Computational Visual Media
DOI 10.1007/s41095-xxx-xxxx-x Vol. x, No. x, month year, xx–xx

Article Type (Research/Review)

Rectangling Irregular Videos by Optimizing Spatio-Temporal Warping

Jin-Liang Wu[1], **Jun-Jie Shi**[2], **and Lei Zhang**[2] ✉

Abstract The image and video processing based on geometric principles usually changes the rectangular shape of video frames to be irregular. This paper presents a warping based approach for rectangling the irregular frame boundaries in space and time. To reduce geometric distortion in the rectangling process, we employ content-preserving deformations of mesh grid as well as line structures as constraints in warping the frames. To conform to the original inter-frame motion, we keep feature trajectory distribution as constraints in motion compensation to ensure the stabilization after warping the frames. Such spatial and temporal optimization of warpings enables the output of regular rectangular boundaries of the video frames with less geometric distortion and motion shakiness. The experiments demonstrate that our approach can generate plausible video rectangling results in a variety of applications.

Keywords rectangling, warping, content-preserving.

1 Introduction

In recent years, geometry techniques have been widely used in image and video processing, such as image resizing or retargetting [30], perspective editing [7, 31], video stabilization [19], panoramic stitching [18, 21, 36–38], *etc*. Unlike traditional image and video processing based on pixel operation, geometry-driven methods typically adopt mesh grid structure in the image plane and manipulate the grid points to drive the processing of the enclosed patches, which usually enables more flexible and coherent processing on the image and video content [12]. Generally, these methods achieve the desired balance between efficiency and effectiveness, but they usually have to warp the boundary shape of images or video frames when directly operating grid points, i.e., changing the rectangular shape to be the one with irregular boundaries. Figure 1 shows two examples that have irregular boundaries generated in the processing of fish-eye video correction and panoramic video stitching. Because most display screens have rectangular resolutions, it is necessary to rectify the images and videos with irregular boundaries back to rectangular boundaries for normal displaying on the common screens. This paper mainly addresses the problem of rectangling the video frames having irregular boundaries.

Obviously, the most direct solution to video rectangling is to crop a rectangular part from the input video, but it will lose a lot of information of video. Some other methods use image and video completion or inpainting methods to fill the gaps between irregular boundaries of frames and the potential rectangles [10, 14, 17, 33]. However, existing image and video completion methods are too brittle for dealing with irregular videos in general, especially for the fish-eye or stitching images videos with a large field of vision, which are prone to disturbing artifacts after completion (see Figure 1).

He *et al*. propose to use image warping for rectangling panoramic images that have irregular boundaries [9]. Such method can well fill the gaps along the irregular boundaries by warping the whole image with content preservation, which is able to generate more natural transitions than image completion methods. This rectangling strategy has also been adopted in processing stereoscopic panorama like the work in [35, 37], which has less consideration on the temporal consistency like feature correspondence and motion. In this paper, we further extend this method to irregular videos by rectangling the frames in spatially and temporally coherent manner. More importantly, our approach can not only deal with panoramic or stitching videos, but also irregular videos from other video processing tasks like perspective editing of fish-eye videos.

The main contribution of our work lies in a warping based approach for rectangling irregular videos. Especially,

1 The 54th Research Institute of CETC, Shijiazhuang, 050050, China.

2 School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China. E-mail: leizhang@bit.edu.cn.

Manuscript received: 2014-12-31; accepted: 2015-01-30.

Fig. 1 Samples of irregular videos. **Top:** shape-corrected editing of fish-eye video frame by [31]. **Bottom:** panorama video frame by [21]. The third column shows the rectangling results by our approach. The fourth column shows the results by video completion method.

the motion-aware deformaton of underlying mesh grid through frames is constrained and optimized in space and time with line struture preservation. This can mitigate the shape distortion in rectangling the irregular videos significantly. The experiments demonstrate the effectiveness and efficiency of our approach in dealing with a variety of irregular videos generated in different applications.

2 Related work

In this section, we briefly review the most related works to our method, which refers to the aspects of image completion,video completion and warping.

(i) Image and video completion. Image completion methods can be employed to fill the holes along the irregular boundaries for rectangling. Generally, these methods are broadly classified into three categories: statistical-based, partial differential equation (PDE) based and exemplar-based methods. Statistical-based methods are mostly used in texture synthesis, where the statistical models like histograms [11], wavelet coefficients [23], *etc.*, are employed to describe the color or structure of the images and fill the holes. But these methods are only good at filling the holes with naive textures. PDE-based methods use a diffusion process to propagate the hole boundary information to the interior regions. The diffusion process is usually described by Laplacian equation [2], Poisson equation [22] or Navier-Stokes equation [1], which are not suitable for processing large holes in the images. Exemplar-based methods borrow some compatible patches from the input image itself, and fill the holes by aligning the patches with appearance consistency between neighboring patches [5, 6, 16]. Generally, exemplar-based methods are more capable of generating high-quality completion results than statistical-based and PDE-based methods, but might incur semantic ambiguity especially for natural images.

