FASTER SEAM CARVING FOR VIDEO RETARGETING

Chenyang Li[1], Ruimin Hu[1,2], Chao Liang[1,2], Chunxia Xiao[3], Weijian Ruan[1]*

[1]National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, China
[2]Collaborative Innovation Center of Geospatial Technology, China
[3]School of Computer Science, Wuhan University, China

ABSTRACT

Video retargeting is to resize a video to a desired resolution or aspect ratio while preserving its salient content without visual distortion. The key to video retargeting is to reconcile spatio-temporal coherence of video frames, and most existing works use seam carving to achieve that by employing the dynamic programming to find optimal seams. However, these methods are too time-consuming due to high computational complexity of the dynamic programming. To this end, we propose a novel method which uses discontinuous and suboptimal seams for seam carving. Concretely, we obtain the discontinuous seams by allowing seams to move freely in homogeneous regions of the frame, which helps preserve the spatio-temporal coherence effectively. Then, the genetic algorithm is employed to find suboptimal seams, so as to reduce computational complexity. Finally, each frame can be retargeted to a new aspect ratio or size by repeatedly carving out seams. Compared to the-state-of-the-art methods, the proposed algorithm achieves comparable results at an average expense of only one third of their running time.

Index Terms— video retargeting, spatio-temporal coherence, seam carving, computational complexity

1. INTRODUCTION

Video retargeting is to resize a video to a desired resolution or aspect ratio while preserving its salient content without visual distortion. With the growth of videos from various devices with different resolutions and aspect ratios, e.g. PCs, tablets, and cellular phones, video retargeting becomes an important technique to adapt these videos to different devices, besides, it is also very helpful in other areas of computer vision, such as video retrieval [14, 19] and video analysis [12]. However, traditional video scaling methods, such as uniform scaling or cropping, usually distorts or discards salient objects. Content-aware video retargeting, which resizes videos or changes their aspect ratios while preserving the shape of salient objects, provides satisfactory watching experience for the same video content on different viewing devices.

Content-aware video retargeting methods are divided into two categories, grid-based methods [3, 9, 10, 13, 16, 21] and seam carving methods [2, 4–6, 20]. Grid-based methods independently divide each frame into grids and align the grids between consecutive frames, each grid is aligned to the corresponding ones in other frames. Based on the alignments, the methods formulate an optimization over all frames. For example, Wang et al. [17] adopted crop-and-warp to grids to obtain spatio-temporal coherence. Li et al. [8] employed grid flow to resize video clips. Wang et al. [18] resized each frame independently and adopted optical flow to optimise the motion of salient objects. Compared with grid-based methods, seam carving methods remove pixels more precisely. The idea is to remove one pixel from every row or column in a frame every time. Rubinstein et al. [15] first introduced seam carving to video retargeting, which adopted surface carving method to reconcile spatio-temporal coherence. Grundmann et al. [7] exploited discontinuous seams for spatio-temporal coherence. When comparing the above two category methods, although grid-based methods have lower computational complexity, seam caving methods perform better for different types of video in preserving spatio-temporal coherence, which is more important for practical applications. To this end, we propose a novel seam carving based method, paying more attention to reducing the computational complexity while inheriting its advantages.

Inspired by discontinuous seam carving algorithm [7], we adopts discontinuous seams to realize more efficient video retargeting results. The discontinuous seams are obtained by allowing seams to move more freely in homogeneous regions of the frame. Fig.1 shows that how our discontinuous seams jump over the salient object without distortion (the player). However, different from [7] which exploits dynamic programming to find optimal seams, genetic algorithm is adopted to find suboptimal ones, which reduces computational complexity greatly while achieving comparable results. Dynamic programming traverses all possible solutions to find a optimal seam, while genetic algorithm searches for a suboptimal one in the finite solution space, which is much faster. A frame is then resized to a new aspect ratio or size by carving out seams repeatedly. The proposed method shows great superiority in balancing visual consistency and computational complexity.
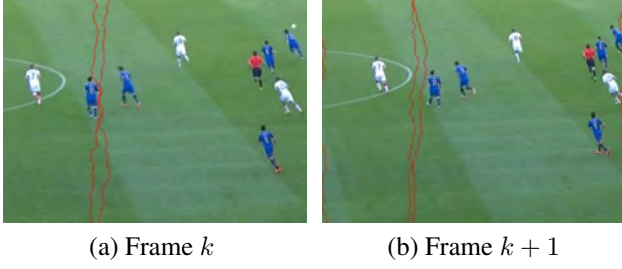
---

*Ruimin Hu is the corresponding author.

