the original image to the screen size, which may introduce the scene distortion with changing the aspect ratio and severely affect the visual appearance effects [1]. Cropping operation can directly select the most interesting parts in images, but the other important information is cut off and lost. These traditional scaling and cropping methods only consider the output size to meet the geometric and resolution constraints without considering the content of images. Therefore, people need to develop an effective intelligent image resizing approach using the image contents to avoid the disadvantages of traditional resizing methods [1], [16].

Seam carving is an intelligent image resizing algorithm for adapting the image content to the display screen without distorting the important objects, which is first proposed by Avidan *et al.* [2]. It is an effective image resizing approach by removing the unimportant pixels using the gradient information and dynamic programming. Rubinstein *et al.* [5] extend the original image seam carving method and uses it for efficient image and video seam carving by introducing the graph cut based energy optimization. A seam is treated as a path of pixel removal with the minimal cost by one pixel for each row or each column in both of these methods. Shamir *et al.* [16] summarize that the visual media retargeting methods usually meet the constraints of keeping the important objects and maintaining the visual appearance without introducing artificial elements for the resultant image and video contents.

The existing image and video seam carving approaches can be roughly classified into three categories [16]. One is the continuous method that the image and video is expressed and processed as a sample of a continuous signal function, such as the warping approach [7] and mesh guided image and video retargeting method [8]. These methods find the mapping relationships from input data to output data with certain constraints. The second type is the discrete retargeting method where the images and videos are represented as the compact 2-D matrices or 3-D video cubes, such as the seam carving methods [2], [5], path-based approach [9], and the shift-map method [11]. In order to adapt the images and videos to the target size, these methods remove and regroup the pixels by blocks to achieve better resizing results. In addition, there are also a few methods that combine several existing methods to obtain the better visual quality result, such as the multioperator media retargeting method [12].

Seam carving algorithm is closely related to the saliency map detection approach [4], [30], [31] for indicating the visually important regions in the image. Cho *et al.* [13] use the importance diffusion map to define the energy of pixels in the neighborhood for the seam removal. The importance diffusion map is added to its neighborhood pixels to preserve the important objects and avoid the distortion of objects. In order to preserve the useful content in the retargeted image, Goferman *et al.* [19] present a content-based saliency detection method to define the significant regions that contain the

Manuscript received September 25, 2012; revised June 26, 2013; accepted July 8, 2013. Date of publication July 22, 2013; date of current version September 11, 2013. 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 61125106, Grant 61272359, and Grant 61072093; the Program for New Century Excellent Talents in University (NCET-11-0789); and the Shaanxi Key Innovation Team of Science and Technology (2012KCT-04). Recommended by Associate Editor J. Han.

J. Shen and D. Wang are with the Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn; windroc@126.com).

X. Li is with the 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, Shaanxi 710119, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2013.2273270

important objects. Recently, Basha *et al.* [24] proposed a seam carving method to resize pairs of stereo images simultaneously. They computed a disparity map as a depth map via the stereopsis algorithm from two stereo images, and resized these two images simultaneously with an extended seam carving algorithm. More recently, Niu *et al.* [26] presented a stereo saliency detection method from pairs of stereoscopic images, which is based on the global disparity contrast and the domain knowledge in stereoscopic photography.

Most seam carving approaches use the low-level visual features [27] to determine the important and salient regions [4], such as the gradient saliency features in [2]. The optimal seams computed by the conventional seam carving approaches [2], [5] are usually from both the foreground and the background pixels. It is still a challenging task to obtain the saliency and significant maps only from a single color image [33]. It is also difficult to determine the important objects using the low-level features especially when the input image has the large depth of field. Mansfield *et al.* [20] propose a scene carving algorithm using the user-defined mask image to simulate the scene depth information. Their algorithm requires an input image and a corresponding depth mask indicated by the user. Their depth mask gives the relative depth information of each object in the scene and thereby separates the foreground objects from the background area in the image. However, different users may have different opinions and interactions to define the important objects and produce the different depth masks by the users. Therefore, the good or poor seam carving results are obtained by using the accurate and inaccurate depth masks. Utsugi *et al.* [22] first attempt to present a content-preserving seam carving approach by using the disparity map from a pair of stereo images. Their method preserves the consistency of contents between the left and right images by fusing the stereo matching results into their seam carving framework. Recently, Basha *et al.* [24] propose a method to resize pairs of stereo images simultaneously. They compute a disparity map as a depth map via the stereopsis algorithm from two stereo images, and resize these two images with an extended seam carving algorithm to preserve the geometric consistency of the input stereo images.

In this paper, we propose a novel depth-aware image seam carving approach using an efficient depth-aware energy optimization. Our approach begins with capturing two images with the Microsoft Kinect depth sensor [23], [25], [29], [32]. One is the input RGB color image and the other is its corresponding depth map. Unlike the existing stereo seam carving [24], their approach requires a pair of stereo images to estimate the approximate depth map using the stereo matching algorithm. The depth information can provide valuable cues for image seam carving, which has been demonstrated by the scene consistent seam carving algorithm for a single image with the user defined depth mask [20]. Our approach belongs to the type of single image seam carving approaches, and we use the true depth map by Kinect depth sensor to guide the seam carving process and define our depth-aware energy function. Our approach can be viewed as an improvement on the state-of-the-art algorithms of single image seam carving, and can also be used for other applications such as the resizing of images from the online website for the mobile devices.

