FAST VOLUME SEAM CARVING WITH MULTI-PASS DYNAMIC PROGRAMMING

Ryosuke Furuta[†]*, Ikuko Tsubaki*[‡]*, and Toshihiko Yamasaki*[†]

[†]The University of Tokyo
{furuta, yamasaki}@hal.t.u-tokyo.ac.jp

[‡]Tokyo University of Technology
tsubakiik@stf.teu.ac.jp

ABSTRACT

In volume seam carving, seam carving for three-dimensional (3D) cost volume, an optimal seam surface can be derived by graph cuts, resulting from sophisticated graph construction. However, the graph cuts algorithm is not suitable for practical use because it incurs a heavy computational load. We propose a multi-pass dynamic programming (DP) based approach for volume seam carving that reduces computation time to 60 times faster and memory consumption to 10 times smaller than those of graph cuts, while maintaining a similar image quality as that of graph cuts. In our multi-pass DP, a suboptimal seam surface is created instead of a globally optimal one, but it has been experimentally confirmed by more than 198 crowd workers that such suboptimal seams are good enough for image processing.

***Index Terms*—** Seam carving, multi-pass dynamic programming, video retargeting, tone mapping

1. INTRODUCTION

This paper proposes a scheme for seam carving (finding seam surfaces) in a three-dimensional (3D) cost volume that features low computation time and memory consumption. Hereafter, we refer to seam carving for 3D cost volume as "volume seam carving." Volume seam carving has been applied to various image processing tasks, such as video retargeting [1], video summarization [2, 3], and tone mapping [4]. The volume seam carving procedure is as follows: A cost volume is first created, then a seam surface is determined that affects the cost less if it is removed. That seam surface is then removed, and the procedure is repeated until the targeted 3D volume is obtained. A seam surface is a two-dimensional (2D) manifold in the cost volume.

To achieve the volume seam carving task outlined above, to date, the graph cuts algorithm [1] has been the only choice because more efficient algorithms such as the dynamic programming (DP) algorithm used in still image seam carving [5] have not been considered for use in this area. Rubinstein et al. [1] stated that volume seam carving cannot be solved with DP and deemed graph cuts the only choice. However, it is well-known that the graph cuts algorithm requires a tremendous amount of computational time and memory when the number of nodes and edges increases. Consequently, in order to make such problems solvable, Rubinstein et al. [1] proposed a multi-resolution method and Chen and Sen [2] employed a video chunking method. However, the significant computation time and high memory usage of the graph cuts algorithm remain critical problems in the era of high-resolution images/videos.

In this paper, a multi-pass DP scheme that realizes volume seam carving with low computation time and memory consumption is proposed. Here, we propose two options: continuous mode and discontinuous mode. In continuous mode, the connectivity of the generated seam surface is ensured even though it is suboptimal, that is, not globally optimal. In discontinuous mode, connectivity is not guaranteed, but a lower energy path can be found. We describe a monotonic manifold as a seam surface even if it is unconnected. In video retargeting, seams may be amenable to discontinuous mode between frames, as discussed by Grundmann et al. [6], Chao et al. [7], and Yan et al. [8]. In contrast, in tone mapping, the connectivity of the seam surface is important, as discussed by [4]. For these reasons, both operational modes are presented here.

In this paper, the proposed method is applied to video retargeting and tone mapping in order to verify its efficacy. We present the results as well as the average computation time and memory consumption that respectively decrease to approximately one-sixtieth and one-tenth those of the graph cuts based solutions. We also verify via large scale subjective evaluation experiments that the suboptimal solution can produce an image quality roughly equivalent to that of graph cuts in these applications. It is important to note that this paper presents a new and general optimization method (multi-pass DP) for various kinds of volume seam carving based applications, not a specific application using multi-pass DP.