(a) Frame k (b) Frame $k+1$

Fig. 1. Discontinuous seams for frame k and frame $k+1$. The two seams just jump over the salient object without distortion.

2. VIDEO RETARGETING BY GENETIC SEAM REMOVAL

Our video retargeting algorithm resizes a video by finding and removing seams in an iterative way. We first calculate spatio-temporal coherence costs for every pixel to obtain the energy map for current frame. We then adopt genetic algorithm to compute suboptimal seams. By removing N seams from each frame we can reduce the width of the video by N columns.

2.1. Temporal Coherence

Suppose we perform the measurement in the kth frame I^k with resolution $m \times n$. The objective is to remove a seam S^k from I^k so that the resulting $(m-1) \times n$ frame I^* would be visually close to the most temporally coherent one, I^e, where is obtained by applying the seam S^{k-1} of previous frame I^{k-1} to current frame I^k. As is shown in Fig.2, for the pixel B in the frame I^{k-1}, the temporal cost is the difference between I^* and I^e if that pixel were removed. Sum-of-squared-differences(SSD) is used to represent the difference. Then the temporal coherence cost $T_c(u_b, u', v_b)$ can be formulated as:

$$T_c(u_b, u', v_b) = \sum_{x=0}^{u_b-1} \left\| I_{x,y}^k - I_{x,y}^e \right\|^2 + \sum_{x=u_b+1}^{m-1} \left\| I_{x,y}^k - I_{x-1,y}^e \right\|^2 \tag{1}$$

where $I_{x,y}^k$ here denotes the gradient magnitude of the pixel at (x, y) in frame I^k.

2.2. Spatial Coherence

Spatial coherence calculates how much spatial error will be introduced after removing a pixel. The cost of removing a pixel here is represented by the variation in the gradient of the intensity. Spatial coherence cost consists of two items, horizontal coherence cost S_p and vertical coherence cost S_q, Spatial coherence is formulated in a linear combination:

$$S_c = \alpha S_p + (1-\alpha) S_q \tag{2}$$

where α is a weight ratio, S_p is the change in horizonal gradient magnitude for removing a pixel. The horizontal cost of
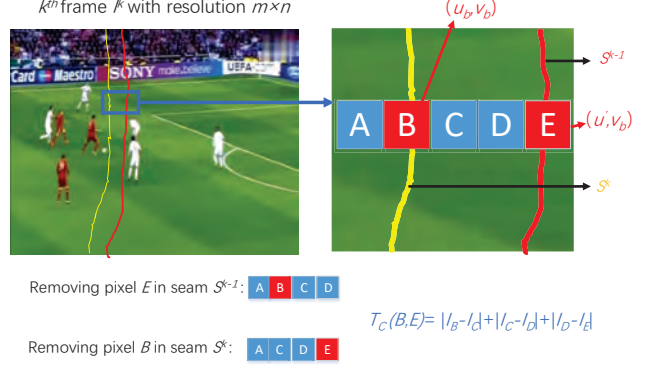

Fig. 2. Temporal coherence cost. The previous seam S^{k-1}(red) is applied to k^{th} frame. Current seam S^k (yellow) removes pixel B results in the row $ACDE$. The optimal temporally coherence seam S^{k-1} removes pixel E, so I^* would contain $ABCD$. We can compute the temporal coherence $T_c(B, E)$ for pixel B: $|I_B - I_C| + |I_C - I_D| + |I_D - I_E|$, where I_B here denotes the gradient magnitude of pixel B.

