Consistent Volumetric Warping Using Floating Boundaries for Stereoscopic Video Retargeting

Shih-Syun Lin, Chao-Hung Lin, *Member, IEEE*, Yu-Hsuan Kuo, and Tong-Yee Lee, *Senior Member, IEEE*

***Abstract*—The key to content-aware warping and cropping is adapting data to fit displays with various aspect ratios while preserving visually salient contents. Most previous studies achieve this objective by cropping insignificant contents near frame boundaries and consistently resizing frames through an optimization technique with various preservation constraints and fixed boundary conditions. These strategies significantly improve retargeting quality. However, warping under fixed boundary conditions may bound/limit the preservation of visually salient contents. Moreover, dynamic frame cropping and frame alignment may result in unnatural object/camera motions. In this study, a floating boundary with volumetric warping and object-aware cropping is proposed to address these problems. In the proposed scheme, visually salient objects in the space-time domain are deformed as rigidly and as consistently as possible by using information from matched objects and content-aware boundary constraints. The content-aware boundary constraints can retain visually salient contents in a fixed region with a desired resolution and aspect ratio, called critical region, during warping. Volumetric cropping with the fixed critical region is then performed to adjust stereoscopic videos to the desired aspect ratios. The strategies of warping and cropping using floating boundaries and spatiotemporal constraints enable our method to consistently preserve the temporal motions and spatial shapes of visually salient volumetric objects in the left and right videos as much as possible, thus leading to good content-aware retargeting. In addition, by considering shape, motion, and disparity preservation, the proposed scheme can be applied to various media, including images, stereoscopic images, videos, and stereoscopic videos. Qualitative and quantitative analyses on stereoscopic videos with diverse camera and considerable motions demonstrate a clear superiority of the proposed method over related methods in terms of retargeting quality.**

***Index Terms*—Content-aware media retargeting, mesh warping, cropping, optimization**

I. Introduction

CONTENT-AWARE retargeting has drawn increasing attention in the field of computer graphics during the last decade. This technique is applied to stereoscopic images, and recently, to stereoscopic videos because of the rapid development of stereoscopic equipment. The studies on stereoscopic video retargeting aims at preserving shapes and disparities of visually salient contents, and maintaining temporal coherence. However, these preservation objectives are difficult and sometimes impossible to achieve without generating distortions and artifacts [1]. When the spatial content of a left video frame is preserved, the corresponding content in the right video frame and in the different time frames suffer distortions caused by inconsistent transformation. Therefore, balancing the preservation requirements and avoiding spatial and temporal artifacts are still challenging problems in stereoscopic video retargeting.

Frame cropping and mesh warping are common techniques of content-aware retargeting. Frame cropping removes insignificant content near frame boundaries, and mesh warping optimizes the deformation between a source and a target data using various preservation constraints. Wang et al. [1] have proven that combining these two operations can improve retargeting quality. They warp frames through an optimization with hard boundary constraints, and then crop the warped frames to explicitly match the desired aspect ratio (Figure 1). In cropping, a critical region of the desired aspect ratio that contains significant contents is determined for each frame. Although significant contents can be preserved in this manner, the critical regions of different positions in frames may result in unnatural motions. To solve this problem, a volumetric warping and cropping method based on the concept of floating boundary is proposed wherein visually salient volumetric objects in the left and right videos are deformed as consistently and as rigidly as possible under content-aware boundary constraints. The strategies of volumetric warping and floating boundary can lead to consistent content preservation and unnatural object/camera motions avoidance.

The basic idea of the proposed method is using floating boundaries, which has not been studied before to the best of our knowledge, and using information of matched volumetric objects in warping and cropping. The information of matched volumetric objects in a pair of videos enables the generation of object significance maps and consistent preservation of visually salient objects. The floating boundaries, that is, deformation optimization using content-aware boundary constraints, allow preservation of spatial shapes and temporal motions of volumetric objects.

In the proposed method, the input stereoscopic video clip is segmented into several volumetric objects, and the corresponding objects in the left and right video clips are assigned with the same significance value in preprocessing. The positions of high-significance contents in each frame are detected by using a weighted principal component analysis and a thresholding-based filter. A soft weight is assigned to a boundary vertex based on the distance to the detected high-significance contents. The use of soft boundary weighting in warping keeps visually salient objects in a fixed critical region

S.-S. Lin, Y.-H. Kuo, and T.-Y. Lee are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan, R.O.C. (e-mail: catchylss@gmail.com, gma2127@gmail.com, tonylee@mail.ncku.edu.tw).

C.-H. Lin is with Department of Geomatics, National Cheng Kung University, Tainan 701, Taiwan, R.O.C. (e-mail: linhung@mail.ncku.edu.tw).

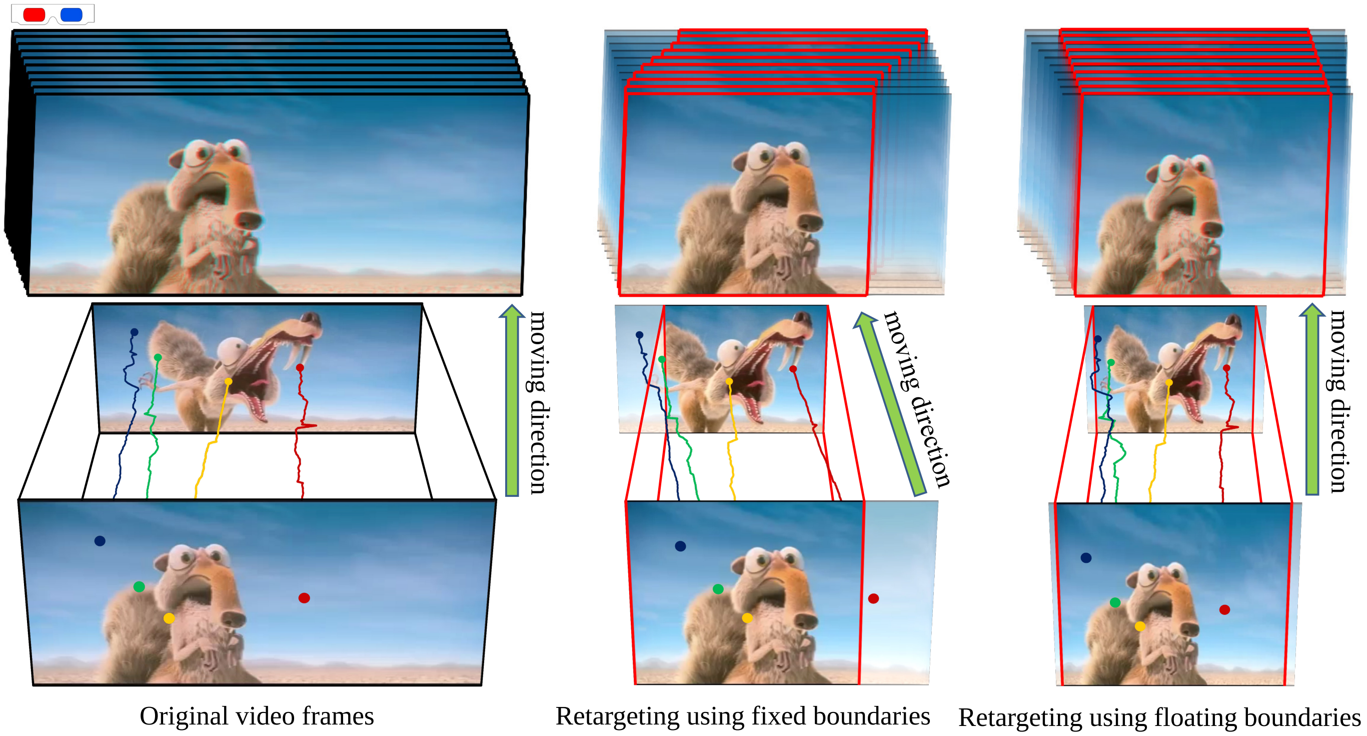

moving direction
moving direction
moving direction
Original video frames
Retargeting using fixed boundaries
Retargeting using floating boundaries

Fig. 1. Retargeting using floating boundaries and a fixed cropping window. Warping and cropping by using fixed boundaries and a dynamic cropping window may result in unnatural object/camera motions (middle). The proposed method which uses content-aware floating boundaries and a fixed cropping window can alleviate this problem (right). The cropping windows are marked by red quadrangles.

of the desired resolution and aspect ratio, which can potentially avoid unnatural temporal motions caused by warping and cropping that uses different critical regions in frames [1]. For example, in Figure 1, the temporal motion of the deformed foreground object is oblique and exhibits shifting effects, which is different from the original motion, when a dynamic cropping window is used. By contrast, the motion shifting effects are efficiently alleviated by using the proposed floating boundary and static cropping window.

