

Deformable Shape Preserving Video Retargeting with Salient Curve Matching

Botao Wang, Hongkai Xiong *Senior Member, IEEE*, Zhiqian Ren, Chang Wen Chen *Fellow, IEEE*

Abstract—Video retargeting is dedicated to resizing the resolution of videos in a content aware manner, and it involves three critical challenges, namely, visual saliency preservation, deformation prevention, and temporal consistency persistence. Existing retargeting algorithms include seam carving based approaches, warping based approaches, and cropping based approaches, which mainly concentrate on saliency preservation and fail to prevent the deformation of salient shapes. This paper proposes a deformable shape preserving video retargeting scheme where salient curves extracted from frames are protected from deformation by minimizing the matching cost of curves in the original frames and the retargeted frames. Correspondingly, a curve matching algorithm is developed to generate the deformation cost which is invariant to translation, rotation and scaling with the Bookstein coordinate transform. In turn, the deformation cost of salient curves will be incorporated into the energy map of seam carving. Furthermore, the proposed scheme defines a temporal energy term to penalize the change of the relative position of curves in consecutive retargeted frames with respect to the original frames. Extensive experiments are validated by the visual comparison, user evaluation, deformation analysis and temporal consistency evaluation, which prove that the proposed scheme outperforms state-of-the-art video retargeting methods.

Index Terms—Video retargeting, seam carving, curve detection, curve matching, dynamic programming.

I. INTRODUCTION

In recent years, the rapid development of handheld devices, e.g. smart phones, tablets and laptops, whose screens come in different aspect ratios and resolutions, puts forward an imperative requirement for the content aware image and video resizing to adapt the resolution of the source video and images to different display formats.

Intuitively, conventional video resizing methods are stretching, cropping, and padding [1], as displayed in Fig. 1. The linear stretching approaches would homogeneously compress or stretches the image content and may cause noticeable distortion. The cropping approaches may not bring distortion to the video, however, some contents in the frames are inevitably sacrificed. Although the padding approaches do not cause distortion or lose contents, it would not make full use of the integration of visual and screen content at all. Obviously,

B. Wang, H. Xiong, and Z. Ren are with the Department of Electronic Engineering, Shanghai Jiao Tong University. email:{botaowang, xionghongkai}@sjtu.edu.cn

C. W. Chen is with the Department of Computer Science and Engineering, State University of New York at Buffalo, NY 14260. e-mail: chenew@buffalo.edu

The work was supported in part by the NSFC, under grants U1201255, 61271218, and 61228101.

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

traditional retargeting techniques can not provide satisfactory visual effects so that content-aware video retargeting are widely addressed.

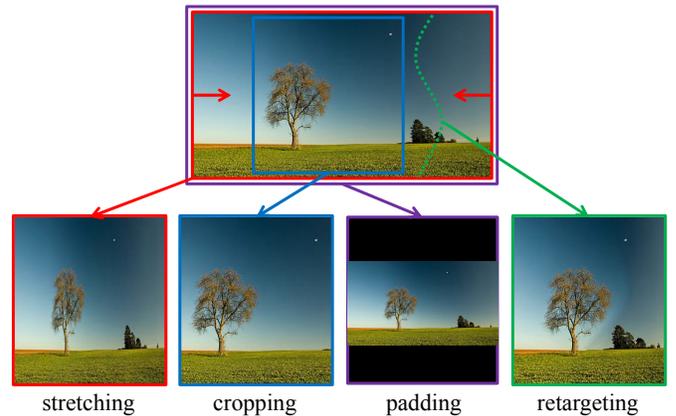


Fig. 1. Conventional image/video resizing approaches, e.g., stretching, cropping and padding, can not obtain satisfactory visual effects in comparison to content-aware image/video retargeting approaches.

In general, there are three critical objectives in video retargeting, known as (1) saliency preservation, (2) deformation prevention, and (3) temporal consistency persistence. On one hand, video retargeting should preserve “important” contents and regions in the retargeted videos which are most attractive to the observers. When manipulating the frames during retargeting, it is vital to prevent deformations and artifacts to the largest extent, including waving, flickering and jittering effects. On the other hand, the retargeted video should exhibit smooth and consistent motion in the temporal domain. To achieve the above objectives, three classes of video retargeting techniques have been developed, known as seam carving based methods, warping based methods, and content-aware cropping based methods.