It is observed that introducing some semantic cues in exemplar-based image completion is very helpful for obtaining plausible results. Such cues can be specified by user interaction [20, 27] or explored from some large-scale dataset like the Internet images [8, 28, 29, 39]. Then exemplar patches are selected and aligned to obey the cues such that the completed holes can well match the context of the entire images.

The above image completion methods can be directly applied in video completion by filling the holes in each frame individually. However, there might be temporal inconsistency between successive frames, especially for the videos with dynamic scenes. To obtain good completion in space and time, motion information needs to be considered in filling the holes of different frames. Jia et al. [14] uses motion tracking to impose the consistent fragments to fill the regions at the same positions in neighboring frames, which is able to generate visually smooth completion results. Alternatively, motion field can also be used when selecting the candidate patches to fill the holes [25, 33]. Although these methods demonstrate good behavior in filling small holes, there are still failure results for processing videos with more complex scenes.

(ii) Image and video warping. Recent image warping methods typically use geometric principles by deforming an embedded mesh grid for the target shape, which are then drive the change of the image content. To keep the original video content, the warping is usually required to be shape-preserving [13, 24], which have been widely used in many applications like image resizing [15], perspective editing [4, 7], *etc.*. The image warping can also be used for rectangling images with irregular boundaries [9], which provides the effect of image completion, and is extented to

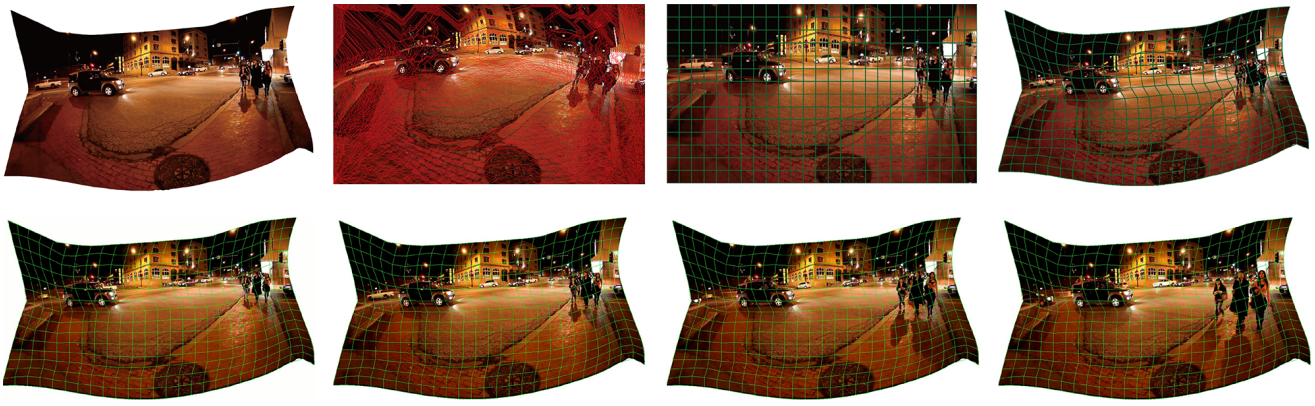

Fig. 2 **Top:** Per-frame quad mesh placement by the method of [9]. The red lines are the inserted seams, and green lines are mesh edges. **Bottom:** The consistent quad mesh grid through four successive frames (the 7th∼10th frames).

process the steroscopic panorama [35, 37]. Actually, the method of [37] claims the first one to deal with rectangling steroscopic stitching images, while the method of [35] extends it to steoscopic panorama vidos. But these two methods focus on the disparity preservation, which lack explict constraints on the motion consistency as well as line structure correspondence between frames, i.e., preserving the original motion in the warping. This might still generate the distortion after warping frames. For video warping, the temporal consistency should be considered in warping each frame, which has been used in video stabilization [19], fish-eye video correction [31] and video stitching [21].

In this paper, we extend the image rectangling method of [9] to process irregular videos based on the spatio-temporal optimization of the frame warpings, which purports to 'fill' the holes between the irregular boundaries and rectangles in the video frames for the rectangling. Especially to ensure the spatio-temporal consistency, we propose the use of line structure preservation and motion preservation between adjacent frames for rectangling video frames. Here, the line preservation enables common line orientation by line matching between adjacent frames, while the motion preservation avoids the motion jitter in warping the original frames. These are the key ingredients of our method that differs from the single image rectangling.

3 Building spatially and temporally consistent grids

Our warping based frames rectangling approach requires a consistent mesh grid structure through all the frames, i.e., placing a mesh grid in each frame that has the same number of grid points and connectivity topology, such that it can drive the frame warping with spatial and temporal coherence. Here, we use the quad mesh to build the mesh grid structure and drive the warping of frames.