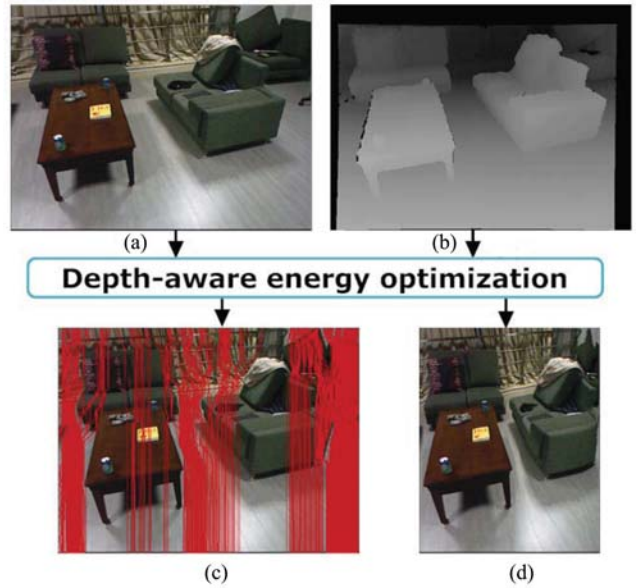

Fig. 1. Workflow of the proposed depth-aware image seam carving method. (a) Input color image. (b) Input depth map. (c) Removed seams. (d) Our result.

In contrast, to the best of our knowledge, our method is the first work to attempt to use the true depth map captured by Kinect depth camera for the task of efficient image seam carving. As a result, the image seam carving algorithm needs to be adapted to be applicable to the color image and its corresponding depth map. According to the input color image and its depth map, our approach is more easier to accurately detect the important and significant regions for calculating the seam carving energy. Based on the assumption that the important regions are changed as little as possible, we incorporate the color image and its corresponding depth map to define a depth-aware energy function. Our depth-aware energy is composed of the significant energy item and the depth energy item. In order to further improve the visual effects of the seam carving results, the just noticeable difference (JND) model is employed to preserve as much as the visual saliency pixels. The experimental results demonstrate that the proposed seam carving approach preserves the important and salient objects with good visual appearance after image resizing.

II. Our Approach

The conventional seam carving methods [2], [5] generally produce satisfactory resizing results for most images. However, as Avidan *et al.* have described in [2], the seam carving algorithm has some limitations and disadvantages. For instance, the seam may cross the salient and important objects, which will cause the unsatisfactory visual artifacts such as the shape and geometry distortion of important objects. Therefore, we use the depth cues as the extra significant information in the energy function. Our algorithm requires the input color image and its corresponding depth map, which is captured by an additional depth camera such as the Kinect sensor. Fig. 1

shows the workflow of the proposed depth-aware image seam carving method. Our seam carving method is based on both the significant energy and the depth energy. Our approach begins with the JND computation and use it to compute the depth-aware significant energy, and then the depth energy is defined by the captured depth map. In order to reduce the computational complexity, we use the multiscale graph cut based energy optimization strategy to accelerate the proposed algorithm.

A. Our Depth-Aware Energy Function

We develop a new seam carving algorithm by minimizing a significant energy using the depth map and the JND model. A gray scale depth map is used to represent the exact depth of the scene in our method. We use one to indicate the nearest points and zero to indicate the points that are too far or the points that we are not able to judge the distance in the captured depth map. The energy of conventional seam carving [2], [5] only considers the local gradient information to define the pixel saliency, and does not consider the global information of image content. The depth map provides the relative position of objects in the scene that easily detects the significant objects in the image. Moreover, the distance between the object and the camera is often associated with the object's significance, that is to say, we pay more attention to the near objects in the scene, and so the depth map provides such valuable cues for efficiently measuring the significance of objects.

Therefore, our depth-aware energy is composed of two items: the JND and depth-based significant energy item and the depth energy item. The significant energy defines the salient objects according to both the JND model and the depth map, while the depth energy measures the important objects by the depth map where the near objects usually have the large energy values. By considering both the significant energy and the depth energy with adjacent pixels after removing some pixels, the problem of finding the unnoticeable seams becomes the problem of finding the seams with the smallest energy. Then, the energy of pixel (i, j) for a vertical seam is relative to the pixel's column of the seam. Thus, our energy function is defined as follows:

$$E(i, j) = E_{sig}(i, j) + c_{dep} \cdot E_{dep}(i, j) \tag{1}$$

where $E_{sig}(i, j)$ is the JND and depth-aware significant energy, and $E_{dep}(i, j)$ is the depth energy. c_{dep} is the weight coefficient of the depth energy item, which is set to 0.2 in our experimental results.

The captured depth map gives the exact depth information of the objects in the scene, and we usually pay more attention to the near objects than the distant objects in the scene. For instance, the pixels with small depth values are mainly the background pixels, and they are more likely to be removed in the seam carving stage. In order to preserve the near significant objects as many as possible, a large depth energy is added to the pixels of the near objects while a small depth energy is added to the pixels of the distant objects in the scene. We have

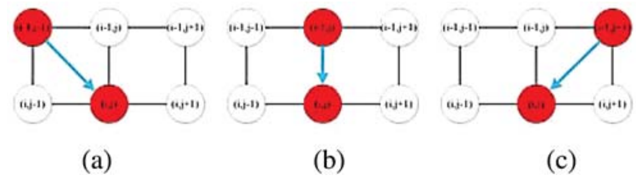

Fig. 2. Illustration of three possible positions for a vertical seam. According to the connection conditions, the pixel of the seam is possible among three positions (arrows in blue): the upper left neighbor, the upper neighbor and the upper right neighbor (red points).

found that the value of the smoothed depth map is enough to define this energy function as follows:

$$E_{dep}(i, j) = exp(-\|D(i, j)\|^2/\sigma) \tag{2}$$

where $D(i, j)$ is the depth value at pixel (i, j) in the depth map, and σ is used to define a smooth energy with the Gaussian weighting function.