2. MULTI-PASS DP

We propose a new approach based on DP for volume seam carving. A cost volume is defined as $C(x_1, x_2, x_3)$, where x_1, x_2, x_3 are integers in the ranges $0 \leq x_1 < n_1$, $0 \leq x_2 < n_2$, $0 \leq x_3 < n_3$, and n_1, n_2, n_3 are resolutions of the cost volume in the x_1, x_2, x_3 directions, respectively. A seam surface that crosses the x_1 axis is defined as $S(x_2, x_3)$. When the seam surface passes coordinate (x_1, x_2, x_3), we describe it as $S(x_2, x_3) = x_1$. $(x_1, x_2, x_3) = (x, y, t)$ for video retargeting,

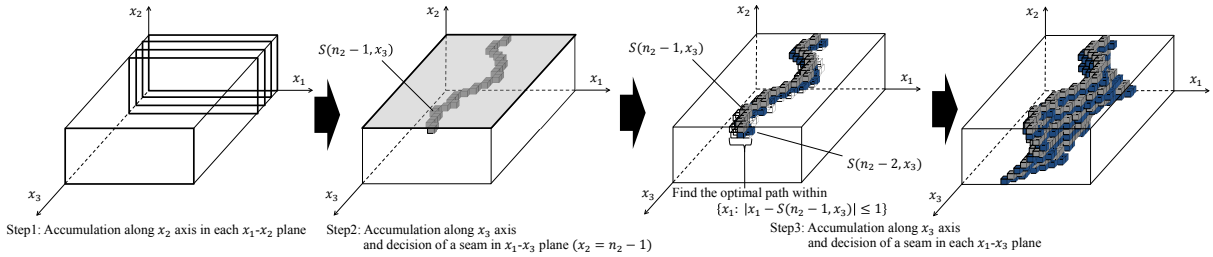

Fig. 1. Continuous DP process. The seam surface is connected in both the x_2 and x_3 directions.

which reduces the x resolution at each frame t. The value of cost volume $C(x_1, x_2, x_3)$ at each coordinate is obtained in the same manner as conventional methods.

2.1. Continuous method

Step 1: Accumulation along the x_2 axis

In this step, cost values are accumulated in each x_1-x_2 plane along the x_2 axis, from $x_2=0$ to n_2-1 using minimization, just like the original seam carving for image retargeting [5]. An accumulated cost function A_1 is obtained by

$$\begin{cases} A_1(x_1,0,x_3) = C(x_1,0,x_3), \\ A_1(x_1,x_2,x_3) = C(x_1,x_2,x_3)+ \\ \qquad \min\limits_{j\in\{-1,0,1\}} A_1(x_1+j, x_2-1, x_3). \quad (x_2>0) \end{cases} \tag{1}$$

Step 2: Accumulation along the x_3 axis and seam determination in the x_1-x_3 plane at $x_2=n_2-1$.

A seam is derived in the x_1-x_3 plane at $x_2=n_2-1$ in this step. Accumulated cost function $A_1(x_1,x_2,x_3)$ is furthermore accumulated in the x_3 direction, from $x_3=0$ to n_3-1, via minimization, and accumulated cost function A_2 is obtained. The value of j selected during the minimization is recorded in path $P_2(x_1, n_2-1, x_3)$.

$$\begin{cases} A_2(x_1,x_2,0) = A_1(x_1,x_2,0). \\ A_2(x_1,x_2,x_3) = A_1(x_1,x_2,x_3)+ \\ \qquad \min\limits_{j\in\{-1,0,1\}} A_2(x_1+j, x_2, x_3-1), \quad (x_3>0) \\ P_2(x_1,x_2,x_3) = \operatorname{argmin}\limits_{j\in\{-1,0,1\}} A_2(x_1+j, x_2, x_3-1)+x_1, \ (x_3>0) \end{cases} \tag{2}$$

The x_1 that minimizes $A_2(x_1,n_2-1,n_3-1)$ for the current x_2 is chosen and assigned to $S(n_2-1,n_3-1)$. The final seam is the optimal path obtained by following P_2 from $x_3=n_3-1$ to 0.

$$\begin{cases} S(x_2,n_3-1) = \operatorname{argmin}\limits_{x_1} A_2(x_1,x_2,n_3-1) \\ S(x_2,x_3) = P_2(S(x_2,x_3+1), x_2, x_3+1), \quad (x_3<n_3-1). \end{cases} \tag{3}$$

Step 3: Accumulation along the x_3 axis and determination of a seam in each x_1-x_3 plane.