removing pixel E (in Fig.3 (a)) can be formulated as:

$$S_p(E) = |I_E - I_D| + |I_E - I_F| - |I_D - I_F| \tag{3}$$

where I_D, I_E, I_F here denotes the gradient magnitude of pixel D, E, F. and S_q is defined to measure the change in vertical gradient magnitudes when transitioning between a pair of pixels in adjacent rows. We consider the pixel E itself and whether the top neighbor of the pixel in question is its left (A), center (B), or right (C) neighbor. Fig.3 (a) corresponds to $S_q(E, A)$, which can be defined as:

$$S_q(E, A) = \left| |I_A - I_D| - |I_B - I_D| \right| + \left| |I_B - I_E| - |I_B - I_D| \right| \tag{4}$$

However, S_q only considers three neighbor pixels of the pixel E in the row above. Our vertical spatial cost allows a pixel to consider all pixels in the row above to get discontinuous spatial seams. For the pixel H (u_h, v_h) in the bottom row, the cost to pixel A $(u_a, v_h - 1)$ in the top row is:

$$S_q'(u_h, u_a, v_h) = \sum_{x=u_a}^{u_h-1} |G_{x,v_h}^q - G_{x,v_h}^d| + \sum_{x=u_a+1}^{u_h} |G_{x,v_h}^q - G_{x-1,v_h}^d| \tag{5}$$

where $G_{x,v_h}^q = |I_{x,v_h} - I_{x,v_h-1}|$ is the vertical gradient magnitude between pixel (x, v_h) and its top neighbor (dashed blue), while $G_{x,v_h}^d = |I_{x,v_h} - I_{x+1,v_h-1}|$ is its diagonal gradient magnitude with the top right neighbor (dashed red).

2.3. Genetic Discontinuous Seam Carving

We use genetic algorithm to find the suboptimal seam, then the retargeting process can be done iteratively. To generate
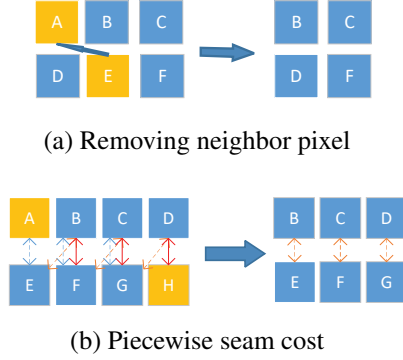
(a) Removing neighbor pixel

(b) Piecewise seam cost

Fig. 3. Spatial coherence costs: (a) Removing a neighbor pixel A in the row above of pixel E. Bottom row DEF becomes DF, therefore the intensity difference before removing E was $|I_D - I_E| + |I_E - I_F|$ and is $|I_D - I_F|$ afterwards. For the vertical coherence cost, the intensity difference between the two rows was $|I_A - I_D|$ and $|I_B - I_E|$ and is $|I_B - I_D|$ afterwards. (b) The cost of discontinuous spatial coherence cost. We consider all pixels in the row above.

valid seams, the genetic algorithm is redesigned to make sure the seam is connected [1]. In our genetic model, an individual only has one chromosome, while the gene composition of the chromosome can be encoded as a seam. We initialize ten individuals as the initial population, and then the population undergoes a certain algebraic evolution to obtain a suboptimal solution. We then decode the suboptimal solution to obtain the suboptimal seam.

The chromosome is defined as below:

$$Z = \{a_1, a_2, ...a_{i=p}, a_{i+1}, ...a_n\} \tag{6}$$

Where n is the height of the frame for vertical seams (for horizontal seams, n is the width of the frame). For every chromosome, we design a pivot $a_{i=p}$, its value presents a real position of a pixel, while other genes' values vary from -1 to 1, indicating three neighbors of the pixel in the neighbor rows. Assuming that we have a chromosome$\{-1, 0, 1, 5, 0, 1, 0, -1\}$, where the pivot is $(4, 5)$, and the seam can be decoded as:$\{(1,5), (2,6), (3,6), (4,5), (5,5), (6,6), (7,6), (8,5)\}$. In each generation, we use the single point crossover operator and the uniform mutation operator to generate new individuals.