B. Depth-Aware Significant Energy

We notice that the depth values of a single object in the depth map either looks similar or changes uniformly, and its first-order gradient value is small in both cases. The area with a large gradient value in the depth map is likely to be an important edge. The depth map provides a real depth boundary of the object, which will greatly improve the detection performance of the saliency map. Therefore, we define the depth-aware significant energy $E_{sig}(i, j)$ by considering both the input color image and its corresponding depth map simultaneously as follows:

$$E_{sig}(i, j) = E_{ig}(i, j) + E_{dg}(i, j) \tag{3}$$

where $E_{ig}(i, j)$ is the significant energy by the color image and $E_{dg}(i, j)$ is the additional depth energy by its depth map.

The additional depth energy of current pixel (i, j) is determined by the seam position on its adjacent pixels. As shown in Fig. 2, there are three possible positions according to the connectivity constraints of seams, and then these energy functions are defined as follows:

$$\begin{aligned} E_{dg}(i, j-1) &= |D(i, j+1)-D(i, j-1)|+|D(i-1, j)-D(i, j-1)| \\ E_{dg}(i, j+1) &= |D(i, j+1)-D(i, j-1)|+|D(i-1, j)-D(i, j+1)| \\ E_{dg}(i, j) &= |D(i, j+1) - D(i, j-1)|. \end{aligned} \tag{4}$$

The human visual system (HVS) is similar to a visual receptor, which is more sensitive to the area that its variation of pixel value is larger than certain threshold value. These areas usually contain most important visual information, which are often the significant information that people are interested in [15]. JND model plays such a role that defines the certain threshold mentioned above. If the differences between the corresponding pixels in two images are all less than the JND values, these two images are assumed to be no difference to human eyes. Therefore, we expect that the important areas that human visual system perceives should have more energy values than other areas, which can increase the energy of important pixels and reduce the energy of unimportant pixels.

Our JND-based significant energy enhances the visual appearance of important areas while attenuates the visual appearance of other areas. If the difference between the value of one pixel and the average value of its local background is less than the JND value, this pixel value is set to the average value. Otherwise, we add or subtract some value for this pixel to enhance the difference. This method increases differences between important pixels and their background ones, and decreases differences between unimportant pixels and their background ones. Our JND-based significant energy E_{ig} is then defined as follows:

$$E_{ig}(i,j) = |J(I(i,j+1)) - J(I(i,j-1))|$$
$$J(i,j) = \begin{cases} I(i,j) - JND(i,j), & \text{if } I(i,j) - I_B < -r(i,j) \\ I_B, & \text{if } |I(i,j) - I_B| < r(i,j) \\ I(i,j) + r(i,j), & \text{otherwise} \end{cases}$$
$$r(i,j) = c_{jnd} \cdot JND(i,j) \tag{5}$$

where $J(\cdot)$ is the JND-based image significant operator [15]. $I(i, j)$ is the pixel value at (i, j), and I_B is the average intensity value of local background. $r(i, j)$ is the value to be added or subtracted, and c_{jnd} is a control parameter.

Now we explain how to compute the JND value of an input image. The existing JND model mainly considers two factors: luminance adaptation (LA) and contrast masking (CM). The first factor LA simulates the process of HVS according to the background, which is defined as follows [21]:

$$JND(i,j) = LA(i,j) + CM(i,j) - c_{lc} \cdot \min(LA(i,j), CM(i,j))$$

$$LA(i,j) = \begin{cases} 17\left(1 - \sqrt{\frac{\bar{I}(i,j)}{127}}\right) + 3 & \text{if } \bar{I}(i,j) \le 127 \\ \frac{3}{128}(1 - \bar{I}(i,j) - 127) + 3 & \text{otherwise} \end{cases}$$
$$\bar{I}(i,j) = \frac{1}{32}\sum_{i=1}^{5}\sum_{i=1}^{5} I(x-3+i, y-3+j) \cdot B(i,j) \tag{6}$$

where c_{lc} is a control parameter that equals to 0.3 and $B(i, j)$ is a weighted low-pass filter.

The second factor CM indicates the influence among different pixels, which is composed of two components [21]: edge masking (EM) and texture masking (TM). Based on the assumption that the noise of texture structure is usually larger than the one of edge structure, and thus many texture structures should be treated as neither edges nor saliency areas. Therefore, the JND value of these texture areas should be larger than the one of edge areas. We first decompose the image I into two components ($I = U + V$): the structure image U and the texture image V. We then calculate the EM value from the structure image and the TM value from the texture image as follows:

$$CM(i,j) = EM^U(i,j) + TM^V(i,j)$$
$$EM^u(i,j) = C_s^U \cdot \beta \cdot W_e; \quad TM^u(i,j) = C_s^V \cdot \beta \cdot W_t \tag{7}$$

where β is a control parameter that is set to 0.117, and C_s is the largest luminance difference in a 5×5 neighborhood, which is determined by the maximum value in four different direction filters. We set $W_e = 1$ and $W_t = 3$ in our implementation.