II. Related Work

Numerous content-aware retargeting methods have been proposed recently. These methods can be classified into *cropping*, *seam carving*, and *warping* based on the characteristics of the adaptation algorithms. In cropping, a cropping window or critical region is determined, such that the amount of visually salient contents within the critical region is maximized [8] or the aesthetic value defined according to the stereoscopic photography principles is maximized [9] while the temporal coherence is preserved [1]. The advantage of cropping methods is the distortionless adaptation of the contents near the data center. In seam carving [10]–[15], a one-pixel width continuous or discontinuous seam line or surface with minimal significance is iteratively carved or inserted to reduce or enlarge the input data to the desired aspect ratio. This technique allows high flexibility in removing pixels, and thus, it can deal with data that contains many homogenous regions.

Compared with the methods that discretely remove seams or crop borders of an image/frame, warping-based methods that deform data using various constraints are potentially suitable for data containing dense information. Wang et al. [1] incorporate motion-aware constraints with the mesh warping to preserve visually salient motions in video retargeting. Consecutive frames are aligned by estimating inter-frame camera motion and by constraining the relative positions of the aligned frames to preserve temporal coherence and reduce waving artifacts. In addition, similar to the multi-operator retargeting technique [16], they integrate cropping and warping into the retargeting framework wherein the cropping removes temporally recurring contents and the warping utilizes available homogenous regions to absorb deformations from warping. In their later work [17], the scalability problem caused by global optimization over the entire space-time volume is solved without compromising resizing quality. Although these warping-based methods can provide good retargeting results for numerous cases, the recent study [18] reports that the object occupying several quads may suffer from inconsistent deformation. This problem may lead to apparent distortions, particularly of structure lines. Therefore, Lin et al. [18] propose an object-preserving warping technique that uses object motions instead of pixel motions in warping to address this problem. Similarly, Li et al. [19] propose a spatiotemporal grid flow that segments a video clip into spatiotemporal grids wherein the consistency of the content associated with a spatiotemporal grid is preserved during warping, and Yuan et al. [20] propose a volume-based metric and solve the video retargeting in graph representation.

Recently, the mesh warping techniques are applied to stereoscopic image retargeting [21]–[24]. The key to this retargeting is to preserve pixel disparities in addition to object shapes.

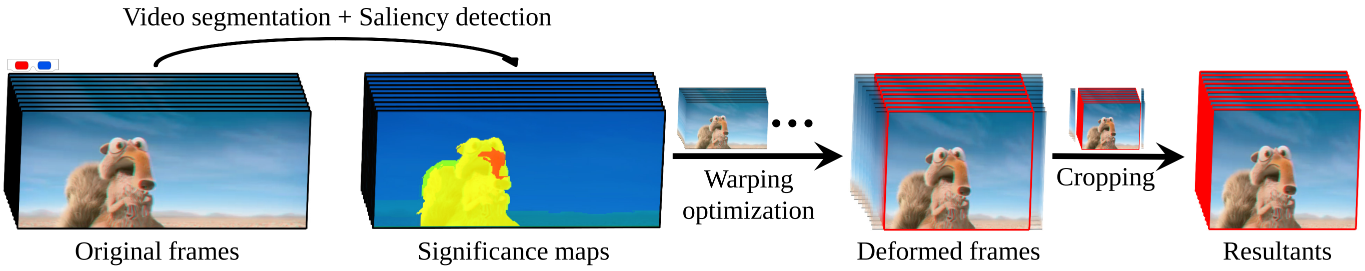


Fig. 2. Schematic workflow of the proposed approach. From left to right: original frames, significance map generation, volumetric warping, and cropping.

Based on the idea of mesh warping, the methods [21]–[24] optimize a mesh deformation with the aid of depth layer information, significance maps, and disparity maps. Therefore, both the disparities and shapes of high-significance objects can be preserved. Lang et al. [25] and Chang et al. [21] discuss the perceptual aspects of stereo vision and their applications in content manipulation. They provide a set of disparity mapping operators with a warping function to achieve desirable disparity distributions.

In the present study, the objectives of preserving visually salient contents, including shapes, motions, and disparities, are the same as those in [18], [23], and the idea of integrating warping and cropping in the retargeting scheme is the same as that in [17]. However, the proposed method has substantial differences from these methods. First, instead of using pixel coherence and fixed boundaries in warping, the uses of volumetric object coherence and floating boundaries allow the determination of a fixed critical region and the preservation of visually salient objects. Second, by using content-aware boundary constraints, the cropping window in frames is static rather than dynamic, which can efficiently alleviate unnatural temporal motions caused by the different positions of critical regions. Third, the proposed preservation constraints and significant map of volumetric objects can provide good shape, motion, and disparity preservation.

III. Methodology

Figure 2 shows the schematic workflow of the proposed method which consists of three main steps, namely, *significance map generation*, *volumetric warping*, and *cropping*. The basic idea is to resize stereoscopic videos by utilizing the information of volumetric objects. Therefore, the input stereoscopic video clip is firstly partitioned into several volumetric objects by using the video segmentation technique [26]. To realize content-aware retargeting, saliency detection [2] is adopted to evaluate pixel saliency in the left and right video clips. Significance measurement for segmented objects is then performed to generate significance maps. Each volumetric object is assigned a significance value to address the problem of inconsistent object deformation. In the next step, the proposed volumetric warping that uses floating boundaries is performed. A grid mesh is created to cover the left and right videos, and the volumetric objects are forced to undergo as-rigid-as-possible and as-consistent-as possible deformation by using various constraints, including spatiotemporal, disparity, and content-aware boundary constraints. During warping, the frame boundaries are floating (or unfixed) and high-significance objects are potentially retained in a region with the desired aspect ratio. Therefore, in the next step, a simple cropping is performed to remove the contents outside the determined critical region, meaning that the cropping window is fixed in frames. The generation of significance maps is described in Section III-A, and the volumetric warping and cropping which are described in Sections III-B and III-C, respectively.

A. *Significance map generation*

One of the basic ideas in our method is using volumetric object information in warping. This idea requires data segmentation and object-based significance measurement processes. In the proposed method, the frames of the left and right video clips are interleavingly placed and combined into a single video clip in order to segment them consistently. Following the approach in [23], the combined video clip is partitioned into several volumetric objects by using the hierarchical graph-based segmentation technique [26]. In this manner, the volumetric objects in the left and right video clips and the correspondences between them can be obtained. Each volumetric object is assigned with the average saliency values of the pixels within that object. The pixel saliency is estimated by the saliency detection approach [2]. Note that this study does not focus on the video segmentation and saliency detection. Any advanced method such as the recent segmentation method [27] and saliency detection method [7] can be adopted in the proposed scheme for better results.

With this object-based significance measurement, a high-significance object can be deformed as rigidly as possible during warping, and inconsistent object deformation reported in [28] can be efficiently alleviated. To further preserve significant contents, the saliencies of high-significance objects are enhanced, whereas those of low-significance objects that potentially belong to the background are suppressed. Such process is achieved by performing saliency enhancement with the information of object volumes as follows:

$$w_i = \begin{cases} \frac{s_i - \Omega_{min}}{s_{max} - \Omega_{min}} & , \text{ if } \quad s_i > \Omega_{min}; \\ 0.1 & , \text{ otherwise,} \end{cases} \tag{1}$$

where s_i is the significance value of the segmented object o_i, w_i is the enhanced value of s_i, and $w_i \in [0.1, 1]$; Ω represents the significance value normalized by the object volume (that is, the number of pixels within that object); and s_{max} and Ω_{min} denote the maximum of significance values and the minimum of normalized significance values, respectively. Figure

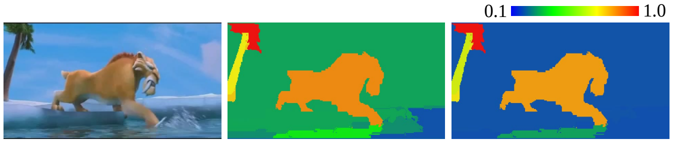


Fig. 3. Significance map generation. Left: input video frame; middle: significance map; right: enhanced significance map. Significance values are visualized by colors ranging from blue (lowest significance) to red (highest significance).

3 presents an example of stereoscopic video segmentation and significance map generation. The foreground objects and the background in the video clip are separated, and the object significance values are evaluated and enhanced by using Eq. (1). Compared with the unenhanced significance map, the foreground objects with the enhanced saliency values can be better preserved during warping.

B. *Volumetric warping using floating boundaries*

A uniform cubic grid mesh $\mathbf{M} = (\mathbf{Mesh}^L, \mathbf{Mesh}^R)$, which covers the input stereoscopic video, is created to be the control mesh in warping. This control mesh consists of two sub-meshes $\mathbf{Mesh}^{\{L,R\}} = \{\mathbf{V}^{\{L,R\}}, \mathbf{E}^{\{L,R\}}, \mathbf{Q}^{\{L,R\}}\}$ with vertex positions $\mathbf{V}$, edges $\mathbf{E}$, and quads $\mathbf{Q}$, to control warping of the left and right video clips. In addition, a set of segmented objects $\mathbf{O} = \{O_1, ..., O_{n_o}\}$ and their corresponding significance values $\{w_1, ..., w_{n_o}\}$ obtained in the preprocessing are used in warping, where n_o represents the number of segmented objects. To preserve the spatial shapes, disparities, and temporal motions of the stereoscopic video clips, three constraints, namely, *volumetric object preservation*, *disparity preservation*, and *floating boundary* are defined with an optimization solver.