Because the input video has irregular frame boundaries, we have to construct a virtual domain fit to a rectangle, in which the mesh vertices can be correctly placed and used to embed the frame content. We adopt the mesh placement scheme in [9] to set up the quad mesh in each frame. Concretely, a set of extra seams are inserted into the irregular frames to construct a virtual rectangular domain (see Figure 2 **top**). Then, the method employs the local warping technique to deform the rectangular domain, where the quad mesh is deployed and warped back to the original irregular frame. This procedure can guarantee a valid quad mesh displacement in the irregular frame. For an input video $\mathcal{F} = \{I^t\}_{t=1}^T$, we denote the quad mesh in the t-th frame as $\mathcal{M}^t = \{\mathcal{V}^t, \mathcal{E}^t\}$, where $\mathcal{V}^t = \{V_{i,j}^t\}$ is the set of mesh vertices and $\mathcal{E}^t = \{< V_{i,j}^t, V_{i\pm1,j\pm1}^t >\}$ is the set of mesh edges in the t-th frame.

We assume the frames having fixed boundaries in the temporal sequence in this paper. Then, the mesh edges $\mathcal{E}^t$ commit a common topology through the frames by their connectivities, i.e., the corresponding vertices sharing the same edge connections. However, the temporal difference between adjacent frames possibly makes the corresponding vertex positions suffering slight movements, i.e., $V_{i,j}^t \neq V_{i,j}^{t+1}$. To enable a consistent grid structure between adjacent frames, we need to further rectify the vertex positions to build a unified grid, which defines each of its vertices to have the same position through the frames. Here, we use the average mesh vertices of adjacent frames to construct the unified grid, i.e., $\sum_{t=1}^T V_{i,j}^t/T$, where T is the total number of video frames. Then, the corresponding vertices have the same positions in the space.

Consequently, by carefully setting the grid size, we can obtain a valid mesh deployed over all the frames (see

Figure 2 **bottom**). In the following sections, we still use $\{V_{i,j}^t\}$ to denote the unified mesh vertices in every frame I^t, by which the irregular boundaries are warped to the rectangular shape.

4 Motion-aware content-preserving frame rectangling

With the embedded grid, we next find the optimal image warps that deform each frame to a rectangle as well as preserving the content of the input video. We denote the mesh vertices in the deformed frame as $\hat{\mathcal{V}}^t = \{\hat{V}_{i,j}^t\}$, of which their positions determine the image appearance and distortion after rectangling irregular frames. In our setup, the preferred content-preserving frames warping refers to two aspects: i) deformation on the frame with less geometric distortion; ii) keep the original video motion with less shakiness after frames warping. Hence, we propose a novel energy function to solve an optimal motion-aware content-preserving frame warping for rectangling the irregular boundaries. In the sequel, we elaborate the details of the energy function terms and its optimization.

4.1 Energy function

The overall energy function of our approach contains the following four terms to realise the desired frame warping towards rectangling the boundaries in a spatially and temporally consistent manner.

Shape preservation. To preserve the local content of the original frame, we require the warping to induce as less geometric distortion as possible after rectangling.

Here, we follow the idea of as-similar-as-possible transformation, and define the shape preservation term as:

$$E_s^t(\hat{\mathcal{V}}^t) = \sum_{p \in Q_{i,j}} \|\hat{V}_{i,j}^t - (\hat{V}_{i,j+1}^t + u_p(\hat{V}_{i+1,j+1}^t - \hat{V}_{i,j+1}^t) + v_p \mathrm{R}_{90}(\hat{V}_{i+1,j+1}^t - \hat{V}_{i,j+1}^t))\|^2 \quad (1)$$

where u_p and v_p are scales in the local coordinates system of quad $Q_{i,j}$ enclosed by $\{V_{i,j}^t, V_{i+1,j}^t, V_{i+1,j+1}^t, V_{i,j+1}^t\}$ in the original frame, and R_{90} is a 2×2 anti-clock rotation matrix, defined by $\mathrm{R}_{90} = [0, -1; 1, 0]$ (see Figure 3). The derivation of Equation (1) can be found in [19], which ensures a similarity transformation on the frame to have the minimal geometric distortion.

Line preservation. As human visual system is much sensitive to the lines structure, we also add the line preservation term as in [9] to keep the orientation of the lines after warping. But our approach differs with theirs by considering the line correspondence between adjacent frames, such that the corresponding lines of different frames should be warped in a temporally consistent manner.

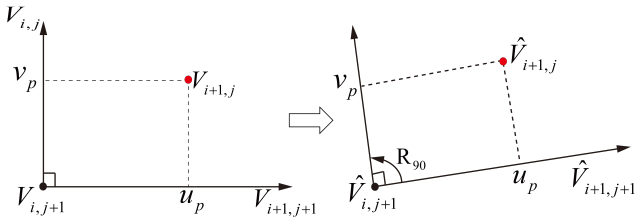

Fig. 3 Shape preserving by as-similar-as-possible transformation with the same local coordinates.