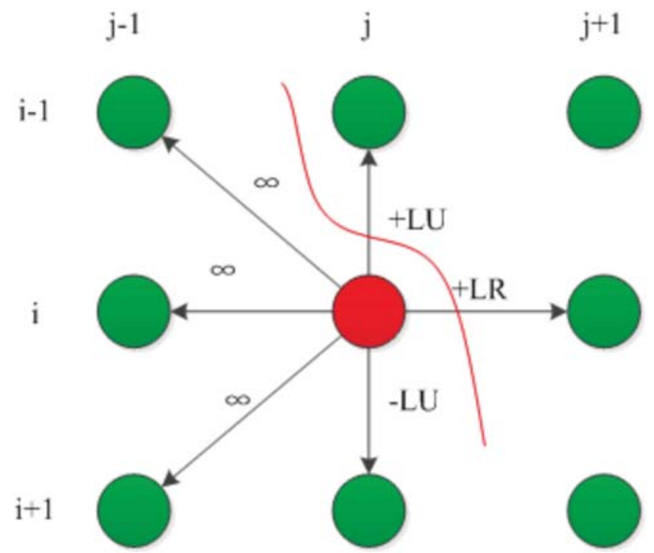

Fig. 3. Structure of graph cut based energy optimization. The red point (i, j) is the current pixel, and the green points are its neighboring pixels. There are edges that started from the red point and ended in the green points. Three of them with left arrows have infinite weights. The red line is the cutting boundary, and pixels on the left side of it are the seam to be removed.

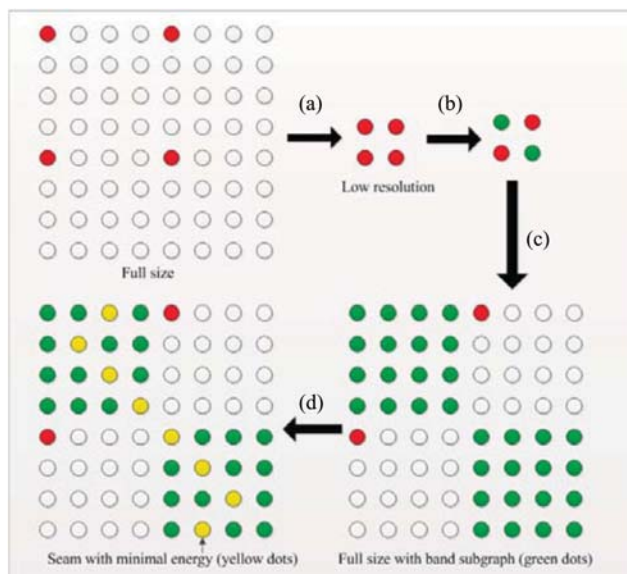

Fig. 4. Graph cut based multiscale optimization. (a) Downsample. (b) Optimal seam (green dots). (c) Upsample. (d) Seam optimization (yellow dots).

C. Optimal Seam by Graph Cut Based Optimization

We use the graph cut based optimization method to minimize our energy function. The image is first represented by a 2-D graph as shown in Fig. 3. The vertexes in the graph represent the image pixels and the edge weights are determined by our energy function. In order to find a seam with the minimal energy, the graph is divided into two subgraphs S and T, and the sum of the edge weights from S to T is also minimized. The pixels on the left of the boundary construct the seam with minimum energy. Our energy function consists of two items: the significant energy and depth energy. The

TABLE I
EXECUTION TIME STATISTICS WITH DIFFERENT DOWN-SAMPLING SCALES s

Scales	$s = 1$	$s = 2$	$s = 4$	$s = 8$	$s = 16$	$s = 32$
Time	1155.400	135.970	22.794	8.008	10.271	24.828
Results						

(a) (b) (c) (d)

Fig. 5. Comparison results between the improved seam carving method [5] and our approach on the Kinect data. (a) Input color image. (b) Input depth map. (c) Result by [5]. (d) Our result.

significant energy is calculated by the difference between adjacent pixels, which represents the energy after removing the seams. Our depth energy denotes the importance of current pixel, and then we add the average depth value between two adjacent vertexes to the edge weight. In Fig. 3, each vertex in the graph has eight neighboring vertexes. For the vertex at position (i, j), there are six edges starting from it. According to the constraint condition of connectivity, we set a very large weight to the edges that are connected with $(i-1, j-1)$,$(i, j-1)$ and $(i+1, j-1)$. Other edge weights are defined as follows:

$$\begin{aligned}
+LU &= |I(i-1, j) - I(i, j+1)| + 0.5 \cdot (D(i, j) + D(i-1, j)) \\
+LR &= |I(i, j-1) - I(i, j+1)| + 0.5 \cdot (D(i, j) + D(i, j+1)) \\
-LU &= |I(i, j-1) - I(i+1, j)| + 0.5 \cdot (D(i, j) + D(i+1, j))
\end{aligned} \tag{8}$$

where $+LU$, $+LR$, $-LU$ are the edge weights from (i, j) to $(i-1, j)$, from (i, j) to $(i, j+1)$, and from (i, j) to $(i+1, j)$, respectively. Thus, we use the graph cut based optimization algorithm to find the seam with minimum energy.

D. Multiscale Optimization Acceleration

The aforementioned seam carving algorithm removes one seam in each iteration, and the computation complexity is related to the constructed graph size of input image. When the input image is very large, the vertices and edges of the constructed graph are too large to be computed efficiently. In order to improve the computational efficiency, we use the multiscale graph cut based optimization algorithm [3] to accelerate our seam carving algorithm. Fig. 4 gives the structure of the efficient multiscale optimization approach. Our approach begins by finding the optimal seam for minimizing the energy in the down-sampled image. We then up-sample the result to the original image size and construct a multiscale graph structure, which is far small than the original image in size. Therefore, our multiscale optimization achieves better computational performance, which is further analyzed in Table I.