Seam carving was initially introduced in [2] for image retargeting, where an image is shrunk by iteratively removing the minimum energy seam along vertical or horizontal connected path with dynamic programming. Rubinstein et al. extended it to video retargeting in 2008 [3], and the temporal consistency was preserved by iteratively removing the minimum energy surface obtained by graph cut from the image volume. In 2010, Grundmann et al. [4] claimed that the temporal smoothness of the seams is sufficient but not necessary to obtain temporally coherent videos. Correspondingly, a discontinuous spatio-temporal seam carving algorithm aims to calculate temporally discontinuous seams and allows for carving around fast mov-

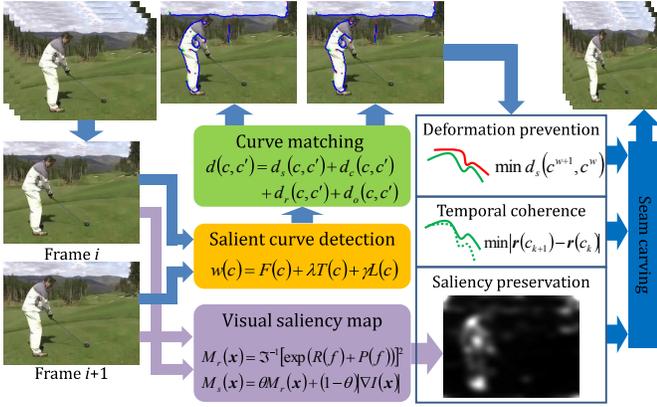


Fig. 2. The diagram of the proposed shape preserving video retargeting scheme.

ing salient regions. In 2011, Mansfield et al. [28] introduced scene carving, which combines a relative depth map with conventional seam carving algorithm to retarget images while maintaining the correct depth ordering. In 2013, the seam carving based video retargeting was improved to preserve the temporal smoothness of seams by matching the frame pixels with the keypoints [5]. Obviously, seam carving based methods change the aspect ratio of a video by compressing unsalient regions where the distortion is less noticeable to the viewer. However, they often bring about severe deformation when the video exists few textureless regions.

Motivated by warping based image retargeting [6] [7], Wolf et al. transformed the original video to the target resolution by solving a sparse linear system of equations [8]. In 2009, Wang et al. suggested to align the consecutive frames by estimating the inter-frame camera motion, and then warp the aligned frames [9]. Later, a non-uniform, pixel-accurate warping was dedicated to interactively defined features which involve video saliency, edge preservation, and scene cut detection [10]. To overcome a variety of constraints and parameters, Niu et al. proposed a warp propagation video retargeting scheme [11] which achieves temporal consistency by introducing a motion history map that propagates information about moving objects between frames. Wang et al. [29] decomposed the video retargeting problem into the spatial component and temporal (motion) component, which can be solved independently. In the spatial domain, each frame is resized individually without the consideration of motion information. In the temporal domain, they keep the pathlines in the optical flow field of the retargeted video as close as the original frames. The major drawback of warping based methods is the waving and jittering effect caused by the global warping on frames.

Both seam carving based and warping based techniques are also regarded as heterogenous approaches [12]. As an alternative, content-aware cropping based techniques aim to preserve temporal coherence and visual saliency at the cost of some contents [13]-[16]. Based on cropping and scaling, Liu et al. resized the frames by minimizing the information loss [14]. Further, Wang et al. advanced to crop the frames to eliminate recurring regions and warp the cropped frames

to mask deformation [15]. In common, the cropping based approaches sacrifice the contents in the borders of the frames to protect salient regions from distortion. Essentially, it has been observed that the most unpleasant distortion in existing video retargeting schemes is caused by the deformation of salient curves in the frames [17]. Moreover, it is sufficient but not necessary to preserve temporal consistency by keeping the smoothness of the seams in the frames. Motivated by the perspective of curves, this paper is dedicated to solving the problem of distortion in video retargeting.

The proposed video retargeting algorithm is dedicated to protecting the shapes of the salient curves from deformation, and attaining temporal consistency in videos by keeping the relative position of the corresponding curves in consecutive frames. Its diagram is demonstrated in Fig. 2. The contribution of this paper is two-fold.