A seam in each x_1-x_3 plane is derived in this step, starting from $x_2=n_2-2$ and reducing x_2 by one. First, the accumulated cost function A_2 is updated at each x_1-x_3 plane:

$$A_2(x_1,x_2,x_3) = \begin{cases} A_2(x_1,x_2,x_3), & (|x_1 - S(x_2+1,x_3)| \le 1) \\ \infty, & (otherwise). \end{cases} \tag{4}$$

This update has the effect of making the seams between each consecutive x_2 connected.

Second, accumulated cost function $A_2(x_1,x_2,x_3)$ and path $P_2(x_1,x_2,x_3)$ are obtained by accumulating $A_1(x_1,x_2,x_3)$ in the x_3 direction, from $x_3=0$ to n_3-1, in the same way as Eq. (2). The x_1 that minimizes $A_2(x_1,x_2,n_3-1)$ is then chosen and assigned to $S(x_2,n_3-1)$. The final seam is the optimal path obtained by following P_2 from $x_3=n_3-1$ to 0 in the same way as in Eq. (3). Subsequently, x_2 is reduced by one, and the derivation of the seam in the next x_2 is repeated, starting with Eq. (2). The complete seam surface is obtained when the process ends at $x_2=0$.

Fig. 1 shows the continuous DP process. The gray path shows the seam obtained at $x_2=n_2-1$. By successively creating a seam in each x_1-x_3 plane that is connected to the seam in the previous x_2, the seam surface becomes totally connected in the x_2 direction. In addition, the seam obtained in each x_2 is connected in the x_3 direction because j is selected from among $\{-1,0,1\}$ in Eq. (2). Hence, the obtained seam surface $S(x_2,x_3)$ is guaranteed to be connected.

2.2. Discontinuous method

Steps 1 and 2 are virtually the same as in Section 2.1. The only difference is that path P_1 in the x_2 direction is recorded in Eq. (1) to give

$$P_1(x_1,x_2,x_3) = \operatorname{argmin}_{j\in\{-1,0,1\}} A_1(x_1+j, x_2-1, x_3)+x_1, \quad (x_2>0). \tag{5}$$

Step 3: Seam determination in each x_1-x_2 plane

A seam in each x_1-x_2 plane is derived in this step. By selecting a path from the paths in P_1 that crosses seam $S(n_2-1,x_3)$ in the x_1–x_3 plane at $x_2=n_2-1$, a seam can be independently obtained in each x_1-x_2 plane.

$$S(x_2,x_3) = P_1(S(x_2+1,x_3), x_2+1, x_3), \qquad (x_2<n_2-1) \tag{6}$$

The complete seam surface is derived by obtaining a seam in every x_1-x_2 plane.

Fig. 2 shows the process followed by the discontinuous method. Although the gray seam in $x_2=n_2-1$ is connected in the x_3 direction because of Eq. (2), any seam obtained as an intersection of the seam surface and other x_1–x_3 planes are not guaranteed to be connected because Eq. (3) is calculated independently at each x_3. The connected seam in the x_1-x_3 plane at $x_2=n_2-1$, however, has the effect of making the seam surface prone to connecting in other x_1-x_3 planes.