Our fitness function has higher fitness value with lower energy individual:

$$f(x) = \frac{2(h-x)}{h-l} \tag{7}$$

where l is the lowest energy and h the highest energy of individuals in current generation, and x presents the energy value of the specific pixel in the energy map.

(a) original frame   (b) our results   (c) [7]'s results   (d) [15]'s results   (e) [11]'s results
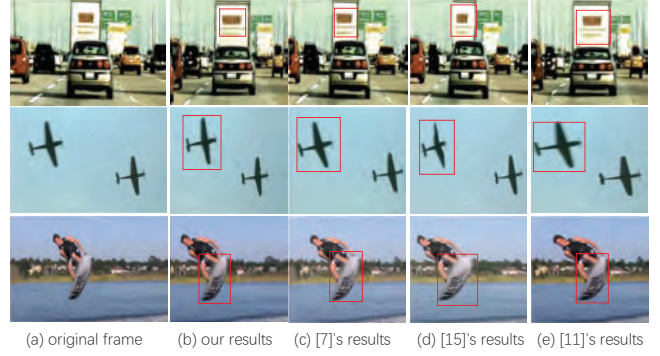
Fig. 4. Retargeting results for different methods. (a) the original frames. (b) our results. (c) Grundmann's results. (d) Rubinstein's results. (e) Oliveira's results.

Table 1. Comparisons with three seam carving based methods on 1500 preferences from 50 subjects. The values in each row indicate the corresponding method's preferences compared with others.

	Ours	[7]	[15]	[11]	Total
Ours	-	**162**	**175**	**184**	**521**
[7]	88	-	169	177	434
[15]	75	81	-	130	286
[11]	66	73	120	-	259

3. EXPERIMENTAL RESULTS

In our experiments, we first compare our retargeting results with other seam caving retargeting methods; we then evaluate the computational complexity. We also show the effect of genetic algorithm's different parameters on retargeting results. For fair comparison, our test videos are mostly same as [7]. All the experiments were performed on a PC with: Intel i7-4720HQ processor with four cores at 2.6GHz, and 8GB DDR4 RAM.

Qualitative comparisons: Our qualitative comparisons of retargeting results are shown in Fig.4. For the first row, our method and [7] perform well in preserving the shape of the truck, while [15] distorts the edges. The plate on the truck saying "YELLOW" is clearer in our retargeted frame, while [11] makes the plate blurred. For the second row, [15] makes the left aircraft deformed and produces many visual artifacts, and [7] cuts off too many regions around the edge of

Table 2.Run time cost for videos with different resolutions.

Video resolution	340*280	480*260	640*256
Ours	21s	35s	51s
Grundmann [7]	68s	118s	175s
Rubinstein [15]	196s	100s	1469s

(a) original frame   (b) p = 10   (c) p = 20   (d) p = 30   (e) p = 40

Fig. 5. Retargeting results for different initial populations.

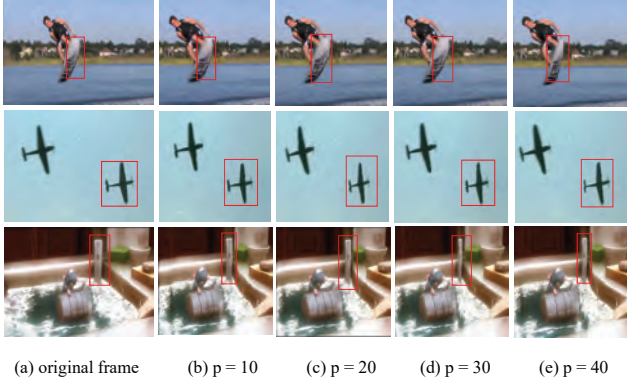(a) original frame   (b) n = 3   (c) n = 5   (d) n = 7   (e) n = 9

Fig. 6. Retargeting results for different generations.