Concretely, we first detect the corresponding lines across multiple frames by using line matching techniques like [32] (see Figure 4). We denote the lines as $L_h = \{l_h^t\}$, where h is the line index at the t-th frame. Each line of L_h should have the same orientation after frame warping. The line orientation range $[-\frac{\pi}{2}, \frac{\pi}{2})$ is quantized into $M = 50$ bins, and designate the lines to the corresponding bins to obtain $\{\theta_m\}_{m=1}^M$. Here, the corresponding lines between adjacent frames should share the common orientation angle in warping each frame individually, which is determined by the line matching as shown in Figure 4. This enables the lines to preserve their common orientation after the frame warping. Hence, the line preservation term is defined as

$$E_l^t(\hat{V}, \{\theta_m\}) = \frac{1}{N_L^t} \sum_j \|C_j(\theta_{m(j)}) \mathbf{e}_{q(j)}\|^2 \quad (2)$$

where N_L^t is the total number of lines in the t-th frame, $q(j)$ indicates the quad containing the line segment, $\mathbf{e}_{q(j)}$ is the difference vector of the end points of the line segment, and C_j is the rotation matrix corresponding to the orientation angle $\theta_{m(j)}$.

Fig. 4 The lines structure and matching between two successive frames. The corresponding lines are denoted by the same indices.

Motion preservation. The process of warping frames inevitably causes motion jitter by changing the original inter-frame motion inconsistently. Hence, we have to introduce a motion preservation term that follows the original motion as much as possible, which is a major difference from the method of [9]. Here, we use the trajectories detected by the pyramidal Lucas-Kanade tracking methods like [3, 26] to collect the feature points and represent the motion based on the corresponding trajectories that starts at the t-th frame and ends at the $t+s$-th frame, denoted by $T_j = \{P_j^k\}_{k=t}^{t+s}$. Here, P_j^k indicates the feature point of the j-th trajectory in the k-th frame. Then, we want the inter-frame transformation of feature points to be rigid, such that motion structure can be preserved after frame warping. Hence, we have the following motion preservation term based on constraining the configuration of feature points:

$$\begin{aligned} E_m^k(\hat{V}) &= \sum_j ||(P_j^k - P_j^{k-1}) - \mathrm{R}_t^k \cdot (\hat{P}_j^k - \hat{P}_j^{k-1})||^2 \\ &= \sum_j ||(P_j^k - P_j^{k-1}) - \mathrm{R}_t^k \cdot (\mathcal{B}_j^k(\hat{V}^k) - \mathcal{B}_j^{k-1}(\hat{V}^{k-1}))||^2 \end{aligned} \tag{3}$$

where $\mathcal{B}_j^k(\cdot)$ is the bilinear interpolation operator to represent each trajectory point enclosed by a quad with its four vertices, and R_t is a rotation matrix that preserves the relative positions of feature points. Intuitively, the motion preservation term imposes a rigid motion structure when warping the frames, such that the inter-frame motion can follow the original motion to avoid the jitter.

Boundary constraints. The aim of frame warping is to obtain a rectangular boundary for each frame. So we have to add the boundary constraints on the transformed vertices, which has the same form as in [9]. Let $\hat{V}_i^B = (x_i, y_i)$ be the vertices along the boundary of the rectangular frame, then we have

$$E_B(\hat{V}) = \sum_{\hat{V}_i \in L} x_i^2 + \sum_{\hat{V}_i \in R} (x_i - w)^2 + \sum_{\hat{V}_i \in T} y_i^2 + \sum_{\hat{V}_i \in B} (y_i - h)^2 \tag{4}$$

where $\{L, R, T, B\}$ is the left, right, top and bottom boundary of the target rectangle, and w and h are the width and height of the rectangle.

With the above four terms, our frame rectangling energy function is defined as

$$E(\hat{V}) = E_S + \alpha \cdot E_L + \beta \cdot E_M + \gamma \cdot E_B \tag{5}$$

where α, β and γ are the weighting factors to control the trade-off among the terms. Typically, γ is set to be a large value for obtaining the rectangular boundary shape. In the experiments of this paper, we set $\alpha = 5$, $\beta = 10$ and $\gamma = 50$. Then, the warped frames are determined by changing the vertices to the new positions and driving the deformation of the corresponding quads, which makes the resultant videos having rectangular boundaries.

4.2 Optimization

The energy function of Equation (5) has a non-linear formulation with respect to its variants $\{V_{i,j}^t\}$, $\{\theta_{m(k)}\}$ and $\{\mathrm{R}_t^k\}$. These variants should be solved in a unified manner to obtain the optimal frame warpings, which are then used to drive the deformation of grid positions for the rectangular boundary.