First, a deformable shape preserving video retargeting scheme is proposed where salient curves extracted from frames are protected from deformation by minimizing the matching cost of curves in the original frames and the retargeted frames. The salient curve detection algorithm obtains the salient curves in the frames by maximizing the curve weight that consists of a boundary response term, a smoothness term, and a length term. The deformation term is incorporated into the energy map of seam carving so as to penalize the deformation of salient curves in the retargeted frames. Specifically, a curve matching algorithm is developed by transforming the curves in the Bookstein coordinate, which is invariant to translation, rotation and scaling. It can generate the deformation cost by calculating the matching costs of the curves in the retargeted frames and in the original frames. In this sense, the shapes of the salient curves in the video can be protected by minimizing the deformation term.

Second, the temporal consistency of video is maintained by keeping the relative position of the matched curves in consecutive retargeted frames with respect to the original frames. To be concrete, we match the curves in consecutive frames and encode the relative position of each curve in the current frame with respect to the corresponding curve in the previous frame. To improve the curve matching accuracy, a centroid cost, a scale cost and an orientation cost are incorporated in the matching cost of two curves in consecutive frames, so that false matched curves that have similar shape but different sizes, orientations and positions can be excluded. During retargeting, we keep the relative position of the curves in the retargeted frames to be consistent with the ones in the original frames. Namely, a temporal term is introduced that penalizes the change of the relative position of curves in the retargeted frames with respect to the original frames. By minimizing the variation of the relative position of curves, temporal consistency of the retargeted video can be preserved.

The rest of the paper is organized as follows: Section II presents the salient curves detection algorithm; Section III describes the curve matching algorithm; Section IV illustrates the proposed shape preserving seam carving approach; Section V shows the experimental results, and Section VI concludes the paper.

II. SALIENT CURVES DETECTION

In an image, salient curves would attract viewers' attention more than others [19]-[21]. Although it is difficult, or rather impossible, to mathematically define a salient curve, it is characterized of a few properties: First, a salient curve is a connected path that exhibits abrupt changes in grayscale or color, because human perceptual system is sensitive to high frequency components in an image, where intense variations occur; Second, salient curves reveal sufficient length; Third, salient curves tend to be smooth. As follows, a salient curve detection algorithm is designed to extract salient curves from frames.

A. Curve representation

The basic elements of curves are depicted as 16 oriented segments $\mathcal{S} = \{S_1, S_2, \dots, S_{16}\}$ in Fig. 3. Since an oriented segment connects two adjacent points in a curve, there are two equivalent representations of curves: the segment-based representation and the point-based representation.

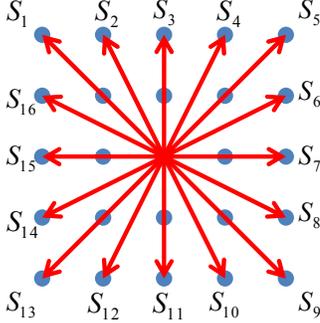


Fig. 3. The basic elements of salient curves: sixteen oriented segments $\mathcal{S} = \{S_1, S_2, \dots, S_{16}\}$.

In the segment-based representation, given the initial point of a curve \mathbf{x}_1 , the curve is denoted as a sequence of oriented segments: $c = (s_1, s_2, \dots, s_{n-1})$, where $s_i \in \mathcal{S}$, $s_i = \mathbf{x}_{i+1} - \mathbf{x}_i$, $i = 1, \dots, n-1$. It is invariant to translation, that is, the shape of a curve does not change with choice of \mathbf{x}_1 . Thus, the segment-based representation can be adopted to evaluate the shape of curves in Fig. 4 (a). In the point-based representation, a curve is represented as a sequence of points: $c = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, where n is the number of points in the curve and $\mathbf{x}_i = (x_i, y_i)$ is the coordinate of the i -th point. Since it is easy to access the points in the curves, the point-based representation can be adopted to evaluate the grayscale intensities of a curve in Fig. 4 (b).