Table 3.Run time cost for different retargeting factors.

Retargeting factor	10%	20%	30%
Ours	21s	38s	56s
Grundmann [7]	68s	132s	181s
Rubinstein [15]	196s	475s	768s

frame. Our method removes the regions around the edge and the middle of two aircrafts, maintaining better spatial coherence. For the third row, our method and [7] preserve the shape of the player, while [15] distorts the player's head, and [11] distorts the skateboard. [11] produces too much artifacts. The results show that our method performs better.

Fig.5 and Fig.6 show the effect of genetic algorithm parameters. Fig.5 shows retargeting results for different initial populations. p here denotes initial population. Retargeting result is best when $p = 30$ for the first row, while 10 for the second, and 10 for the third. Likewise, Fig.6 shows the retargeting results for different generations. The best is $n = 5$ for the first row, while 7 for the second, and 9 for the third.

Quantitative comparisons: We conduct a user study to subjectively evaluate the performance. We invite 50 subjects of different ages to participate in the user study. The retargeted videos are presented in a random order to avoid personal bias. We compare our method with [7], [15] and [11] over five different videos and perform 6 paired comparisons for each video. we receive $5 \times 6 \times 50 = 1500$ answers in total and each method is pairwise compared by $3 \times 5 \times 50 = 750$ times. Table 1 shows the count of preferred retargeted videos of each method by paired comparisons. Note that in Table 1 each entry means that the method in row r is ranked better n times than the method in column c. Table 1 shows that our method outperforms the other three methods in 69.4% (521/750) of the paired comparisons. The breakdown is that our method is preferred in 64.8% (162/250) with [7], in 70.0% (175/250) comparisons with [15], and in 73.6% (184/250) with [11].

Besides, we demonstrate effectiveness of our algorithm on reducing computational complexity in two aspects. First
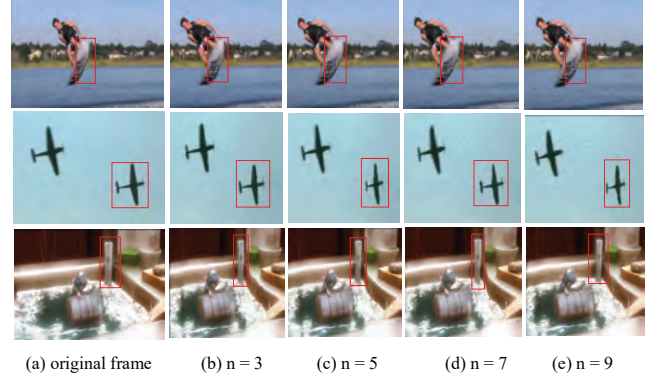
we make an experiment to test these algorithms' performance for videos with different resolutions. The retargeting factor is 10%, and the total number of frames is 15. The experiment results are shown in Table 2. For every tested video, We can observe that our algorithm runs about three times faster than [7], while [15] is the slowest. Second, we make an experiment to test these algorithms' performance for a certain video with different factors. As is shown in Table 3, the experiment results show the run time cost for different factors. For every factor, our algorithm runs about three times faster than [7].

4. CONCLUSIONS

In this paper, we proposed a novel seam carving based algorithm for video retargeting. By adpoting discontinuous seams, the proposed method preserves the spatio-temporal coherence effectively. Besides, genetic algorithm is adopted to find suboptimal seams, reducing computational complexity greatly. Extensive experiments demonstrate the proposed method performs favorably against the state-of-the-art method, especially in balancing visual consistency and computational complexity. In the future, we will further reduce the time complexity of our method to meet the real-time requirement.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, 26(3):10, 2007.

[2] Tali Dekel Basha, Yael Moses, and Shai Avidan. Stereo seam carving a geometrically consistent approach. *IEEE transactions on pattern analysis and machine intelligence*, 35(10):2513–2525, 2013.