Volumetric object preservation constraint. The proposed warping scheme aims to find a deformed grid mesh $\tilde{\mathbf{M}}$, in which the quads in a high-significance volumetric object are deformed as consistently and as rigidly as possible. To achieve this objective, two energy terms, namely, *object preservation* and *grid bending*, are defined. Following the energy terms in [18], the object preservation energy is defined as measuring the rigidity of a volumetric object as follows:

$$\psi_{Op}(\mathbf{M}) = \sum_{O_i \in \mathbf{O}} w_i \times \sum_{\mathbf{e}_j \in O_i} \left\| \tilde{\mathbf{e}}_j - \mathbf{T}_{ij} \tilde{\mathbf{C}}_i \right\|^2 , \tag{2}$$

where w_i is the enhanced significance value of object O_i; and $\tilde{\mathbf{e}}_j$ and $\tilde{\mathbf{C}}_i$ represent a deformed edge and the deformed representative edge of object O_i, respectively. The representative edge functions as a deformation pivot of the edges in an object. The edge closest to the center of the volumetric object is selected as the representative edge. $\mathbf{T}_{ij}$ is the similarity transformation between $\mathbf{e}_j$ and $\mathbf{C}_i$. This energy measures the changes in geometric relations of edges within a volumetric object. Thus, this energy can potentially avoid inconsistent object deformations.

The grid bending term is used to prevent the occurrence of skewed artifacts on the control mesh. Based on the measurement of quad orientation distortion proposed by Wang et al. [18], grid bending is defined as measuring the line bending of the cubic grid mesh, that is,

$$\begin{aligned} \psi_{Lb}(\mathbf{M}) = & \sum_{\{i,j\} \in \mathbf{E}_x} \left\| \tilde{v}_{i_y} - \tilde{v}_{j_y} \right\|^2 + \sum_{\{i,j\} \in \mathbf{E}_y} \| \tilde{v}_{i_x} - \tilde{v}_{j_x} \|^2 \\ & + \sum_{\{i,j\} \in \mathbf{E}_z} \| \tilde{v}_i - \tilde{v}_j \|^2 , \end{aligned} \tag{3}$$

where $\mathbf{E}_x$, $\mathbf{E}_y$, and $\mathbf{E}_z$ are the sets of x-, y-, and z-direction edges in the control mesh, respectively.

The total shape and motion preservation constraint is defined by summing up individual energy terms with weights as follows:

$$\psi_{VP} = \begin{cases} \alpha \times \psi_{Op} + (1-\alpha) \times \psi_{Lb} \text{ , for internal vertices;} \\ \psi_{Op} + k_{max} \times \psi_{Lb} \qquad \text{ , for boundary vertices,} \end{cases} \tag{4}$$

where α is the weighting factor for internal vertices, which controls the rigidity and consistency of object deformations. In the implementation, α is set to 0.5. The weighting factor k_{max} is assigned with a large value to force boundary edges/vertices to be straight during warping. In the implementation, k_{max} is set to 100. Note that these two terms are designed with volumetric objects in the space-time domain, and thus, the temporal motion preservation is considered in these terms.

Disparity preservation constraint. The disparity preservation constraint is used to preserve pixel disparities and avoid vertically shifting effects happened between the corresponding pixels in the left and right frames. The previous studies [21], [22], [24] define this constraint as the difference between the disparity values of the corresponding pixels in the original and deformed frames, which requires the information of disparity maps. The study [23] further considers consistent deformation, and disparity constraints are formulated according to the deformation consistency of objects in left and right frames. These constraints can efficiently preserve disparities. However, we observe that disparities can be roughly maintained without these advanced constraints in the proposed scheme because that a volumetric object in the left and right videos is assigned a significance value and deformed consistently during warping (see Eq. (2)). The volumetric and consistent warping in the proposed method reduces the needs of emphatic disparity constraints. Therefore, a simple constraint is adopted, which simply measures the distance between each pair of corresponding vertices in the left and right grid meshes as follows:

$$\psi_{DP}(\mathbf{M}) = \sum_{i=1}^{n_v} \left\| \tilde{v}_i^L - \tilde{v}_i^R \right\|^2 , \tag{5}$$

where $\tilde{v}_i^L$ and $\tilde{v}_i^R$ represent the deformed vertices in the left and right frames, respectively.

Boundary constraint. Previous warping-based methods [1] resize frames through an optimization with hard boundary constraints and then crop insignificant regions near frame boundaries. Similarly, in [17], soft boundary constraints are used in the first step of warping to search for a suitable resizing resolution. Hard boundary constraints are then used in the second step of warping to adapt data to explicitly match the determined aspect ratio, meaning that the frame boundaries are fixed during warping. After warping, a critical region of the target aspect ratio that contains significant contents is determined for each frame. Critical regions with different positions in frames are aligned by translating frames, and then the outer regions are discarded in cropping. Although the combination of warping and cropping significantly improves retargeting quality, the use of fixed boundaries may restrict the preservation of visually salient objects, and the frame alignment may result in frame panning and unnatural temporal motions, particularly for videos containing considerable object/camera motions.

In this study, floating boundary and static cropping window are applied in warping and cropping to alleviate the aforementioned problems. The basic idea is to retain high-significance contents in a region of the target aspect ratio by using content-aware boundary constraints during warping. Specifically, a soft weight is assigned to a boundary vertex based on the frame content. A boundary vertex that is close to (or far from) high-significance contents is assigned with a high (or low) weight. This boundary weighting scheme aims to keep high-significance contents in the desired critical region and push low-significance contents outside the critical region. Assume that the source stereoscopic video clip with $m \times n$ resolution is resized into a new clip with $m' \times n'$ resolution. The boundary vertex constraints are defined as follows:

$$\psi_{FB}(\mathbf{M}) = \begin{cases} \lambda_{t,i} \times \left\| \tilde{v}_{i_y} - 0 \right\|^2 & \text{, if } v_i \text{ is on the top boundary;} \\ \lambda_{b,i} \times \left\| \tilde{v}_{i_y} - m' \right\|^2 & \text{, if } v_i \text{ is on the bottom boundary;} \\ \lambda_{l,i} \times \| \tilde{v}_{i_x} - 0 \|^2 & \text{, if } v_i \text{ is on the left boundary;} \\ \lambda_{r,i} \times \| \tilde{v}_{i_x} - n' \|^2 & \text{, if } v_i \text{ is on the right boundary,} \end{cases} \tag{6}$$

where λ_i is the weight of boundary vertex v_i.

To calculate the boundary weights, the positions of high-significance contents in each frame are determined via a weighted principal component analysis (wPCA) and a thresholding-based filter. The filter is used to extract high-significance objects and the wPCA is utilize to obtain the overall distribution of significant contents. Each pixel in a frame is assigned with a weight to represent its significant contribution to wPCA and to estimate the geometric median (or weighted mean) which is regarded as a robust center of an arbitrary point set [29]. Given a pixel set $\mathbf{P} = \{(x_i, y_i)\}_{i=1}^{n_p}$, an efficient method to compute the principal components of the point set $\mathbf{P}$ is to diagonalize the covariance matrix of $\mathbf{P}$. In matrix form, the covariance matrix of $\mathbf{P}$ is written as follows:

$$\mathbf{C}(\mathbf{P}) = \frac{\sum_{p_i \in \mathbf{P}} w_i (p_i - \bar{p})(p_i - \bar{p})^{\mathrm{T}}}{\sum_i w_i}, \tag{7}$$

where $\bar{p}$ is the weighted mean that is defined as $\bar{p} = \sum w_i p_i / \sum w_i$, and w_i is the weight of pixel p_i. The pixel significance value described in Section III-A is set as the pixel weight. The low-significance pixels with significant values less than 0.2, that is, the pixels that potentially belong to the background, are excluded in the calculation of wPCA. The eigenvectors and eigenvalues of the covariance matrix are computed by using the matrix diagonalization technique, that is, $\mathbf{V}^{-1}\mathbf{C}\mathbf{V} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix with the eigenvalues of $\mathbf{C}$, and $\mathbf{V}$ is the orthogonal matrix with the corresponding eigenvectors. In geometry, eigenvalues and eigenvectors are related to an ellipse that represents the distribution of high-significance contents.