B. Curve weight

The saliency of a curve $c = (\mathbf{x}_i)_{i=1}^n$ is measured by its weight, which is defined as

$$\begin{aligned} w(c) &= F(c) + \lambda T(c) + \gamma L(c) \\ &= \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) + \lambda \frac{1}{n-2} \sum_{i=2}^{n-1} t(s_{i-1}, s_i) \\ &\quad + \gamma \log(n), \end{aligned} \quad (1)$$

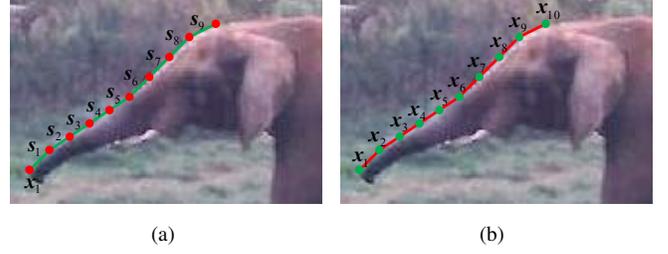


Fig. 4. Curve representation. (a) segment-based representation, (b) point-based representation.

which consists of three terms: the intensity term $F(c)$, the smoothness term $T(c)$ and the length term $L(c)$. λ and γ are constants that balance the three terms.

The intensity term

$$F(c) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) \quad (2)$$

is the mean boundary response of the points in the curves. The boundary response of an image is measured by a function $f(\mathbf{x})$, which gives a noisy indication of whether point \mathbf{x} is located in the boundary. In experiment, we use the gradient magnitude of the image as the boundary response.

$$f(\mathbf{x}) = \sqrt{I_x(\mathbf{x})^2 + I_y(\mathbf{x})^2}, \quad (3)$$

where

$$I_x(\mathbf{x}) = I(x+1, y) - I(x-1, y), \quad (4)$$

$$I_y(\mathbf{x}) = I(x, y+1) - I(x, y-1). \quad (5)$$

Given the input image $I(\mathbf{x})$, $I_x(\mathbf{x})$ and $I_y(\mathbf{x})$ are, respectively, the gradient of the image with respect to x and y . Because only pursuing high boundary-like energy in curve detection would often lead to jagged and twisted curves [22], a smoothness term is defined as:

$$T(c) = \frac{1}{n-2} \sum_{i=2}^{n-1} t(s_{i-1}, s_i). \quad (6)$$

where $t(s_{i-1}, s_i)$ is a function that measures the orientation consistency of two adjacent oriented segments

$$t(s_i, s_{i-1}) = \exp(-|s_i - s_{i-1}|). \quad (7)$$

To prevent abrupt orientation changes in two adjacent oriented segments, the smoothness term can be scaled by

$$t(s_i, s_{i-1}) = \begin{cases} \exp(-|s_i - s_{i-1}|), & \text{if } |s_i - s_{i-1}| \geq T_H, \\ -\infty, & \text{if } |s_i - s_{i-1}| < T_H. \end{cases} \quad (8)$$

where T_H is the threshold of inconsistency in orientation of two consecutive segments. In experiments, we set $T_H = 2$.

The length term

$$L(c) = \log(n) \quad (9)$$

rewards the length of a curve, while the gradient of $L(c)$ decreases with respect to the length of c . Intuitively, once a curve is short, we can exert more reward to the curve weight. As long as the curve has sufficient length, we should impose a little reward on it. The logarithm length term can accurately convey our expectation on the length of salient curves.

C. Curve detection

Existing salient curve detection approaches, e.g. dynamic programming, greedy approximation and best-first search, are of high computational complexity which limits the video retargeting application. Here, we design an efficient curve detection algorithm which is illustrated in Fig. 5.

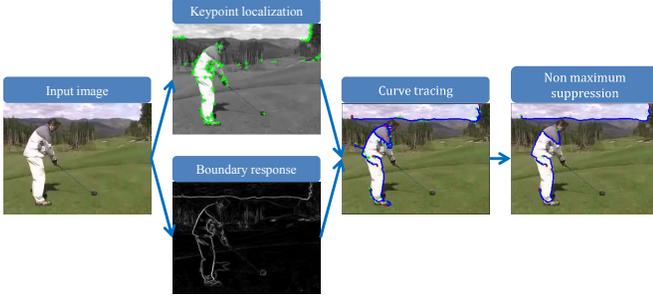


Fig. 5. The underlying curve detection procedure.