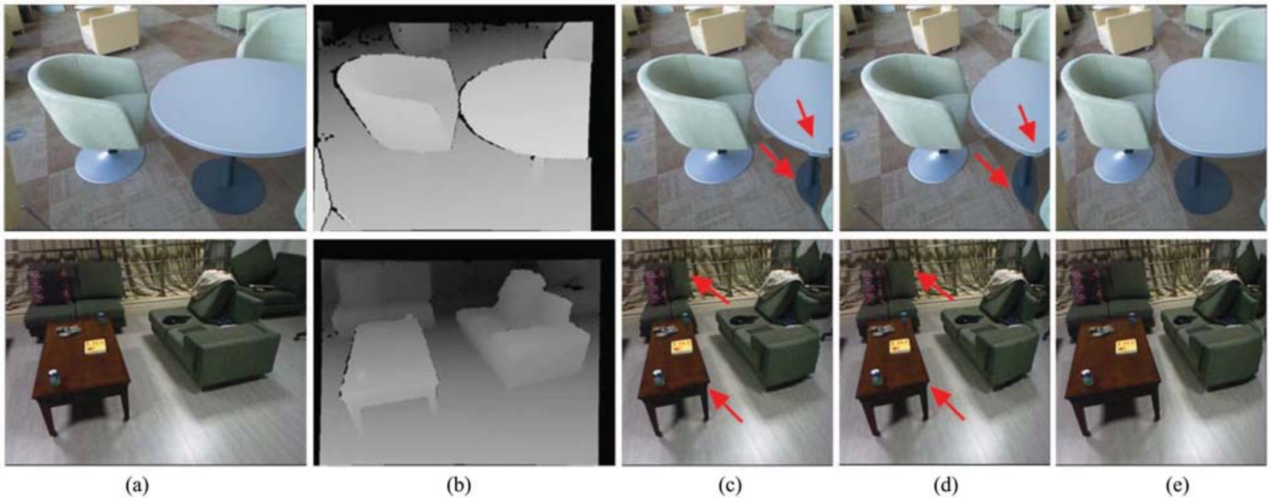

Fig. 6. Comparison between the seam carving methods [2], [5] and our method on the Kinect images. (a) Input color image. (b) Input depth map. (c) Result by [2]. (d) Result by [5]. (e) Our result.

By building the graph on a down-sampling image, the number of vertexes and edges is reduced significantly. Then the amount of computation is also reduced greatly. We calculate 170 seams from an input image of size 427×427 in Table I. The results and its corresponding running time statistics with different down-sampling scales are shown in Table I. It can be clearly seen that the retargeting results have little difference in visual appearance within a certain range of the down-sampling scales, while the computational efficiency is greatly improved. However, as shown in Table I, if the down-sampling scale is too large when it is greater than eight, the computational time will increase again. This is because when the down-sampling scale increases too much, then the size of the low resolution image becomes too small and the computational time to calculate the seams also becomes short. However, the size of the up-sampling graph becomes too large, the time for finding the optimal seam also becomes long. If the down-sampling scale is too large, the overall computational time will also increase. Therefore, the down-sampling scale is set to $s = 8$ to achieve the optimal computational performance in all our experiments.

III. Experimental Results

We have implemented our depth-aware image seam carving algorithm on a PC with Intel Core 2 Quad 2.66 GHZ CPU and 4 GB RAM. We use the Microsoft Kinect depth camera to capture the RGB color images and its corresponding depth maps. However, we can obtain the depth maps through a variety of different ways. For instance, the depth map can also be computed from a pair of stereoscopic images with some stereo matching algorithm [6]. We have tested our approach with these two types of depth maps, and the results in this section demonstrate that our method is not only suitable for true depth maps obtained by the modern depth cameras, but also applicable for the latter type of estimated depth maps.

We first compare our depth-aware image seam carving method with the improved seam carving approach in [5], and the depth maps are captured by the Microsoft Kinect depth camera. [1] We use the recently released OpenNI [2] to get the RGB color images and depth data from the Kinect system, and the size of the captured images is 640×480 in our implementation. Fig. 5 gives the comparison results between our approach and the improved seam carving method [5] with the Kinect images captured by ourselves. The image [Fig. 5(a), top] is resized to 60%, and the image [Fig. 5(a), bottom] is resized to 70%. From the given results, we have found that the important objects with large depth values are better preserved in our results, while the results by [5] [Fig. 5(c)] have unrealistic distortion of objects for the reason that their approach does not use the real depth map information to guide the seam carving process. Both of these two scene images have the complex backgrounds, and the curtain [Fig. 5(a), top] influences the estimation of the salient objects by [5], which causes the seam cutting across the human face and body parts. The optimal seams computed by our depth-aware energy method are mainly from the background pixels, and so both the human faces and the foreground objects are preserved well in our results. Similarly, as shown in Fig. 5(a), we are more interested in the objects of computer screen and the person in front of the computer than the objects of bookcase and background behind. Therefore, our approach produces better visual quality result [Fig. 5(d), bottom] for preserving the nearby person and the computer screen than the result [Fig. 5(c), bottom] by [5].