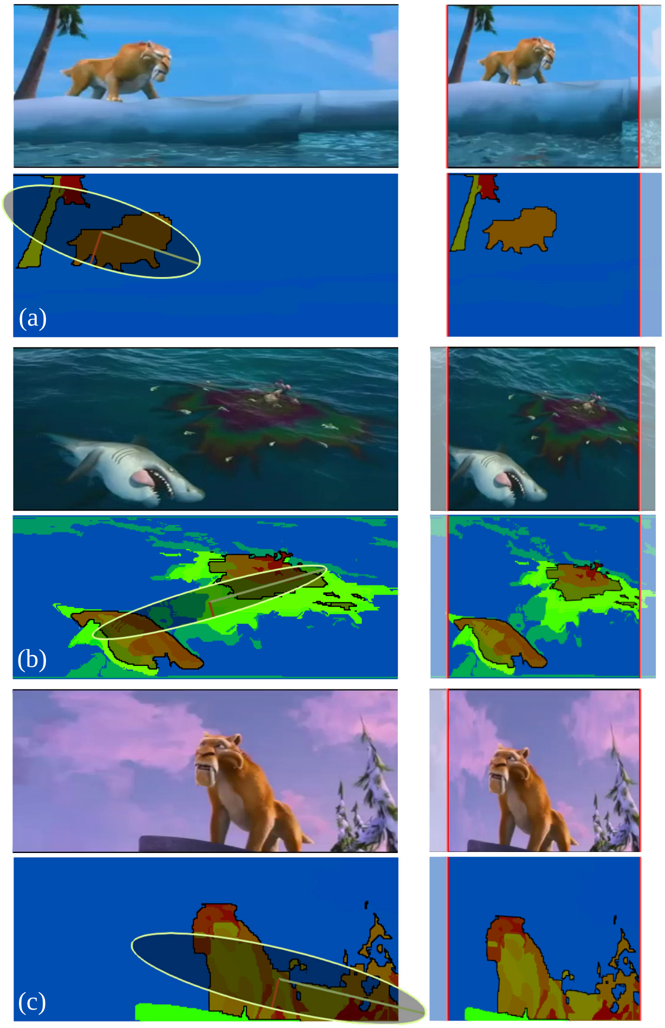


Fig. 4. Warping by using floating boundaries. Left: original video frames; right: retargeting results. The foreground objects in the left (a), middle (b), and right (c) of the video frames are tested. The ellipses represent significant content distributions, and the regions enclosed by black curves represent high-significance objects. The red lines represent cropping windows.

In filtering, the objects with significance values greater than a defined threshold T_s are extracted as the significant objects. T_s is a turnable parameter and the default value is 0.5 in the implementation. The weight of a boundary vertex is defined as a function of the minimal distance from the boundary vertex to the determined wPCA ellipse and the extracted significant objects, that is,

$$\lambda_i = \begin{cases} (c + 1/dist_i)^{\kappa/dist_i} & \text{, if } v_i \text{ lies outside the ellipse;} \\ \infty & \text{, if } v_i \text{ lies inside the ellipse,} \end{cases} \tag{8}$$

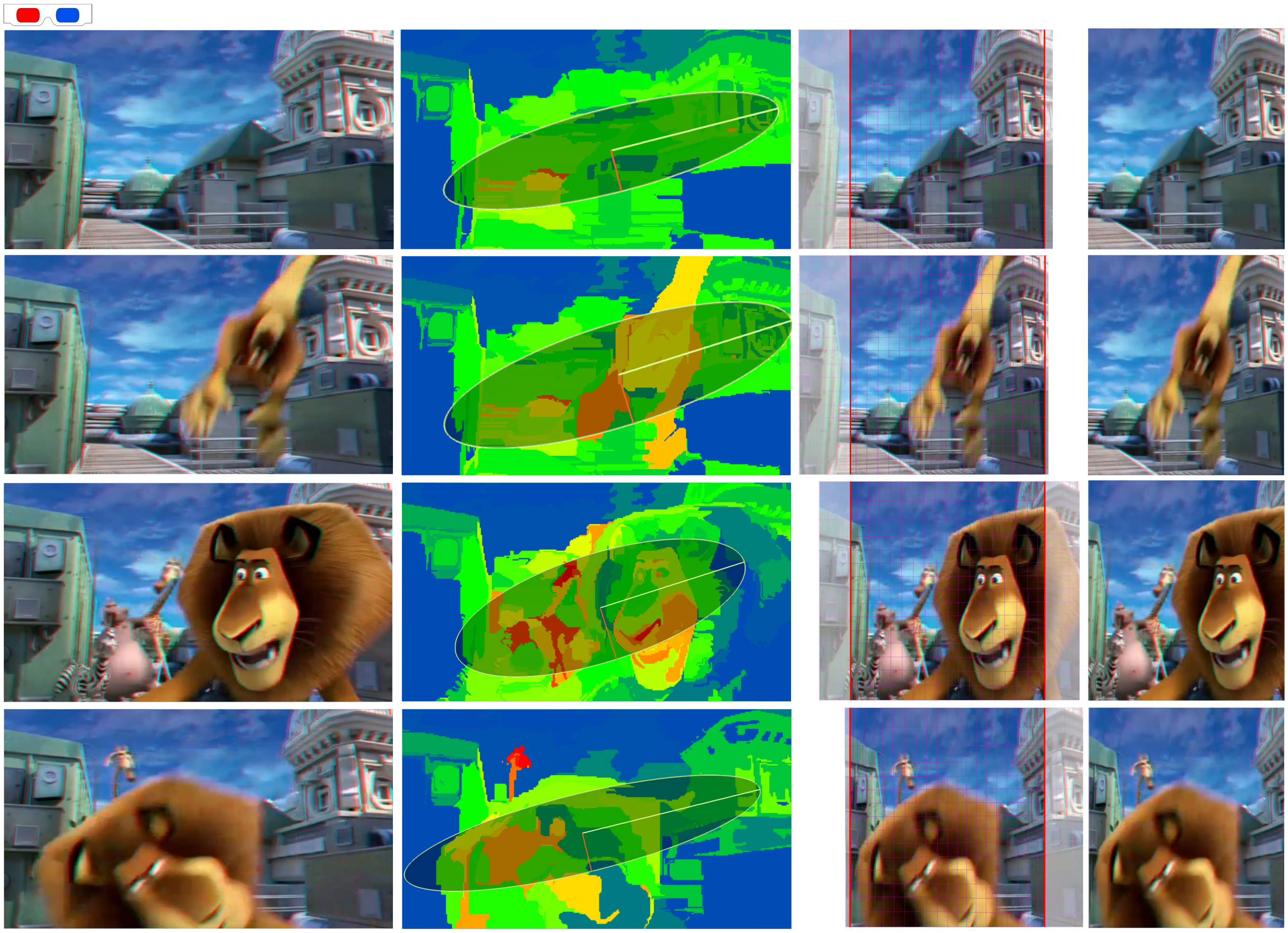


Fig. 5. Results of the proposed approaches. From left to right: original video frames, significance maps, content detection results, warping using floating boundaries (the critical regions are marked with red lines), and retargeting results.

where c is fixed constant and κ is a turnable parameter. In the implementation, $c = 2.7$ and $\kappa = 60$ based on our empirical observation. $dist_i$ represents the distance from boundary vertex v_i to the ellipse and detected significant objects. If boundary vertex v_i lies within the ellipse, an extremely large weight is assigned to this vertex to preserve the significant contents near the boundaries.

Figure 4 shows the warping results by using floating boundaries. Three cases are tested, in which the significant objects lie in the left, middle, and right of the video frames. The results show that the distribution of significant contents and the locations of high-significance objects are accurately detected. Therefore, these detected contents can be preserved during warping.

Optimization with floating boundary. By combining shape, disparity, and boundary constraints, the optimization for the content-aware deformation $\tilde{\mathbf{V}} = \{\tilde{\mathbf{V}}^L, \tilde{\mathbf{V}}^R\}$ is formulated as follows:

$$\arg\min_{\tilde{\mathbf{V}}} \quad (\psi_{VP} + \psi_{DP} + \psi_{FB}). \tag{9}$$

In the implementation, we fix the boundary vertices at the top and bottom for horizontal resizing, that is, an extremely large weight is assigned to λ_t and λ_b. Similarly, an extremely large weight is assigned to λ_l and λ_r for vertical resizing. In optimization, a least-squares linear system $\mathbf{A}\tilde{\mathbf{V}} = \mathbf{b}$ with a sparse designed matrix $\mathbf{A}$ is obtained from Eq. (9). This least-squares system has the optimal solution $\tilde{\mathbf{V}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}$, and thus, the deformed vertices of quad meshes $\tilde{\mathbf{M}}^L$ and $\tilde{\mathbf{M}}^R$ in the left and right video clips can be obtained. To further consider temporal coherence, a smoothing operation using Bzier curves is performed on the z-direction edges of the control mesh, that is, $\{i, j\} \in \mathbf{E}_z$, thus implying that temporal smoothing is applied to the control cubic mesh rather than the motion trajectories.

C. Cropping

Warping with content-aware boundary constraints maintains high-significance contents within the critical region of the desired aspect ratio. Therefore, the critical region determination and frame panning processes are avoided. We simply remove video clip contents outside the critical region during cropping.