It consists of four steps: (1) keypoint extraction, (b) boundary response estimation, (c) curve tracing, and (4) non-maximum suppression. There are two reasons to start salient curves from keypoints: to reduce computational complexity and to facilitate curve matching. In general, keypoints are located on edges and boundaries, so they can be used to identify possible locations of salient curves quickly. In experiment, we adopt Shi-Tomasi corner detector [23] to extract keypoints. The boundary response is defined by Eq. (3). In turn, curves are traced in the boundary response map from the keypoints. Assume a curve $c = (\mathbf{x}_1, \dots, \mathbf{x}_m)$, or equivalently $c = (\mathbf{x}_1, \mathbf{s}_1, \dots, \mathbf{s}_{m-1})$, the optimal one-segment extension of c is

$$\mathbf{s} = \underset{\mathbf{s} \in \mathcal{S}}{\operatorname{argmax}} f(\mathbf{x}_m + \mathbf{s}) + \lambda t(\mathbf{s}_{m-1}, \mathbf{s}). \quad (10)$$

Specifically, the next segment of a curve that ends at \mathbf{x}_m is the one that has the maximum weight in the candidate segment set \mathcal{S} . To compress overlapping curves, a non-maximum suppression is used where the curves are re-arranged in descending order of their weights.

$$\{c_i\}_{i=1}^n, \text{ s.t. } w(c_i) \geq w(c_{i+1}), i = 1, \dots, n-1 \quad (11)$$

where n is the number of curves. Iteratively, the curve with the largest weight joins the final salient curve set, and all curves that overlap with it are removed. Fig. 6 illustrates the examples of the underlying salient curve detection.

III. CURVE MATCHING

To prevent salient curves from deformation in retargeting, we propose a shape matching algorithm which can measure the similarity of two curves. In order to be invariant to translation, rotation and scaling, it matches two curves in the Bookstein coordinate [26][17]. To be concrete, the Bookstein coordinate encodes the location of a point \mathbf{x}_i in a curve $c = \{\mathbf{x}_i\}_{i=1}^n$ relative to the start point \mathbf{x}_1 and the end point \mathbf{x}_n as $B(\mathbf{x}_i|\mathbf{x}_1, \mathbf{x}_n)$. We shall simplify the notation of the



Fig. 6. The test image samples from the underlying salient curve detection.

Bookstein coordinate as $\mathbf{x}_i^B = B(\mathbf{x}_i|\mathbf{x}_1, \mathbf{x}_n)$ if the start point and the end point are irrelevant.

The start point of the curve $\mathbf{x}_1 = (x_1, y_1)$ is mapped to $\mathbf{x}_1^B = (-0.5, 0)$, and the end point of the curve \mathbf{x}_n is mapped to $\mathbf{x}_n^B = (0.5, 0)$ in the Bookstein coordinate. The Bookstein coordinate of any point $\mathbf{x}_i = (x_i, y_i)$ in the curve can be denoted as $\mathbf{x}_i^B = (x_i^B, y_i^B)$

$$x_i^B = \frac{(x_n - x_1)(x_i - x_1) + (y_n - y_1)(y_i - y_1)}{(x_n - x_1)^2 + (y_n - y_1)^2} - 0.5, \quad (12)$$

$$y_i^B = \frac{(x_n - x_1)(y_i - y_1) - (y_n - y_1)(x_i - x_1)}{(x_n - x_1)^2 + (y_n - y_1)^2}. \quad (13)$$

Thus, the curve $c = \{\mathbf{x}_i\}_{i=1}^n$ can be represented in the Bookstein coordinate as $c^B = \{\mathbf{x}_i^B\}_{i=1}^n$. Fig. 7 shows the examples of the Bookstein coordinate representation of curves.