Here, we resort to a two-step scheme with respect to $\{V_{i,j}^t\}$, and $\{\theta_{m(k)}\}$ and $\{\mathrm{R}_t^k\}$ separately, to optimize the energy function. Concretely, we iteratively compute the optimal solution of one variant by fixing the other variants as follows:

1. Fixing the values of line orientation $\{\theta_{m(k)}\}$, Equation (5) becomes a quadratic function with respect to $\{V_{i,j}^t\}$. Then, we compute its normal equation by setting the gradient to be zero.

2. With the computed values of mesh vertices $\{V_{i,j}^t\}$, we update the line orientation by computing the new $\{\theta_{m(k)}\}$ and rotation matrix R. The best line orientation $\{\theta_{m(k)}\}$ can be computed by optimizing Equation (2) with iterative Newton's method. The best rotation matrix R can be computed by singular value decomposition (SVD) of Equation (3).

The above two steps can be iteratively performed until the change of positions of mesh vertices below a prescribed threshold. Finally, we obtain the new vertex positions which change the frames to be with the rectangular shape.

5 Experiments

We have implemented our algorithm and tested it on a variety of irregular videos that have the nearly fixed boundaries through the frames sequence. The purpose of these tests is to verify the effectiveness of our algorithm, especially for the spatio-temporal consistency between frames. The testing videos are produced by perspective editing on the fish-eye videos [7], panoramic stitching of videos captured by multiple unstructured cameras [21], *etc.*, which usually have irregular boundary shapes that need to be rectangled for display adaption (see the examples in Figure 1, Figure 6 and Figure 7).

All the experiments were performed based on a machine with an Intel Core i5-2400 3.1GHz CPU and 8G RAM. Next, we show some results on rectangling irregular videos and comparison with other methods. In our first two examples, we adopt distortion-corrected fish-eye videos

Fig. 5 Rectangling results by frame-by-frame rectangling (**top**) and our algorithm (**bottom**). Visual flicking occurs in the frame-by-frame rectangling results, especially for the car enclosed by green ellipse.

as input, then compare the video completion method and naive cropping method with our approach. In the second two examples, the inputs are the panoramic videos that are from stitching multiple videos. Both have irregular boundaries that need to be rectified for the boundary shapeo of the rectangle. We encourage the readers to watch the accompanying demo video, which includes dynamic exhibition of rectangling results by our algorithm and comparison with other methods.

5.1 Rectangling results

The irregular boundaries of Figure 6 are generated in editing the fish-eye videos for correcting the spherical distortion, which are usually smooth within the irregular frames. Our approach can well recover the regular rectangle shape as well as preserving the structure, especially salient lines, after rectangling the boundaries.

The irregular boundaries of Figure 7 are generated by stitching regular videos captured by multiple cameras, which usually consist of piecewise straight lines in the panoramic videos. In this case, our approach can also align the irregular boundaries to the rectangle without distortion of the structure especially near the boundaries. Our approach usually takes about 2.5 seconds to process one frame, which involves 4 iterations to obtain good rectangling results. Most of the computation time is consumed in quad mesh placement and iterative optimization for solving Equation (5).

The use of spatially and temporally optimized warping achieves good temporal consistency in the transition of adjacent frames, which avoids visual flicking if we simply use the frame-by-frame rectangling based on the method of [9] (see Figure 5). It can be seen that the frame-by-frame rectangling results have obvious flicking, while our algorithm enables temporally consistent rectangling results.

5.2 Comparison with other rectangling methods

To demonstrate the superiority of our approach, we also compared our results with the ones obtained by some other video rectangling methods like disparity-preserving image rectangularization (DPR) [35], the classical video completion method [33] and the naive cropping method. Here for the DPR method, we ignore the disparity constraint in the warping energy terms. Because DPR has less constraints on the line structure as well as line correspondence between frames, it might generate distortion especially as shown in Figure 6. The video completion method of [33] is also able to generate regular frames with rectangle shape by filling the gaps. Typically, this kind of methods attempt to find a set of patches or volumetric blocks from the video itself to cover the gaps with spatial-temporal coherence on color and structure. Therefore, the rectangled video possibly suffers block repetition in the gaps, especially for regions with salient structure (see Figure 6(a) and 7(a)). The naive cropping is simple to realise, and can avoid distortion in rectangling the irregular boundaries. But it usually incurs over-cropping especially for videos with very irregular boundaries, such that there will be loss of visual information.

On the contrary, our approach directly deforms the frames to the rectangular boundaries with both spatial and temporal coherence that enables smooth transition between successive frames. From the results of Figure 6 and 7, it can be seen that the irregular boundaries are well rectangled with as shape-preserving as possible appearance, which provides appealing results after frames rectangling.