IV. Experimental Results and Discussion

We implement and evaluate our method on a PC with 3.4 GHz quad-core CPU and 4 GB RAM. A stereoscopic

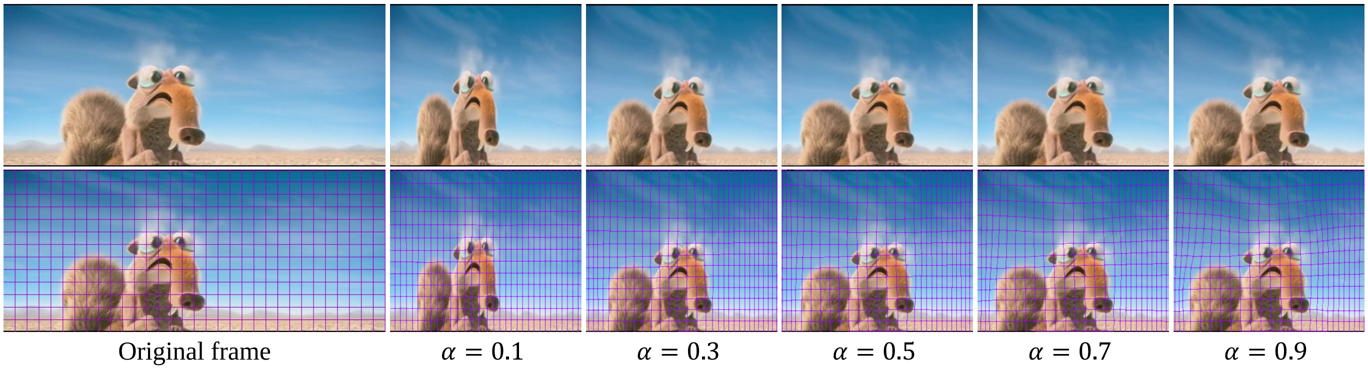


Fig. 6. Warping with different values of parameter α. From left to right: original frame; warping results with the parameter settings, $\alpha = 0.1, 0.3, 0.5, 0.7,$and 0.9.

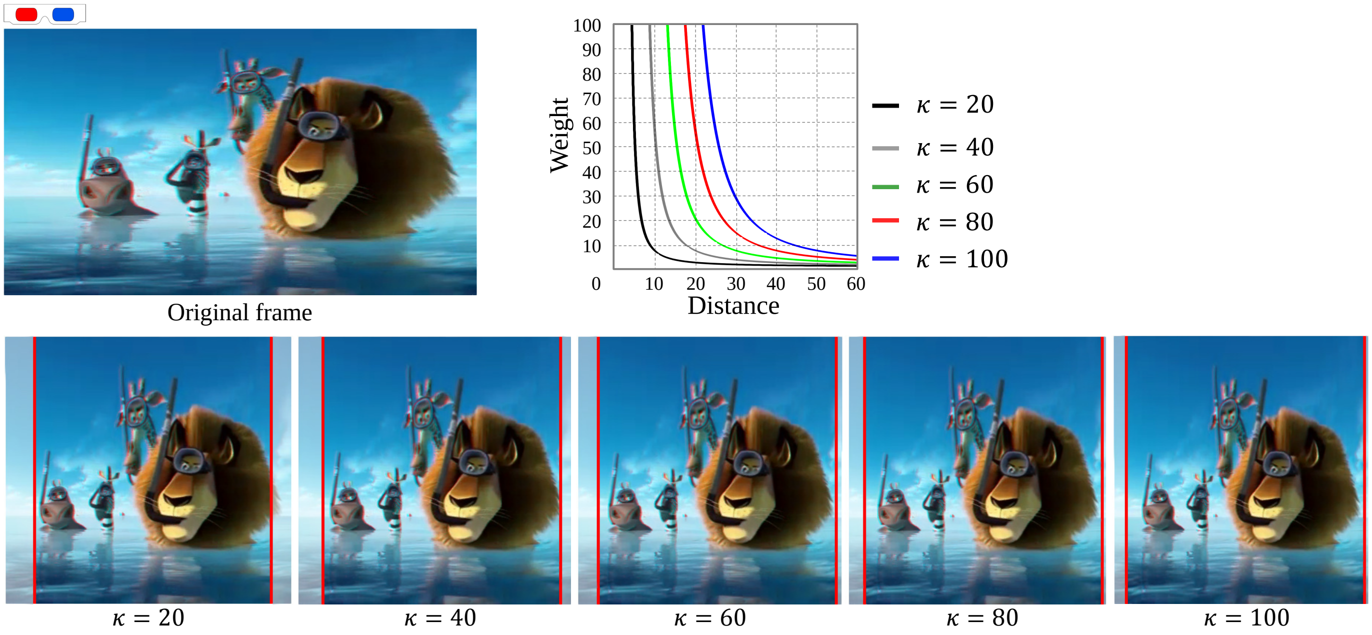
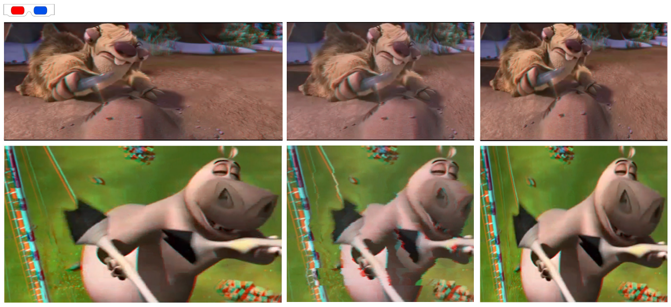


Fig. 7. Warping by using different values of boundary weighting parameter κ. From left to right: warping using $\kappa = 20, 40, 60, 80,$ and 100. The x-axis is the distance between a boundary vertex and the detected significant objects and the y-axis is the weight calculated by Eq. (8).

video containing multiple scenes can be generally divided into several clips. Each clip that represents a single scene is resized individually, and thus, coherently resizing the entire video is unnecessary. For a 640×360 resolution stereoscopic clip with 486 frames, the average computation time for warping and cropping is 7.2414 seconds, that is, resizing a frame takes 0.0149 seconds. To demonstrate the feasibility of the proposed content-aware boundary constraints, a video clip containing evident foreground objects and considerable motions is tested. Figure 5 shows the results of all processes including video segmentation, significance map generation, significant content detection, warping using soft boundary constraints, and cropping with a fixed critical region. The results show that the significant contents are accurately detected, and the detected contents are mostly kept within the desired cropping window during warping because of the proposed content-aware boundary weighting scheme. In addition, without the frame alignment and panning processes, the fixed critical region and cropping window results in easy volumetric cropping and temporal motion preservation (the regions marked with red lines in the 3^{rd} column of Figure 5).

Parameter setting. The parameter α in Eq. (4) and the parameter κ in Eq. (8) are the main parameters in our method. α is the weighting factor for internal vertices, which controls the rigidity and consistency of object deformations. κ is the boundary weighting parameter that controls the strength of boundary constraints. To test the sensitiveness of the retargeting results to these two parameters, various settings of parameter values are tested. The results are shown in Figures 6 and 7. The results in Figure 6 indicate that a large value of α can force high-significance objects to be rigid and consistent, and a small value can avoid skewed artifacts. In the implementation, the default value of α is set to 0.5 to consider the aforementioned factors. The results in Figure 7 show that our cropping results are slightly sensitive to the parameter κ. A small value (κ <20) leads to under-constrained deformation, and a large value (κ >100) results in over-constrained warping. κ=60 is suitable for various cases based on our empirical observation.

Comparisons. Various stereoscopic video clips are selected to evaluate and compare our method with related methods. Some of the selected video clips possess multiple moving objects, unconscious camera shaking, simultaneous camera and object motions, or dense information. Several representative cases are shown in Figures 4-7 and 9, and the other cases are included as attachments in accompanying documents. All results are automatically generated by using the following default parameters: grid resolution is 20×20 pixels; α and k_{max} in Eq. (4) are set to 0.5 and 100, respectively; c and κ in Eq. (8) are set to 2.7 and 60, respectively; and the threshold T_s for significant object extraction is set to 0.5. Please refer to the results and comparisons in the accompanying and supplemental videos, particularly for temporal and stereoscopic effects that are difficult to visualize in still frames. The proposed warping method that considers shape, disparity, and motion preservation is designed for stereoscopic videos. Therefore, comparisons in terms of spatial disparity and temporal motion preservation are conducted. With regard to motion preservation, the proposed method is compared with recent methods that combine warping and cropping in the retargeting scheme [1], [17]. To ensure objectivity in comparison, the resolution of the grid mesh in our method is set to be the same as those in these two methods. Figure 9 shows the comparison results. We can observe that combining warping and cropping can preserve spatial shapes efficiently. However, shifting effect occurs on the deformed trajectories in the results of [1], [17], compared with the original trajectories. This effect is attributed to the use of dynamic critical region and the frame panning process. By contrast, the shifting effect and unnatural temporal motions are efficiently alleviated by using the strategies of floating boundary and fixed cropping window in warping and cropping, respectively. The comparisons between the stereo seam carving method [14] and our method are shown in Figure 8. The results meet the conclusions in [18] that warping-based methods have good results for images/videos containing dense information and seam carving methods are suitable for that containing many homogeneous regions.