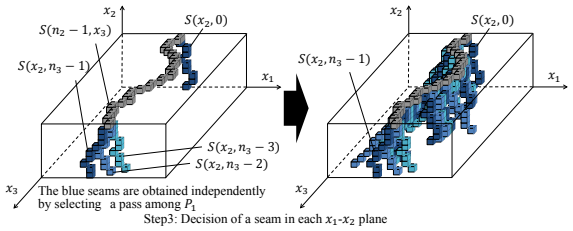

Fig. 2. Step 3 of discontinuous DP process.

Fig. 3. Results of video retargeting.

3. APPLICATIONS AND RESULTS

In order to show that the proposed method is general, we applied our proposed multi-pass DP to two different applications. All experiments were performed on an Intel Core i7-2600 3.4 GHz CPU machine with 16 GB of RAM, using methods implemented in C++.

3.1. Video retargeting

3.1.1. Experimental results

Video retargeting was achieved by applying the proposed method to the energy functions presented by Rubinstein et al. [1]. We adopted the forward energy proposed by [1].

We tested the proposed method with eight videos. The results for five videos are shown in Fig. 3[1]. The resolutions (width, height, number of frames) of the original videos 1–5 in Fig. 3 are (352, 288, 300), (148, 144, 131), (282, 288, 91), (540, 280, 99), and (540, 280, 97), respectively. For comparison, the results for seam surfaces derived by graph cuts are shown in Fig. 3(d). A simple multi-resolution method with a downsampling scale of 4 was adopted for the images in Fig. 3(d). No multi-resolution or downsampling was adopted for Figs. 3(b) and (c) because it was not necessary.

Videos 1 and 2 are scenes in which foreground objects are moving and the background is stable. The other videos are scenes in which both foreground and background are moving. In video 1 and 2, noticeable deterioration is not created in Figs. 3(b) and (d). In contrast, the foreground or background is shaking between frames in Fig. 3(c) because connectivity is not guaranteed in (c). In videos 4 and 5, some players are distorted, as shown in Fig. 3(b) and (d). It is clear that the continuous method is suited for videos that have little motion, and the discontinuous method is suited for videos whose motion is strenuous.

[1]The others are shown in the supplemental material.

Table 1. Comp. time and memory consumption for Fig. 3.

	(a) Computation time (s)		(b) Maximum memory consumption (MB)	
	Multi-pass DP (cont/ disc.)	Graph cuts	Multi-pass DP (cont./ disc.)	Graph cuts
Video 1	207 / 213	4,988	124 / 241	2.4 GB
Video 2	13 / 13	347	15 / 27	358
Video 3	83 / 85	4,645	41 / 77	721
Video 4	445 / 449	43,419	65 / 123	1.2 GB
Video 5	436 / 436	24,115	64 / 121	866
Average	237 / 239	15,503	62 / 118	1.1 GB

The computation time and maximum memory consumption are shown in Tables 1(a) and (b), respectively. By replacing graph cuts with multi-pass DP, the computation time is reduced to 1.5%. The maximum memory consumption is reduced to 5.6% by replacing graph cuts with continuous DP, and to 11% with discontinuous DP, even without multi-resolution or down-sampling. Note that the computation time and memory usage of our multi-pass DP increases by an order of $O(n)$, where n is the number of voxels in the cost volume. On the other hand, the computation time and memory usage of the graph cuts based method grow more rapidly with n.

3.1.2. Subjective analysis

We conducted subjective evaluation experiments to verify that the quality of the videos retargeted by our multi-pass DP is almost equal to that of graph cuts with respect to human perception. We used a crowdsourcing service and 163 participants took part in the test. The flow of the test was as follows: First, subjects were asked to watch four videos (the original and three videos retargeted by our continuous method, our discontinuous method, and graph cuts) on our web page, in which the original video is displayed at the center of the top row and the three retargeted videos are arranged horizontally on the row beneath. After that, subjects were asked to "choose the best video in terms of quality." They could replay the videos any number of times and were able to choose one of four choices, either "Video A/B/C is the best" or "Cannot notice the difference." The three retargeted videos and four choices were randomly ordered to remove any bias. We used the eight videos mentioned in Section 3.1.1 along with two other videos for dummy questions, in which three identical retargeted videos were displayed on the lower row. Therefore, subjects were asked to answer ten questions including two dummy questions. The order of the ten questions was also random. The reason why we add the dummy questions is to filter out the noisy results of dishonest subjects. In the supplemental material, we show the results of the subjects that correctly selected the choice "cannot notice the difference" in dummy questions: in other words, the subjects that were not fooled by the dummy questions.