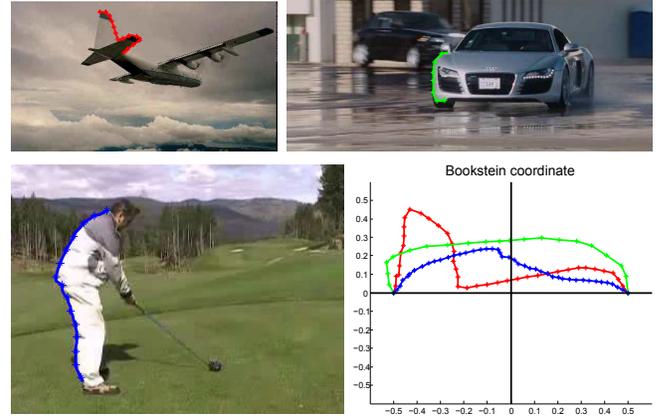


Fig. 7. Bookstein coordinate representations of salient curves. Three curves are extracted from three different images, and represented in the Bookstein coordinate in a rotation, translation and scale invariant manner.

Given two curves $c = (\mathbf{x}_i)_{i=1}^n$ and $c' = (\mathbf{x}'_i)_{i=1}^m$, their deformation cost is measured in the Bookstein coordinate with $c^B = (\mathbf{x}_i^B)_{i=1}^n$ and $c'^B = (\mathbf{x}'_i^B)_{i=1}^m$

$$d_s(c, c') = \frac{1}{2n} \sum_{i=1}^n \min_{1 \leq j \leq m} |\mathbf{x}_i^B - \mathbf{x}'_j^B| + \frac{1}{2m} \sum_{i=1}^m \min_{1 \leq j \leq n} |\mathbf{x}'_i^B - \mathbf{x}_j^B| \quad (14)$$

Each point $\mathbf{x}_i^B \in c^B$ is matched to a point $\mathbf{x}_j^{B'}$ in c'^B , such that

$$j = \operatorname{argmin}_{1 \leq j \leq m} |\mathbf{x}_i^B - \mathbf{x}_j^{B'}|. \quad (15)$$

Hence, the deformation cost of two curves is the mean distance of matched points in the curves with the Bookstein coordinate.

Considering that the orientation, location, and scale of two curves are normalized in the Bookstein coordinate, two curves which have similar shapes but different orientations, locations and scales may be mismatched. It is demonstrated in Fig. 8. To eliminate the mismatch, three cost terms are introduced into the curve matching cost, i.e., the orientation cost, the centroid cost and the scale cost.

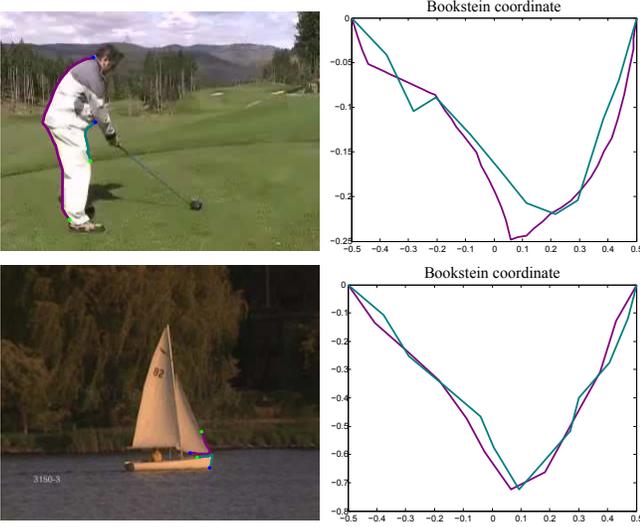


Fig. 8. Mismatched curves that have similar shapes but different orientation, scale and location. The left image shows the curves in the image domain, and the right image shows the curves with the Bookstein coordinate.

Based on the fact that the displacement of corresponding curves in two consecutive frames is small, the centroid cost is to bound the maximum displacement of the centroids of two curves. The centroid of a curve $c = (\mathbf{x}_i)_{i=1}^n$ is defined as

$$\bar{\mathbf{x}}_c = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i. \quad (16)$$

Hence, the centroid cost of two curves c and c' is

$$d_c(c, c') = \begin{cases} \frac{|\bar{\mathbf{x}}_c - \bar{\mathbf{x}}_{c'}|}{T_c}, & \text{if } |\bar{\mathbf{x}}_c - \bar{\mathbf{x}}_{c'}| < T_c \\ \infty, & \text{otherwise.} \end{cases} \quad (17)$$

where T_c is the maximum displacement of the centroids of two curves.