Original video frames Basha et al. (2013) Proposed method

Fig. 8. Comparison of the stereo seam carving approach [14] and our method. From left to right: original video frames, retargeting results of Basha et al. [14], and our results.

In addition to qualitative analysis, we conduct a quantitative analysis by using correlation coefficients that represent the statistical relationship between two data sets. First, the centers of objects in frames are selected as feature points, and the motion trajectory of an object is composed of the feature points in frames for temporal coherence evaluation. The number of trajectories used in the quantitative analysis is the number of objects in a video clip. To estimate the shifting effects of the motion trajectories, the offsets of all feature points $X : \{p_i^k - p_{mean}^k\}_{i=1\sim n_f}^{k=1\sim n_t}$ and $\tilde{X} : \{\tilde{p}_i^k - \tilde{p}_{mean}^k\}_{i=1\sim n_f}^{k=1\sim n_t}$ in the original and retargeted video clips are evaluated, where p_{mean}^k and $\tilde{p}_{mean}^k$ represent the means of the nodes in the motion trajectories of the original and retargeted clips, respectively; and n_f and n_t denote the number of feature points and the number of nodes in a trajectory, respectively. The correlation coefficient between these two data sets are defined as $CCef_T(X, \tilde{X}) = Cov(X, \tilde{X})/\sigma_X \sigma_{\tilde{X}}$, where $Cov(X, \tilde{X})$ represents the covariance between X and $\tilde{X}$, and σ_X and $\sigma_{\tilde{X}}$ denote the standard deviations of the data sets X and $\tilde{X}$, respectively. In this experiment, the center points of the objects are the feature points used to evaluate motion preservation. The analysis results are shown in Table I. In the 1^{st}, 4^{th}, 6^{th} and 10^{th} cases, in which the video clips contain multiple moving objects or considerable motions, our results (average $CCef_T = 0.972$) exhibit better performance compared with those of [1] (average $CCef_T = 0.833$) and [17] (average $CCef_T = 0.858$). This superior performance is caused by the use of floating boundaries and the alleviation of shifting trajectories. In the other cases, in which the video clips contain a single moving object with simple motions, our results (average $CCef_T = 0.988$) are similar to those of [1] (average $CCef_T = 0.947$) and [17] (average $CCef_T = 0.974$).

For disparity preservation, our method is compared with recent stereoscopic image retargeting methods [21], [23]. Similarly, the resolution of the control mesh in our method is the same as those in these two methods, and the correlation coefficient is adopted in this analysis. The correlation coefficient in this experiment is defined as $CCef_D(X, \tilde{X}) = Cov(X, \tilde{X})/\sigma_X \sigma_{\tilde{X}}$, where $X : \{p_i^L - p_i^R\}_{i=1\sim n_f}$ and $\tilde{X} : \{\tilde{p}_i^L - \tilde{p}_i^R\}_{i=1\sim n_f}$, and p_i^L and p_i^R are the pixels in the left and right video frames, respectively; and n_f denote the number of selected feature points. The results shown in Figure 10 indicate that our method using floating boundaries and a fixed cropping window can better preserve both the shapes and disparities of visually salient objects, compared with the methods [21], [23]. The analysis results shown in Table II also indicate that our method exhibit better performance in maintaining disparity preservation (average $CCef_D = 0.950$), compared with the methods [23] (average $CCef_D = 0.909$) and [21] (average $CCef_D = 0.887$).

Besides, disparity maps of the stereoscopic retargeting results are generated using the Semi-Global Matching (SGM) (SGM) method [30] for the purpose of visual comparisons. The disparity maps in Figure 11 show that distortion occurs in the retargeting results because of warping. However, the disparity maps of our results have less distortion than that of the related methods [21], [23] because of the floating boundary and cropping strategies. Based on the qualitative and quantitative analyses of the motion and disparity preservation, we conclude that our method is superior to the related methods

TABLE I
Quantitative analysis of temporal motion preservation. The correlation coefficients $CCef_T$ of all feature points in the original and retargeted video clips generated by Wang et al. [1], Wang et al. [17], and the proposed method are compared and presented in this table. "Avg." and "Std." represent the average and standard deviation of the correlation coefficients.

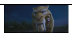
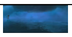
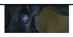
Method											Avg.	Std.
Wang et al. [1]	0.847	0.847	0.977	0.822	0.963	0.839	0.994	0.984	0.922	0.826	0.902	0.069
Wang et al. [17]	0.808	0.950	0.990	0.842	0.962	0.836	0.963	0.988	0.994	0.946	0.928	0.067
Proposed method	0.981	0.973	0.986	0.967	0.994	0.956	0.996	0.989	0.993	0.987	0.982	0.012

TABLE II
Quantitative analysis of disparity preservation. The correlation coefficient $CCef_D$ is used to evaluate the retargeting results generated by Chang et al. [21], Lin et al. [23], and the proposed method. "Avg." and "Std." represent the average and standard deviation of the correlation coefficients.

Method											Avg.	Std.
Chang et al. [21]	0.958	0.941	0.877	0.965	0.823	0.812	0.769	0.934	0.960	0.826	0.887	0.070
Lin et al. [23]	0.958	0.986	0.863	0.987	0.789	0.847	0.846	0.963	0.994	0.859	0.909	0.072
Proposed method	0.991	0.997	0.946	0.971	0.867	0.989	0.970	0.971	0.993	0.802	0.950	0.061

TABLE III
Result of Survey A. Our method is compared with the video retargeting methods [1], [17], in terms of shape and temporal coherence preservation.

	Pairwise comparison 1	Pairwise comparison 2
Wang et al. [1]	474 (37.0%)	N/A
Wang et al. [17]	N/A	496 (38.7%)
Proposed method	806 (63.0%)	784 (61.3%)
Total	1280 (100.0%)	1280 (100%)

TABLE IV
Result of Survey B. Our method is compared with the stereoscopic image retargeting methods [21], [23], in terms of disparity preservation.

	Pairwise comparison 1	Pairwise comparison 2
Chang et al. [21]	385 (36.3%)	N/A
Lin et al. [23]	N/A	401 (37.8%)
Proposed method	675 (63.7%)	659 (62.2%)
Total	1060 (100.0%)	1060 (100%)

in terms of content preservation, particularly for stereoscopic videos containing considerable object/camera motions.

User study. The survey system provided by Rubinstein et al. [31] is used in the user study. In this system, paired comparison method is adopted, in which the participants are shown two results side by side at a time and asked to choose the one they liked better. The user study consists of two main parts: 1) compare our method with the related video retargeting methods [1], [17], in terms of shape and temporal coherence preservation (denoted by Survey A); 2) compare our results with that of [21], [23], in terms of disparity preservation (denoted by Survey B). The Survey A has 128 participants with age ranging from 21 to 48 years old, and the Survey B involves 106 participants with age ranging from 22 to 42 years old. In the comparison, the images/videos having the attributes that can be mapped to major retargeting objectives, namely, shape preservation, disparity preservation, and temporal motion preservation, are utilized. The test datasets for Surveys A and B are made up of 10 videos and images, respectively. The survey results and the number of votes are shown in Tables III and IV. The results indicate that more than 60% participants refer our results to that of the related methods [1], [17], in terms of shape and temporal motion preservation. Similarly, there are more than 60% votes in favor of our results, which indicate that our method has better performance than the related methods [21], [23], in terms of disparity preservation.

Limitations. Content-aware retargeting is based on the saliency detection and content preservation constraints. Similar to other methods, our method may shrink or crop significant contents when an incorrect saliency map is used. Moreover, our method may over-constrain or under-constrain contents during warping when video frames are filled with significant objects or when all pixels or segmented objects have similar significance values. In such scenarios, users are provided with an interface to specify important contents that should be preserved, or when over-constraining or under-constraining occurs, linear rescaling is performed instead of mesh warping.

Accurate video segmentation is challenging. Our method may encounter difficulties from failed video segmentation. For example, a significant object cannot be preserved efficiently when this object is grouped with several insignificant objects. However, this case rarely occurs for a segmentation algorithm. Failed segmentation generally occurs when an incorrect saliency detection result is used in partition. As mentioned earlier, this problem can be solved by using a user interface that can manually mark important contents. Regarding the problem of over-segmentation (the most common problem in segmentation), our results are slightly sensitive to this case. In over-segmentation, an object is partitioned into several sub-objects. The shapes of these sub-objects are preserved individually during warping. The retargeting quality is slightly decreased, in terms of deformation consistency, compared with warping an entire object.

V. Conclusions

This study introduces a novel content-aware retargeting method for stereoscopic videos. The proposed content-aware soft boundary weighting scheme and content preservation constraints enforce visually salient volumetric objects to undergo as-rigid-as-possible and as-consistent-as possible deformation during warping. Moreover, the fixed critical region and cropping window lead to easy volumetric cropping and better temporal motion preservation, compared with the retargeting using different critical regions in frames. The results of the experiments, comparisons, as well as the qualitative and quantitative analyses demonstrate the superiority of the proposed method over related methods in terms of retargeting quality. Besides, The proposed scheme can be applied to various media, including images, stereoscopic images, videos, and stereoscopic videos with the consideration of shape, motion, and disparity preservation.