Evaluation. Actually, there is no standard to evaluate the performance of different rectangling methods. All

Fig. 6 Comparison of rectangling distortion-corrected fish-eye videos. From top to bottom: input frames of two examples; rectangling results by our approach, DPR [35], video completion method [33] and naive cropping results.

the above rectangling methods can generate the videos with rectangular boundaries. For the more comprehensive evaluation on the rectangling results by different methods, we conducted a user study to evaluate the visual quality as used in [37]. We asked 37 participants to grade the rectangling results by different methods within the score range $0 \sim 5$. We collected 7 examples as the cases in the user study, which had the suitable watching time for the participants that did not cause visual fatigue. Here, four cases are from the examples in Figure 6 and Figure 7, and the other three cases are from the examples in the supplementary file. Especially, we asked two questions in the evalution: (1) visual comfort when watching the rectangling videos and (2) free of artifacts that affect the perception of the video content. Here, the visual comfort contains the subjective feeling on the consistency of adjacent frames, which is important for the video rectangling results. We recorded the average scores of different methods as shown in Table 1. It can be seen that our method is able to generate the plausible results with better visual comfort and less artifacts.

5.3 Limitations and discussion

Although our approach enables useful effects, it is not without limitations. Our approach requires spatially and temporally consistent grids to drive the coherent frames warping. At present, we employ the simple average position of quad mesh for setting the grids. Although it is able to obtain the consistent mesh placement for videos with small motion between adjacent frames, it might fail in dealing with videos with large motion. In this case, we have to manually correct the position of quad mesh for the final consistent mesh placement. Besides, in the case of dynamic boundaries through the frames sequence, e.g., frames generated by using video stabilization methods, our approach may fail to place the expected grids, which possibly causes disturbing artifacts after frames rectangling. In this case, a potential solution is to borrow the cross parameterization technique from geometry processing