[3] S Battiato, G. M. Farinella, G Puglisi, and D Ravi. Saliency-based selection of gradient vector flow paths for content aware image resizing. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 23(5):2081–2095, 2014.

[4] Wei Lun Chao, Hsiao Hang Su, Shao Yi Chien, Winston Hsu, and Jian Jiun Ding. Coarse-to-fine temporal optimization for video retargeting based on seam carving. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2011.

[5] David D Conger, Hayder Radha, and Mrityunjay Kumar. Seamlets: Content-aware nonlinear wavelet transform. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 1450–1453. IEEE, 2010.

[6] Ryosuke Furuta, Ikuko Tsubaki, and Toshihiko Yamasaki. Fast volume seam carving with multi-pass dynamic programming. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1818–1822. IEEE, 2016.

[7] M Grundmann, V Kwatra, Mei Han, and I Essa. Discontinuous seam-carving for video retargeting. In *Computer Vision and Pattern Recognition*, pages 569–576, 2010.

[8] B. Li, L. Y. Duan, J. Wang, R. Ji, C. W. Lin, and W. Gao. Spatiotemporal grid flow for video retargeting. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 23(4):1615–28, 2014.

[9] Shih-Syun Lin, Chao-Hung Lin, Yu-Hsuan Kuo, and Tong-Yee Lee. Consistent volumetric warping using floating boundaries for stereoscopic video retargeting. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(5):801–813, 2016.

[10] Shih Syun Lin, Chao Hung Lin, Icheng Yeh, Shu Huai Chang, Chih Kuo Yeh, and Tong Yee Lee. Content-aware video retargeting using object-preserving warping. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1677–1686, 2013.

[11] Saulo A. F. Oliveira and Ajalmar R. Rocha Neto. An improved genetic algorithms-based seam carving method. In *Computational Intelligence*, pages 1–6, 2016.

[12] Jian Qin, Jun Chen, Zheng Wang, Jiyang Zhang, Xinyuan Yu, Chunjie Zhang, and Qi Zheng. *User-Driven Sports Video Customization System for Mobile Devices*. Springer International Publishing, 2015.

[13] Zhan Qu, Jinqiao Wang, Min Xu, and Hanqing Lu. Context-aware video retargeting via graph model. *IEEE Transactions on Multimedia*, 15(7):1677–1687, 2013.

[14] Weijian Ruan, Jun Chen, Chao Liang, Yi Wu, and Ruimin Hu. Object tracking via online trajectory optimization with multi-feature fusion. In *ICME*, pages 1231–1236. IEEE, 2017.

[15] Michael Rubinstein, Ariel Shamir, and Shai Avidan. Improved seam carving for video retargeting. *Acm Transactions on Graphics*, 27(3):16, 2008.

[16] Liang Shi, Jinqiao Wang, Lingyu Duan, and Hanqing Lu. Sports video retargeting. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 981–982. ACM, 2009.

[17] Yu Shuen Wang, Hui Chih Lin, Olga Sorkine, and Tong Yee Lee. Motion-based video retargeting with optimized crop-and-warp. *Acm Transactions on Graphics*, 29(4):1–9, 2010.

[18] Yushuen Wang, Jenhung Hsiao, Olga Sorkine, and Tongyee Lee. Scalable and coherent video resizing with per-frame optimization. *international conference on computer graphics and interactive techniques*, 30(4):88, 2011.

[19] Zheng Wang, Ruimin Hu, Junjun Jiang, Junjun Jiang, Chao Liang, and Jinqiao Wang. Scale-adaptive low-resolution person re-identification via learning a discriminating surface. In *International Joint Conference on Artificial Intelligence*, pages 2669–2675, 2016.

[20] Bo Yan, Kairan Sun, and Liu Liu. Matching-area-based seam carving for video retargeting. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(2):302–310, 2013.

[21] Tzu Chieh Yen, Chia Ming Tsai, and Chia Wen Lin. Maintaining temporal coherence in video retargeting using mosaic-guided scaling. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 20(8):2339, 2011.