References

[1] Y.-S. Wang, H.-C. Lin, O. Sorkine, and T.-Y. Lee, “Motion-based video retargeting with optimized crop-and-warp,” *ACM Trans. Graph.*, vol. 29, no. 4, pp. 90:1–90:9, 2010.

[2] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[3] J. Li, Y. Tian, T. Huang, and W. Gao, “Multi-task rank learning for visual saliency estimation,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, no. 5, pp. 623–636, 2011.

[4] W. Kim, C. Jung, and C. Kim, “Spatiotemporal saliency detection and its applications in static and dynamic scenes,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, no. 4, pp. 446–456, 2011.

[5] J.-S. Kim, J.-Y. Sim, and C.-S. Kim, “Multiscale saliency detection using random walk with restart,” *IEEE Trans. Circuits Syst. Video Techn., to appear*, 2013.

[6] W. Kim and C. Kim, “Spatiotemporal saliency detection using textural contrast and its applications,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 24, no. 4, pp. 646–659, 2014.

[7] M. Song, C. Chen, S. Wang, and Y. Yang, “Low-level and high-level prior learning for visual saliency estimation,” *Inf. Sci.*, vol. 281, pp. 573–585, 2014.

[8] L. Zhang, M. Song, Y. Yang, Q. Zhao, C. Zhao, and N. Sebe, “Weakly supervised photo cropping,” *IEEE Trans. Multi.*, vol. 16, no. 1, pp. 94–107, 2014.

[9] Y. Niu, F. Liu, W. chi Feng, and H. Jin, “Aesthetics-based stereoscopic photo cropping for heterogeneous displays,” *IEEE Trans. Multi.*, vol. 14, pp. 783–796, 2012.

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

[11] A. Shamir and S. Avidan, “Seam carving for media retargeting,” *Commun. ACM*, vol. 52, no. 1, pp. 77–85, 2009.

[12] C.-K. Chiang, S.-F. Wang, Y.-L. Chen, and S.-H. Lai, “Fast jnd-based video carving with gpu acceleration for real-time video retargeting,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 19, no. 11, pp. 1588–1597, 2009.

[13] B. Yan, K. Sun, and L. Liu, “Matching-area-based seam carving for video retargeting,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 23, no. 2, pp. 302–310, 2013.

[14] T. Basha, Y. Moses, and S. Avidan, “Stereo seam carving a geometrically consistent approach,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2513–2525, 2013.

[15] B. Guthier, J. Kiess, S. Kopf, and W. Effelsberg, “Seam carving for stereoscopic video,” in *11th IVMSP Workshop: 3D Image/Video Technologies and Applications*, 2013, pp. 1–4.

[16] M. Rubinstein, A. Shamir, and S. Avidan, “Multi-operator media retargeting,” *ACM Trans. Graph.*, vol. 28, no. 3, pp. 23:1–23:11, 2009.

[17] Y.-S. Wang, J.-H. Hsiao, O. Sorkine, and T.-Y. Lee, “Scalable and coherent video resizing with per-frame optimization,” *ACM Trans. Graph.*, vol. 30, no. 4, pp. 88:1–88:8, 2011.

[18] S.-S. Lin, C.-H. Lin, I.-C. Yeh, S.-H. Chang, C.-K. Yeh, and T.-Y. Lee, “Content-aware video retargeting using object-preserving warping,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 10, pp. 1677–1686, 2013.

[19] B. Li, L.-Y. Duan, J. Wang, R. Ji, C.-W. Lin, and W. Gao, “Spatiotemporal grid flow for video retargeting,” *IEEE Trans. Image Processing*, vol. 23, no. 4, pp. 1615–1628, 2014.

[20] Z. Yuan, T. Lu, Y. Huang, D. Wu, and H. Yu, “Addressing visual consistency in video retargeting: A refined homogeneous approach,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 6, pp. 890–903, 2012.

[21] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, “Content-aware display adaptation and interactive editing for stereoscopic images,” *IEEE Trans. Multi.*, vol. 13, no. 4, pp. 589–601, 2011.

[22] K.-Y. Lee, C.-D. Chung, and Y.-Y. Chuang, “Scene warping: Layer-based stereoscopic image resizing,” in *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2012, pp. 49–56.

[23] S.-S. Lin, C.-H. Lin, S.-H. Chang, and T.-Y. Lee, “Object-coherence warping for stereoscopic image retargeting,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 24, no. 5, pp. 759–768, 2014.

[24] J. W. Yoo, S. Yea, and I. K. Park, “Content-driven retargeting of stereoscopic images,” *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 519–522, 2013.

[25] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, “Nonlinear disparity mapping for stereoscopic 3d,” *ACM Trans. Graph.*, vol. 29, pp. 75:1–75:10, 2010.

[26] M. Grundmann, V. Kwatra, M. Han, and I. Essa, “Efficient hierarchical graph-based video segmentation,” in *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2010, pp. 2141–2148.

[27] X. Liu, D. Tao, M. Song, Y. Ruan, C. Chen, and J. Bu, “Weakly supervised multiclass video segmentation,” 2014, pp. 57–64.

[28] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, “A shape-preserving approach to image resizing,” *Comput. Graph. Forum*, vol. 28, no. 7, pp. 1897–1906, 2009.

[29] M. Daszykowski, K. Kaczmarek, Y. V. Heyden, and B. Walczak, “Robust statistics in data analysis – a review: Basic concepts,” *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 2, pp. 203 – 219, 2007.

[30] H. Hirschmüller, “Stereo processing by semiglobal matching and mutual information,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, 2008.

[31] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, “A comparative study of image retargeting,” *ACM Trans. Graph.*, vol. 29, no. 6, pp. 160:1–160:10, 2010.

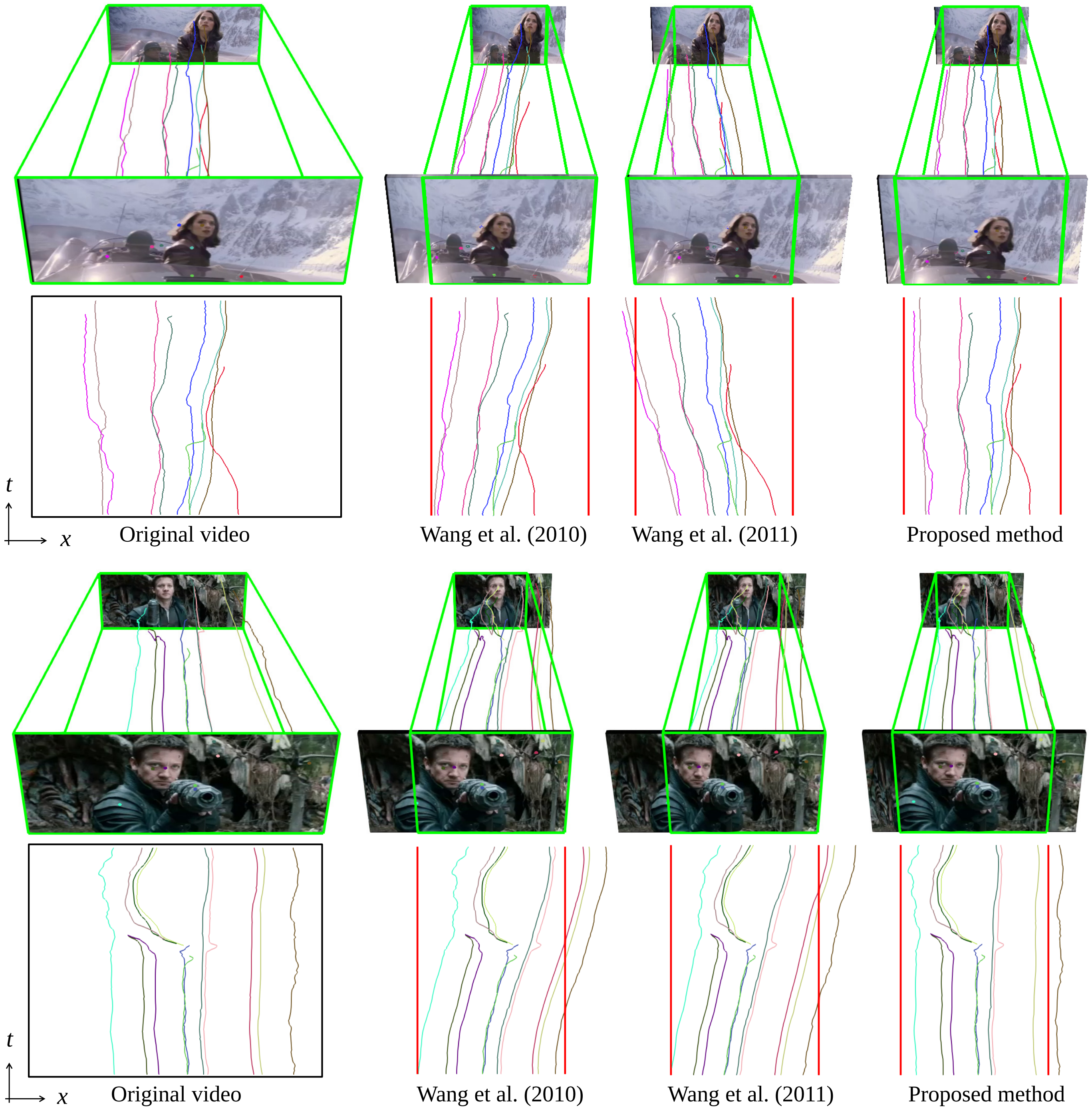


Fig. 9. Comparison of temporal motion preservation. From left to right: original video clip, retargeting results generated by Wang et al. [1], Wang et al. [17], and the proposed method. The video frames, motion trajectories, and cropping windows are shown at the top of each example. The close-up views of the motion trajectories are shown at the bottom of each example.