The scale factor of a curve $c = (\mathbf{x}_i)_{i=1}^n$ is defined as

$$r_c = \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i - \bar{\mathbf{x}}_c|, \quad (18)$$

where $\bar{\mathbf{x}}_c$ is the centroid of the curve. The scale factor is the mean distance of the points on the curve to its centroid.

Consequently, the scale cost of two curve is

$$d_r(c, c') = \begin{cases} \frac{\min\{r_c, r_{c'}\}}{\max\{r_c, r_{c'}\}}, & \text{if } \frac{\min\{r_c, r_{c'}\}}{\max\{r_c, r_{c'}\}} > T_r, \\ \infty, & \text{otherwise.} \end{cases} \quad (19)$$

where $T_r \in (0, 1]$ is the minimum ratio of scales of the two curves. That is, the relative scale of two curves may vary within $(T_r, 1/T_r)$. In experiments, $T_r = 0.8$.

The orientation of a segment $\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$ in the curve $c = (\mathbf{x}_i)_{i=1}^n$ is defined as

$$o(\mathbf{s}_i) = \arctan \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \in [0, 2\pi) \quad (20)$$

Hence, an orientation histogram $h(c) \in \mathbb{R}_+^{N_b}$ of the segments can be attained, which consists of N_b bins which are evenly spaced between $[0, 2\pi)$. Each segment casts a vote for the histogram, which is linearly interpolated between two adjacent bins. Eventually, the histogram is normalized so that $|h(c)| = 1$. In experiments, the number of bins is set $N_b = 8$. The orientation cost of two curves is the absolute difference of the orientation histograms

$$d_o(c, c') = \begin{cases} |h(c) - h(c')|, & \text{if } |h(c) - h(c')| \leq T_o \\ \infty, & \text{otherwise} \end{cases} \quad (21)$$

where T_o is the threshold of the maximum orientation cost. That is, if the orientation distributions of segments in two curves are distinct, they can not be matched either. In experiments, $T_o = 0.2$.

Finally, the matching cost of two curves c and c' is the sum of the deformation cost, centroid cost, scale cost, and orientation cost:

$$d(c, c') = d_s(c, c') + d_c(c, c') + d_r(c, c') + d_o(c, c'). \quad (22)$$

Fig. 9 provides the test image examples where salient curves detected by the proposed curve matching approach are marked in red. For each salient curve, its centroid is illustrated with a yellow asterisk, its scale is illustrated with a green circle, and its orientation histogram with blue lines. It can be seen that the dotted blue lines involve eight orientation bins, and the solid blue lines denote the orientation histogram. The length of a solid blue line is proportional to the intensity of the corresponding bin.



Fig. 9. Salient curves are marked in red. The yellow asterisk illustrates the centroid of the curve, and the radius of the green circle centered at the centroid describes the scale of the curve. The dotted blue lines denote the orientation bins of the curve, and the solid blue lines denote relative weights of the orientation histogram.

IV. SHAPE PRESERVING SEAM CARVING

We shall shrink the width of a video to illustrate the proposed shape-preserving video retargeting algorithm. The

original video is denoted as $\{I_i^W\}_{i=1}^N$, where W is the width of the frames, the i denotes the frame index and N is the number of frames. In the k -th step of retargeting, a seam is removed sequentially from the first frame I_1^{W-k+1} to the last frame I_N^{W-k+1} . In this way, the width of the video can be reduced by one: $\{I_i^{W-k+1}\}_{i=1}^N \rightarrow \{I_i^{W-k}\}_{i=1}^N$.

In seam carving, a vertical seam is a path $\mathbf{p} = (p_i)_{i=1}^H$ from the first row to the last row of the frame, where H is the height of the frame. The i -th row contributes one and only one point to the seam, whose coordinate is (p_i, i) . The points in the seam are connected, so that $|p_i - p_{i-1}| \leq 1$. An energy function $E(\mathbf{p})$ defines the ‘‘importance’’ of a seam \mathbf{p} . Typically, the least significant seam is removed from the frame. In the proposed algorithm, the energy function of a seam is defined as

$$E(\mathbf{p}) = E_s(\mathbf{p}) + \alpha E_d(\mathbf{p}) + \beta E_t(\mathbf{p}), \quad (23)$$

which consists of three terms: the saliency term, the deformation term and the temporal term. The saliency term $E_s(\mathbf{p})$ depicts the visual saliency of seam \mathbf{p} . The deformation term $E_d(\mathbf{p})$ provides the matching cost of salient curves after removing seam \mathbf{p} . The temporal term $E_t(\mathbf{p})$ reflects the misplacement of corresponding curves in consecutive frames. Two constants α and β control the relative weights of the deformation term and the temporal term.