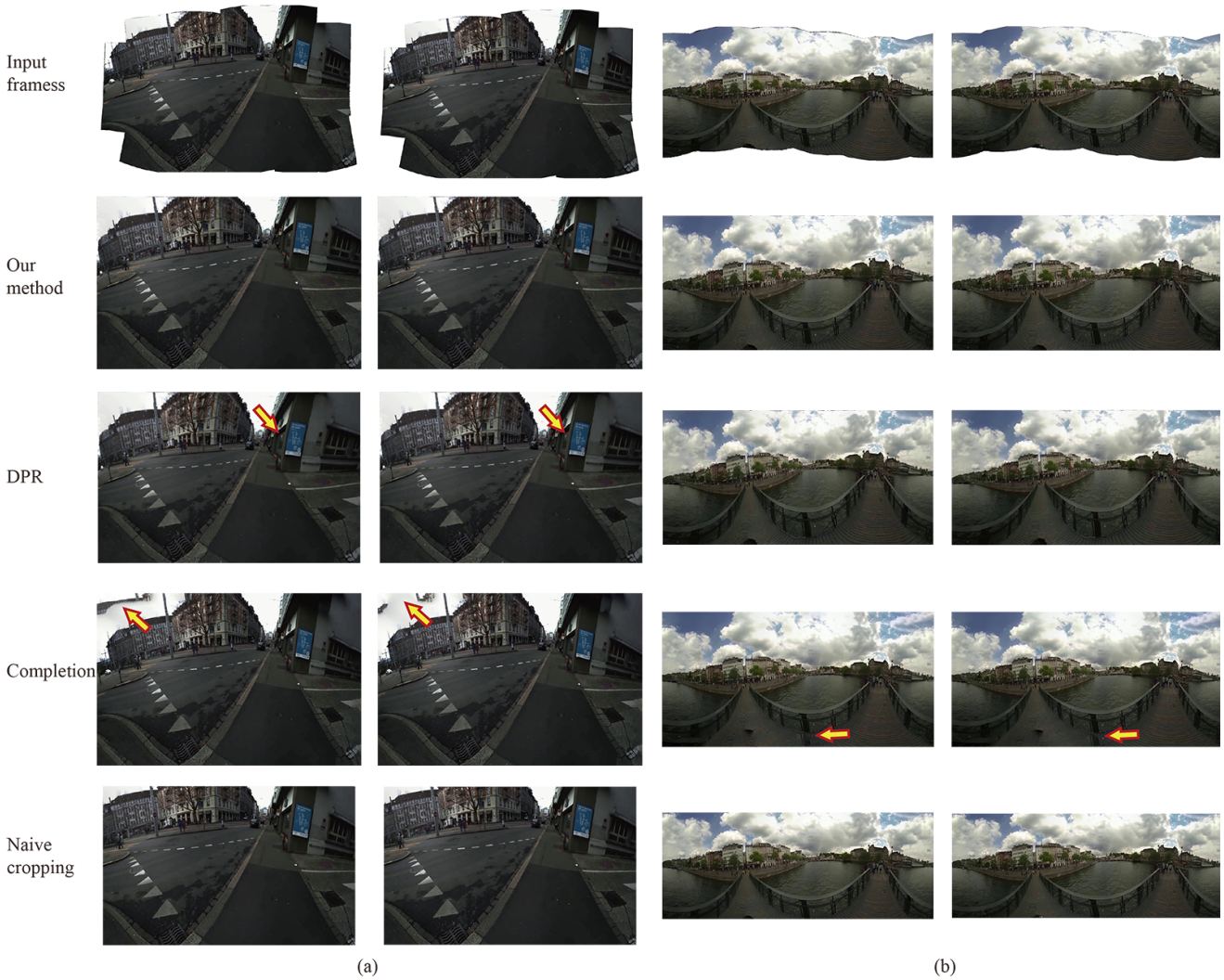

Fig. 7 Comparison of rectangling panoramic videos. From top to bottom: input frames of two examples; rectangling results by our approach, DPR [35], video completion method [33] and naive cropping results.

like [34], which implants extra vertices for consistent grid topology in space and time.

6 Conclusions

We have presented a warping based approach to transform the irregular video frames generated by geometry-based imge and video processing into the ones with rectangular boundaries. Our approach enables spatio-temporal consistence in warping frames towards rectangular boundaries due to the use of shape and motion preservation terms. The experiments demonstrate the efficacy of our approach for rectangling irregular videos.

As the future work, we plan to extend our approach to deal with videos with time-varying boundaries, e.g., videos generated by applying video stabilization algorithms. Besides, it is also promising to accelerate the computational speed of quad mesh placement and iterative optimization steps by using the parallel GPU programming technique.

References

[1] M. Bertalmío, A. L. Bertozzi, and G. Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. In *Proceedings of CVPR*, pages 355–362, 2001.

[2] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of SIGGRAPH*, pages 417–424, 2000.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of ECCV*, 2004.

[4] Robert Carroll, Aseem Agarwala, and Maneesh Agrawala. Image warps for artistic perspective manipulation. *ACM Trans. Graph.*, 29(4):127:1–127:9, July 2010.

Tab. 1 The user study on the rectangling results by our method, DPR [35], video completion method [33] and naive cropping. The numbers indicate the average scores for visual comfort and free of artifacts respectively.

	Fig. 6(a)	Fig. 6(b)	Fig. 7(a)	Fig. 7(b)	Suppl.1	Suppl.2	Suppl.3
Our method	4.61/4.79	4.81/4.78	4.51/4.71	4.53/4.56	4.64/4.72	4.22/4.25	4.57/4.62
DPR	3.65/4.01	3.99/4.13	4.25/4.03	4.14/4.11	4.02/4.09	4.19/4.21	4.11/4.15
Completion	1.43/1.41	3.12/3.43	1.23/1.50	2.52/2.85	2.11/2.21	3.56/3.62	1.89/1.93
Naive cropping	2.91/4.77	2.61/4.47	3.81/4.21	3.57/4.45	2.78/2.82	3.67/3.72	3.72/3.78

[5] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In *Proceedings of CVPR*, pages 721–728, 2003.

[6] Iddo Drori, Daniel Cohen-Or, and Hezy Yeshurun. Fragment-based image completion. *ACM Trans. Graph.*, 22(3):303–312, July 2003.

[7] Song-Pei Du, Shi-Min Hu, and Ralph R. Martin. Changing perspective in stereoscopic images. *IEEE Trans. Vis. Comput. Graphics*, 19(8):1288–1297, August 2013.

[8] James Hays and Alexei A. Efros. Scene completion using millions of photographs. *ACM Trans. Graph.*, 26(3), July 2007.

[9] Kai-Ming He, Huiwen Chang, and Jian Sun. Rectangling panoramic images via warping. *ACM Trans. Graph.*, 32(4):79:1–79:10, July 2013.

[10] Kai-Ming He and Jian Sun. Image completion approaches using the statistics of similar patches. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(12):2423–2435, December 2014.

[11] David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of SIGGRAPH*, pages 229–238, 1995.

[12] Shi-Min Hu, Tao Chen, Kun Xu, Ming-Ming Cheng, and Ralph R. Martin. Internet visual media processing: a survey with graphics and vision applications. *Vis. Comput.*, 29(5):393–405, March 2013.

[13] Takeo Igarashi, Tomer Moscovich, and John F. Hughes. As-rigid-as-possible shape manipulation. *ACM Trans. Graph.*, 24(3):1134–1141, July 2005.

[14] Yun-Tao Jia, Shi-Min Hu, and Ralph R. Martin. Video completion using tracking and fragment merging. *Vis. Comput.*, 21(8-10):601–610, August 2005.

[15] Z. Karni, D. Freedman, and C. Gotsman. Energy-based image deformation. *Comput. Graph. Forum*, 28(5):1257–1268, July 2009.

[16] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Trans. Image Process.*, 16(11):2649–2661, November 2007.

[17] Johannes Kopf, Wolf Kienzle, Steven Drucker, and Sing Bing Kang. Quality prediction for image completion. *ACM Trans. Graph.*, 31(6):131:1–131:8, November 2012.