Table 2. Number of subjects that preferred the retargeted video of each method.

	Multi-pass DP (cont. / disc.)	Graph cuts	Cannot notice the difference	Total
Video 1	24 / 8	39	**92**	163
Video 2	42 / 13	28	**80**	163
Video 3	40 / 17	22	**84**	163
Video 4	14 / 50	25	**74**	163
Video 5	19 / 39	37	**68**	163
Average	27.8 / 25.4	30.2	**79.6**	163

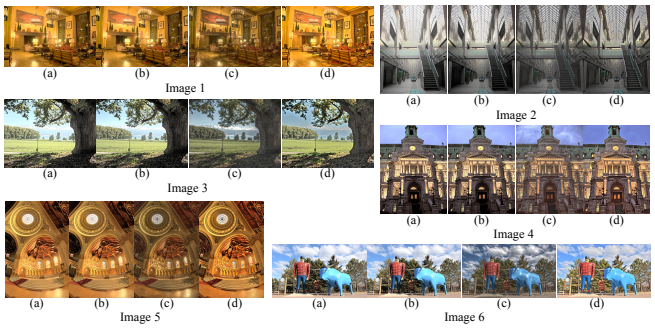

Fig. 4. Tone mapping results. (a) Multi-pass DP (continuous), (b) Graph cuts, (c) LLF [9], and (d) LEPF [10].

The result of the five videos of Fig. 3 is shown in Table 2[1]. We observe that on average, about half of the 163 subjects answered that there was no noticeable difference between the retargeted videos. On average, both of our methods obtain a score that is near that of graph cuts, and the sum of the two methods are significantly more than that of graph cuts. Therefore, in practical use, users can obtain an output that is equivalent to graph cuts by selecting the best result after applying both of our methods to the input video because our methods are very fast to calculate.

3.2. Tone mapping

3.2.1. Experimental results

We applied our method to the tone mapping based on volume seam carving, which was proposed by [4]. We show the detail of this algorithm in the supplemental material.

We tested the continuous method with 10 HDR images obtained from [11], [12], and [13]. The results for six images are shown in Fig. 4[1]. The resolution (width, height) of images 1–6 in Fig. 4 are (1000, 664), (760, 1016), (512, 381), (401, 535), (512, 768), and (1000, 563), respectively. For comparison, the results for seam surfaces derived by graph cuts are shown in Fig. 4(b). Figs. 4(a) and (b) look very similar. Because the average peak signal-to-noise ratio (PSNR) between Figs 4(a) and (b) is 33.5 and the average structural similarity (SSIM) [14] is 0.981, there are no visible differences.

Figs. 4(c) and (d) show the results for the local Laplacian filer (LLF) [9] and local edge-preserving filter (LEPF) [10] methods for comparison. On the whole, the detail is clear in Figs. 4(c) and (d); however, the contrast in Figs. 4(a) and (b) is higher than in Figs. 4(c) and (d). We show the quantitative analysis of these images in the supplemental material.

The computation time and maximum memory consumption are shown in Tables 3(a) and (b), respectively. By replacing graph cuts with continuous DP, the computation time was reduced to 2.2% and the maximum memory consumption was reduced to 51%. The number of iterations was 81 for each image.

Table 3. Comp. time and memory consumption for Fig. 4.