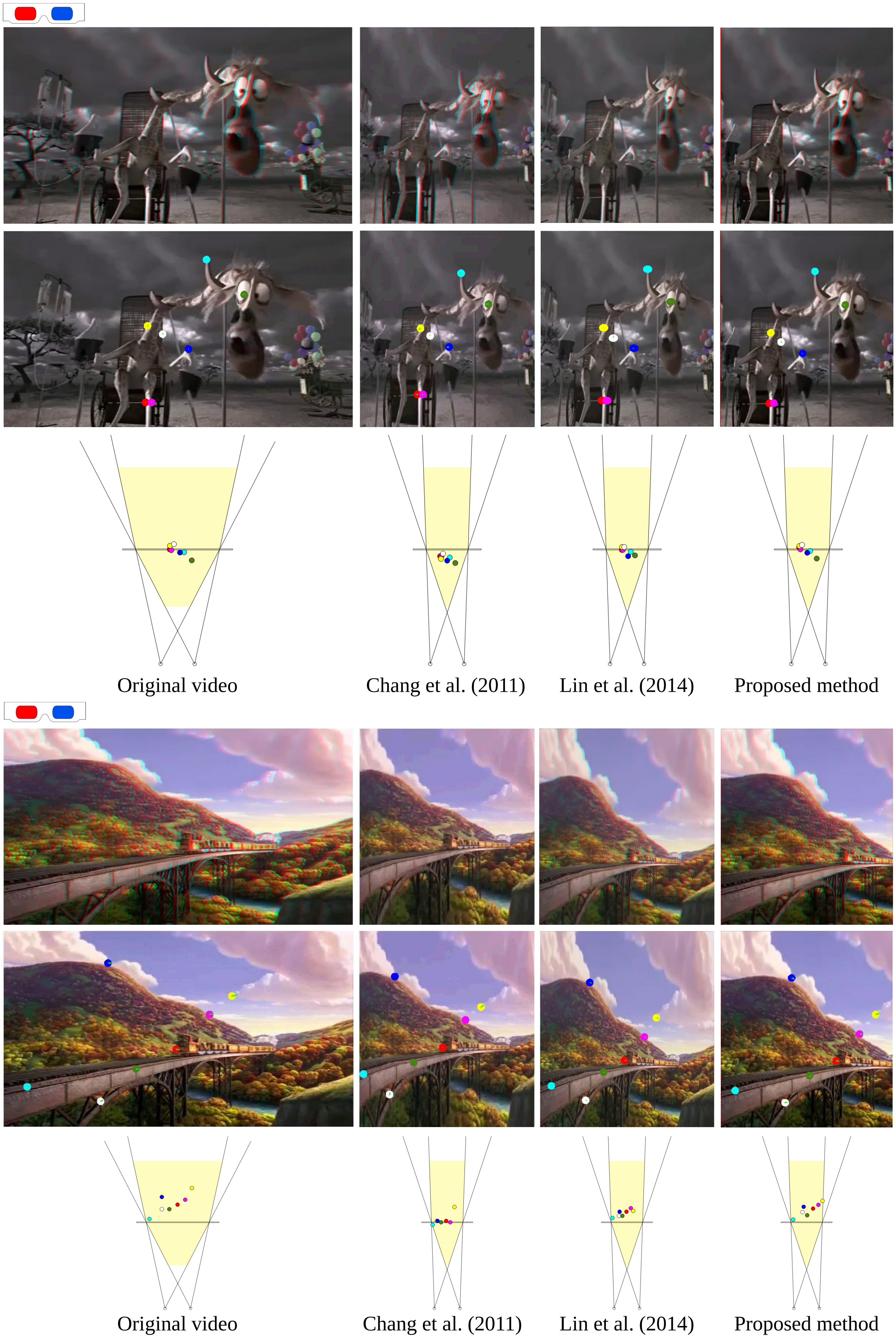


Fig. 10. Comparison of disparity preservation among Chang et al. [21], Lin et al. [23], and the proposed method. The stereoscopic video frames are shown in the first row of each example; the left frames with the selected feature points (marked by colors) are shown in the second row of each example; and the depth distribution of the selected feature points are shown in the third row of each example.

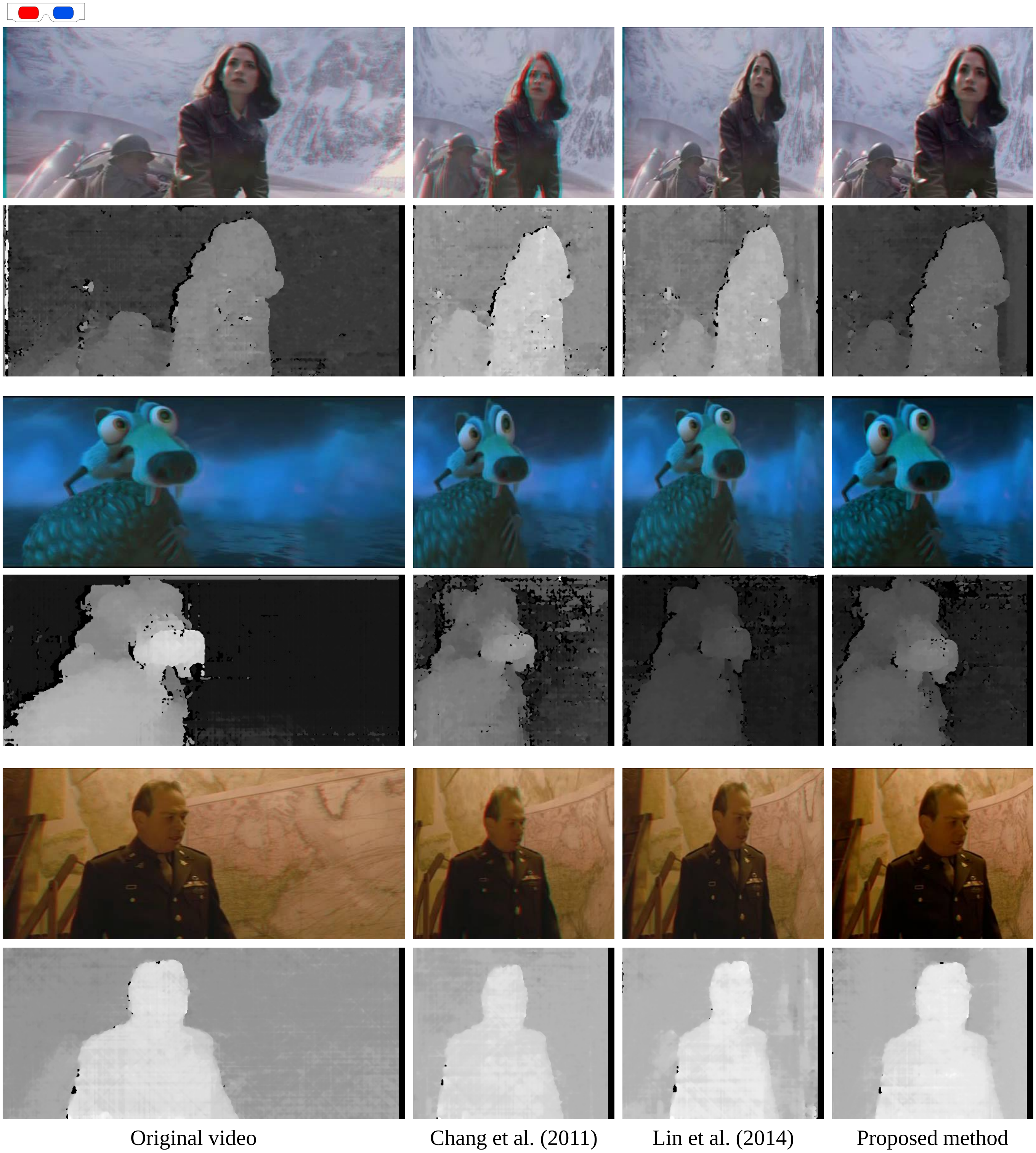


Fig. 11. Comparison of depth distortion. First row: original video frames and their disparity maps; second-fourth rows: the retargeting results and disparity maps of Chang et al. [21], Lin et al. [23], and the proposed method.