[18] Anat Levin, Assaf Zomet, Shmuel Peleg, and Yair Weiss. Seamless image stitching in the gradient domain. In *Proceedings of ECCV*, pages 377–389, 2002.

[19] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. Content-preserving warps for 3d video stabilization. *ACM Trans. Graph.*, 28(3):44:1–44:9, July 2009.

[20] Darko Pavić, Volker Schönefeld, and Leif Kobbelt. Interactive image completion with perspective correction. *Vis. Comput.*, 22(9-11):671–681, September 2006.

[21] F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross. Panoramic video from unstructured camera arrays. *Comput. Graph. Forum*, 34(2):57–68, May 2015.

[22] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Trans. Graph.*, 22(3):313–318, July 2003.

[23] Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision*, 40(1):49–70, October 2000.

[24] Scott Schaefer, Travis McPhail, and Joe Warren. Image deformation using moving least squares. *ACM Trans. Graph.*, 25(3):533–540, July 2006.

[25] Takaaki Shiratori, Yasuyuki Matsushita, Sing-Bing Kang, and Xiao-Ou Tang. Video completion by motion field transfer. In *Proceedings of CVPR*, 2006.

[26] D. Sun, S. Roth, and M.J. Black. Secrets of optical flow estimation and their principles. In *Proceedings of CVPR*, pages 2432–2439, 2010.

[27] Jian Sun, Lu Yuan, Jiaya Jia, and Heung-Yeung Shum. Image completion with structure propagation. *ACM Trans. Graph.*, 24(3):861–868, July 2005.

[28] Miao Wang, Yu-Kun Lai, Yuan Liang, Ralph R. Martin, and Shi-Min Hu. Biggerpicture: Data-driven image extrapolation using graph matching. *ACM Trans. Graph.*, 33(6):173:1–173:13, November 2014.

[29] Miao Wang, Ariel Shamir, Guo-Ke Yang, Jin-Kin Lin, Guo-Wei Yang, Shao-Ping Lu, and Shi-Min Hu. Biggerselfie: Selfie video expansion with hand-held camera. *IEEE Trans. Image Process.*, 27(12):5854–5865, December 2018.

[30] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 27(5):118:1–118:8, December 2008.

[31] Jin Wei, Chen-Feng Li, Shi-Min Hu, Ralph R. Martin, and Chiew-Lan Tai. Fisheye video correction. *IEEE Trans. Vis. Comput. Graphics*, 18(10):1771–1783, October 2012.

[32] Tomas Werner and Andrew Zisserman. New techniques for automated architectural reconstruction from photographs. In *Proceedings of ECCV*, pages 541–555, 2002.

[33] Yonatan Wexler, Eli Shechtman, and Michal Irani. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):463–476, March 2007.

[34] Y. Xu, R.-J. Chen, C. Gotsman, and L.-G. Liu. Embedding a triangular graph within a given boundary. *Vis. Comput.*, 28(6):113:1–113:14, August 2011.

[35] I-Cheng Yeh, Shih-Syun Lin, Shuo-Tse Hung, and Tong-Yee Lee. Disparity-preserving image rectangularization for stereoscopic panorama. *Multimed. Tools Appl.*, 79(6):26123–26138, 2020.

[36] Fang-Lue Zhang, Connelly Barnes, Hao-Tian Zhang, Jun-Hong Zhao, and Gabriel Salas. Coherent video generation for multiple hand-held cameras with dynamic foreground. *Comput. Vis. Media*, 6:291–306, 2020.

[37] Yun Zhang, Yu-Kun Lai, and Fang-Lue Zhang. Stereoscopic image stitching with rectangular boundaries. *Vis. Comput.*, 35(6-8):823–835, 2019.

[38] Yun Zhang, Yu-Kun Lai, and Fang-Lue Zhang. Content-preserving image stitching with piecewise rectangular boundary constraints. *IEEE Trans. Vis. Comput. Graphics*, 2020.

[39] Z. Zhu, Hao-Zhi Huang, Zhi-Peng Tai, K. Xu, and Shi-Min Hu. Faithful completion of images of scenic landmarks using internet images. *IEEE Trans. Vis. Comput. Graphics*, 22(8):1945–1958, August 2016.

Jin-Liang Wu received the Ph.D. degrees in applied mathematics from Zhejiang University, Hangzhou, China, in 2012. He is a senior engineer in The 54th Research Institute of CETC, Shijiazhuang, China. His research interests include image and video processing, data analytics, artificial intelligence.

Jun-Jie Shi received the B.S. degree in computer science from the Beijing Institute of Technology, Beijing, China, in 2017, where he is pursuing the master's degree with the School of Computer Science. His research interest is image and video processing.

Lei Zhang received the B.S. and Ph.D. degrees in applied mathematics from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. He is a Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. His research interests include image and video processing, computer graphics.