(a) Computation time (s)

	Multi-pass DP (continuous)	Graph cuts
Image 1	0.88	18.38
Image 2	4.17	245.29
Image 3	0.28	24.82
Image 4	0.28	3.69
Image 5	0.53	10.99
Image 6	0.76	26.09
Average	1.15	54.88

(b) Max memory consumption (MB)

	Multi-pass DP (continuous)	Graph cuts
Image 1	40	70
Image 2	46	82
Image 3	12	22
Image 4	14	24
Image 5	24	42
Image 6	34	90
Average	28	55

Table 4. Number of subjects that preferred the tone-mapped image of each method.

	Multi-pass DP (continuous)	Graph cuts	Cannot notice the difference	Total
Image 1	23	46	**129**	198
Image 2	**86**	48	64	198
Image 3	28	58	**112**	198
Image 4	30	28	**140**	198
Image 5	36	35	**127**	198
Image 6	13	38	**147**	198
Average	36.0	42.2	**119.8**	198

3.2.2. Subjective analysis

We conducted subjective evaluation tests for the tone mapping results, similar to those for video retargeting in Section 3.1.2. The number of subjects was 198, and only two tone-mapped images were displayed side by side on the web page. Subjects were asked to select one of three choices, either "Image A/B is better" or "Cannot notice the difference." We used the 10 HDR images mentioned in Section 3.2.1 and three other images for dummy questions, in which two identical images were displayed.

The results of six images of Fig. 4 are shown in Table 4[1]. We did not compare our method with the results of LLF [9] in Fig. 4(c) and LEPF [10] in Fig. 4(d). These methods are not volume seam carving based methods, and the purpose of this experiment is to verify that the same image quality can be obtained by our multi-pass DP and graph cuts. We observe that on average, 60.5% of 198 subjects noticed no difference between the two tone-mapped images.

4. CONCLUSION

In this paper, we proposed a fast volume seam carving method based on multi-pass DP and applied it to video retargeting and tone mapping applications. Experimental results showed that our method is approximately 60× faster and consumes 10× less memory than conventional graph cuts based methods.

Our large-scale subjective evaluation experiments showed that more than a half of participants noticed no differences between the results of our method and those of graph cuts.

This work was supported by Grants-in-Aid for Scientific Research (No. 26700008 and No. 16J07267) from JSPS and Microsoft IJARC core10.

5. REFERENCES

[1] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 16, 2008.

[2] B. Chen and P. Sen, "Video carving," in *Eurographics 2008, Short Papers*, 2008.

[3] Zhuang Li, Prakash Ishwar, and Janusz Konrad, "Video condensation by ribbon carving," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2572–2583, 2009.

[4] I. Tsubaki and K. Iwauchi, "Tone mapping based on luminance compression using seam surfaces," *IEICE Trans. Inf. Syst. (Japanese Edition)*, vol. 98-D, no. 4, pp. 606–615, 2015.

[5] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 10, 2007.

[6] Matthias Grundmann, Vivek Kwatra, Mei Han, and Irfan Essa, "Discontinuous seam-carving for video retargeting," in *IEEE CVPR*, 2010.

[7] Wei-Lun Chao, Hsiao-Hang Su, Shao-Yi Chien, W. Hsu, and Jian-Jiun Ding, "Coarse-to-fine temporal optimization for video retargeting based on seam carving," in *IEEE ICME*, 2011.

[8] Bo Yan, Kairan Sun, and Liu Liu, "Matching-area-based seam carving for video retargeting," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 23, no. 2, pp. 302–310, 2013.

[9] S. Paris, S. Hasinoff, and J. Kautz, "Local Laplacian filters: edge-aware image processing with a Laplacian pyramid," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 68, 2011.

[10] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 70–79, 2013.

[11] Frederic Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, no. 3, pp. 419–426, 2003.

[12] Mark D. Fairchild, *The HDR Photographic Survey*, MDF Publications, 2008, http://www.rit-mcsl.org/fairchild/HDRPS/HDRthumbs.html.

[13] H. Yeganeh and Z. Wang, "Objective quality assessment of tone mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, 2013.

[14] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.