The main contribution of our method is that we utilize the true depth map to guide the single image seam carving process. Our approach preserves both the salient regions and the important near objects with large depth values. We further compare our method with the seam carving approaches in [2] and [5] on the captured Kinect data in Fig. 6. Since the original seam carving method [2] employs the image gradients to define the energy function, which will introduce the distortion of objects

[1] http://www.xbox.com/en-us/kinect

[2] *http://openni.org*

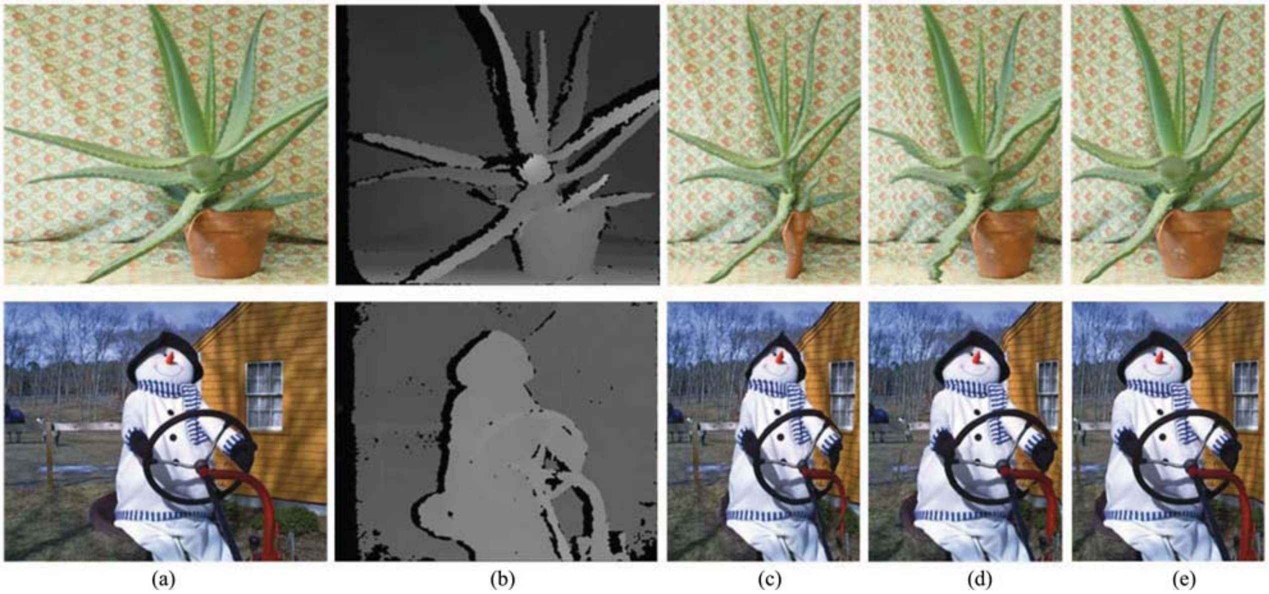

Fig. 7. Comparison between the seam carving methods [5], [24] and our method on the stereoscopic image pairs. (a) Input color image. (b) Estimated depth map. (c) Result by [5]. (d) Result by [24]. (e) Our result.

in the result image when the intensity variation of salient objects is smaller than that of the background region. As shown in Fig. 6, the important objects have severely distorted such as the shape of the blue round table, the dark green sofa and the brown table [Fig. 6(c), arrows in red]. However, the result by the improved seam carving approach [5] looks a little better than the results by [2], which still has the distinct shape distortions in the table and sofa regions [Fig. 6(d), arrows in red]. In contrast, our result exhibits the best visual quality for preserving the salient objects without introducing the shape distortion of the important objects [Fig. 6(e)].

As discussed above, we can obtain an approximate depth map from a pair of stereo images by the current stereo vision algorithm [6]. We use the stereo images in [24] for comparing the image seam carving results between our method and [24]. A disparity map is first computed from a pair of stereo images with the stereo matching approach. Then our method runs on one of the stereo color images and the calculated disparity map that is treated as the depth map in our framework. Our seam carving result is then compared with the result by the stereo seam carving method in [24]. As shown in Fig. 7, two pairs of stereo images are both resized to 60% of the original size. It is clear that our result achieves the better visual quality than both the results by the stereo seam carving method [24] and the improved seam carving approach [5]. The leaf of the aloe in the lower left corner by [24] is severely broken [Fig. 7(d), top], and the flowerpot is also heavily distorted in the result by [5] [Fig. 7(c), top]. In contrast, our result preserves the shapes of these objects much better [Fig. 7(e)], such as the shapes of flowerpot and snowman. In our method, the depth map information plays an important role in finding the significant regions to retain the foreground objects according to the proposed depth-aware energy. Furthermore, our significant energy helps to preserve the distant salient objects even with small depth values.

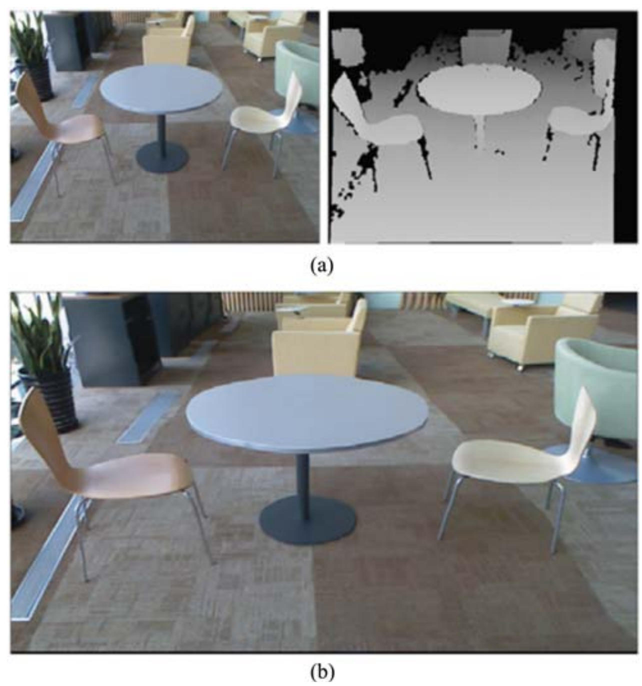

Fig. 8. Enlargement result by our seam carving method on the input image. (a) Input color image and its captured depth map. (b) Our enlargement result.

Until now, we have given the results on how to reduce the width of the input image by removing the unimportant seams. However, our approach can also be used to enlarge the width of image by adding the seams with our energy optimization framework. Our image enlarging algorithm is similar to the single image seaming carving algorithm [5]. This is implemented by first finding the optimal seams for removal and then duplicating these seams to enlarge the width of the original image in our depth-aware energy optimization

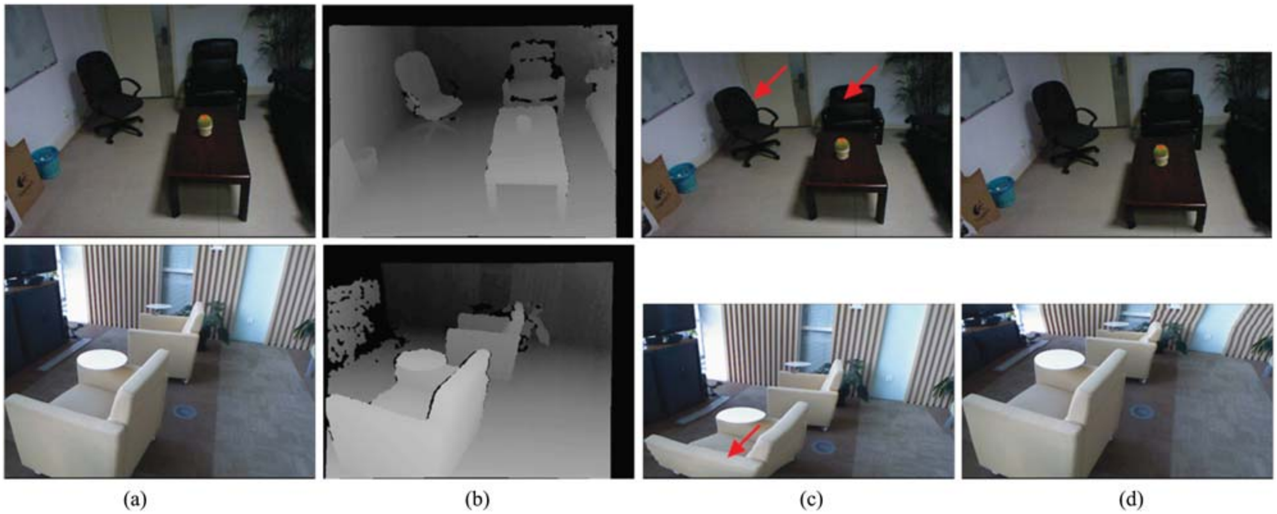

Fig. 9. Illustration of seam carving in the vertical direction between [5] and our method on the Kinect images. (a) Input color image. (b) Input depth map. (c) Result by [5]. (d) Our result.

framework. As shown in Fig. 8, the enlarging result by our approach achieves the satisfying visual effects. Our result preserves the scene consistency of the important objects well, such as the shapes and geometry of the chairs and tables [Fig. 8 (b)]. In addition, since the seams for removal in vertical and horizontal directions are symmetric, our seam carving approach can be applied to resize the height of the image by removing the optimal seams in the vertical direction. Fig. 9 gives an example to illustrate the seam carving results along the vertical direction. The input image is resized by changing the size of height with our optimization method. Our approach achieves better result than the one by [5] [Fig. 9(c), arrows in red] for preserving both the near objects with large depth values and the distant salient objects even with small depth values [Fig. 9(d)].

IV. Conclusion

We proposed a novel depth-aware single image seam carving approach by utilizing the Kinect depth camera in this paper. Our seam carving results for images are comparable to the current state-of-the-art image retargeting algorithms. By incorporating the RGB color image and its corresponding depth map, we developed an efficient depth and JND-based energy optimization approach that achieved a better seam carving performance for preserving the important and salient objects of image contents. The depth energy item in our optimization framework helped to preserve the important objects with small depth values well after seam carving. The multiscale energy optimization strategy enabled the proposed algorithm to work fast and efficiently with ease. The experimental results demonstrated that the proposed approach produced better retargeting results than previous seam carving methods on the image data sets by the Kinect depth camera.

The proposed framework can be extended in several directions. One direction is to further improve the runtime of the seam carving process. The complexity of the proposed method lies in solving the multiscale graph cut based energy optimization equation, which depends on the adjacency structure and the number of image pixels. We believe that the running time can be greatly improved by solving the cache-efficient graph cut algorithm [28] or adopting the graphics processing unit (GPU) acceleration implementation [10]. Another interesting avenue is to further optimize the initial depth map captured by the Kinect depth camera, which will improve the accuracy of the final depth map by inpainting and smoothing the captured depth map. Additionally, we hope to extend our framework for efficient video seam carving by capturing the video depth information, and further develop a GPU-based video seam carving approach for real-time applications in future work.

References

[1] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, and H. Q. Zhou, “A visual attention model for adapting images on small displays,” *ACM Multimedia Syst.*, vol. 9, no. 4, pp. 353–364, 2003.

[2] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” *ACM Trans. Graph.*, vol. 26, no. 3, pp. 10.1–10.9, 2007.

[3] H. Lombaert, Y. Sun, L. Grady, and C. Xu, “A multilevel banded graph cuts method for fast image segmentation,” in *Proc. IEEE ICCV*, 2005, pp. 259–265.

[4] Y. Yu, G. K. I. Mann, and R. G. Gosine, “An object-based visual attention model for robotic applications,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1398–1412, Oct. 2010.

[5] M. Rubinstein, A. Shamir, and S. Avidan, “Improved seam carving for video retargeting,” *ACM Trans. Graph.*, vol. 27, no. 3, pp. 16.1–16.9, 2008.

[6] A. Donate, X. Liu, and E. G. Collins, “Efficient path-based stereo matching with subpixel accuracy,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 183–195, Feb. 2011.

[7] L. Wolf, M. Guttmann, and D. Cohen-Or, “Nonhomogeneous content-driven video-retargeting,” in *Proc. IEEE ICCV*, 2007, pp. 1–6.

[8] Y. Wang, C. Tai, O. Sorkine, and T. Lee, “Optimized scale-and-stretch for image resizing,” *ACM Trans. Graph.*, vol. 27, no. 5, pp. 8.1–8.8, 2008.

[9] D. Simakov, Y. Caspi, E. Shechtman, and M. Iran, “Summarizing visual data using bidirectional similarity,” in *Proc. IEEE CVPR*, 2008, pp. 1–8.

[10] V. Vineet and P. J. Narayanan, “CUDA cuts: Fast graph cuts on the GPU,” in *Proc. IEEE CVPR Workshops*, 2008, pp. 1–8.

[11] Y. Pritch, E. Kav-venaki, and S. Peleg, “Shift-map image editing,” in *Proc. IEEE ICCV*, 2009, pp. 151–158.

[12] M. Rubinstein, A. Shamir, and S. Avidan, “Multi-operator media retargeting,” *ACM Trans. Graph.*, vol. 28, no. 3, pp. 23.1–23.11, 2009.

[13] S. Cho, H. Choi, Y. Matsushita, and S. Lee, "Image retargeting using importance diffusion," in *Proc. IEEE ICIP*, 2009, pp. 977–980.
[14] R. Achanta and S. Susstrunk, "Saliency detection for content-aware image resizing," in *Proc. IEEE ICIP*, 2009, pp. 1005–1008.
[15] C. Chiang, S. Wang, Y. Chen, and S. Lai, "Fast JND-based video carving with GPU acceleration for real-time video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 11, pp. 1588–1597, Nov. 2009.
[16] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. ACM SIGGRAPH Asia Course Notes*, 2009, pp. 11.1–11.13.
[17] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 160.1–160.9, 2010.
[18] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE CVPR*, 2010, pp. 569–576.
[19] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE CVPR*, 2010, pp. 2376–2383.
[20] A. Mansfield, P. Genler, L. Gool, and C. Rother, "Scene carving: Scene consistent image retargeting," in *Proc. ECCV*, 2010, pp. 143–156.
[21] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.
[22] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura, "Seam carving for stereo images," in *Proc. 3DTV-Conf.*, 2010, pp. 1–4.
[23] S. Hadfield and R. Bowden, "Kinecting the dots: Particle based scene flow from depth sensors," in *Proc. IEEE ICCV*, 2011, pp. 2290–2295.
[24] T. Basha, Y. Moses, and S. Avidan, "Geometric consistent stereo seam carving," in *Proc. IEEE ICCV*, 2011, pp. 1816–1823.
[25] A. Janoch, S. Karayev, Y. Jia, J. Barron, M. Fritz, K. Saenko, and T. Darell, "A category-level 3-D object dataset: Putting the Kinect to work," in *Proc. IEEE ICCV Workshops*, 2011, pp. 1168–1174.
[26] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *Proc. IEEE CVPR*, 2012, pp. 454–461.
[27] Y. W. Pang, H. Yan, Y. Yuan, and K. Wang, "Robust CoHOG feature extraction in human-centered image/video management system," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 458–468, Apr. 2012.
[28] O. Jamriška, D. Sýkora, and A. Hornung, "Cache-efficient graph cuts on structured grids," in *Proc. IEEE CVPR*, 2012, pp. 3673–3680.
[29] X. Ren, L. Bo, and D. Fox, "RGB-(D) scene labeling: Features and algorithms," in *Proc. IEEE CVPR*, 2012, pp. 2759–2766.
[30] Q. Wang, P. Yan, Y. Yuan, and X. Li, "Multi-spectral saliency detection," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 34–41, 2013.
[31] Q. Wang, Y. Yuan, and P. Yan, "Visual saliency by selective contrast," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1150–1155, Jul. 2013.
[32] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.
[33] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, Jan. 2013.

Jianbing Shen (M'11-SM'12) received the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2007.

He is currently an Associate Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. He has published more than 30 refereed papers in journals and conference proceedings. His current research interests include computer vision and computer graphics.

Dapeng Wang is currently pursuing the M.S. degree at the School of Computer Science, Beijing Institute of Technology, Beijing, China.

His current research interests include image and video retargeting.

Xuelong Li (M'02-SM'07-F'12) is a Full Professor with the 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, China.