A. Visual saliency term

Saliency map reflects the visual attractiveness of each region in the frame. The gradient magnitude is a local indicator of the texture of the frame which is sensitive to the edges and boundaries, so that it is prone to preventing curves from deformation. Thus, a combination of spectral residual saliency [24] and the gradient magnitude of the frame is adopted as the saliency term. The spectral residual saliency map can be obtained by computing the log-spectrum of the frame and extracting the spectral residual. Specifically, the amplitude spectrum $A(u, v)$ and phase spectrum $P(u, v)$ of the input image are attained with the Fourier transform,

$$\mathfrak{F}[I(x, y)] = A(u, v) \exp[i \cdot P(u, v)], \quad (24)$$

where \mathfrak{F} denotes the 2-D discrete Fourier transform. The log-spectrum of the image is the logarithm of the amplitude spectrum,

$$L(u, v) = \log[A(u, v)]. \quad (25)$$

The average spectrum is obtained by convoluting the log-spectrum with a 3×3 average filter,

$$\bar{L}(u, v) = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} * L(u, v). \quad (26)$$

Hence, the spectral residual is defined to be the difference of the average spectrum and the log-spectrum,

$$R(u, v) = \bar{L}(u, v) - L(u, v). \quad (27)$$

The spectral residual saliency map can be obtained by computing the inverse Fourier transform of the spectral residual and the phase spectrum,

$$M_r(x, y) = |\mathfrak{F}^{-1}[\exp(R(u, v) + i \cdot P(u, v))]|^2. \quad (28)$$

where \mathfrak{F}^{-1} denotes the inverse Fourier transform and $M_r(x, y)$ is the spectral residual saliency map. In sum, the final saliency map is the convex combination of the spectral residual saliency map and the gradient magnitude,

$$M_s(x) = \theta M_r(x) + (1 - \theta)|\nabla I(x)|, \quad (29)$$

where $|\nabla I(x)|$ is the gradient magnitude of the frame, $\theta \in [0, 1]$ is the weight parameter, and $M_s(x)$ is the final saliency map. In the experiment, we favor the edge-preserving gradient magnitude, thus we set $\theta = 0.4$. The test images and their corresponding spectral residual saliency maps, gradient magnitudes and the final visual saliency maps are displayed in Fig. 10.

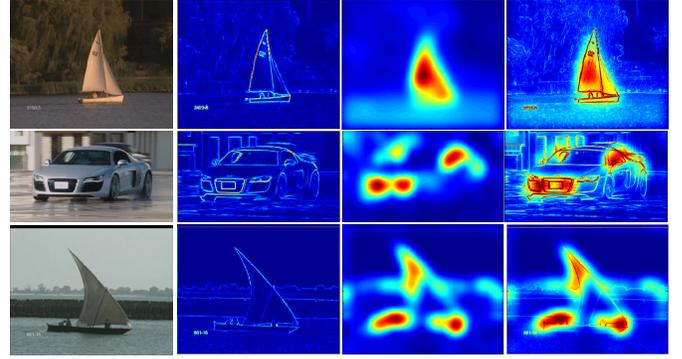


Fig. 10. Visual saliency maps. The first column is the original frames. The second column is the gradient magnitudes. The third column is the spectral residual saliency map. The last column is the final visual saliency map.

Given a saliency map, candidate seams can be extracted with the standard seam carving method. Briefly, the cumulative minimum energy $M_c(x, y)$ of the saliency map $M_s(x, y)$ is computed

$$M_c(x, y) = M_s(x, y) + \min_{i=-1,0,1} M_c(x+i, y-1). \quad (30)$$

A seam can be detected by tracking the minimum energy path in $M_c(x, y)$ starting from a point in the last row,

$$\mathbf{p}(i) = \mathbf{p}(i+1) + \underset{j=-1,0,1}{\operatorname{argmin}} M_c(\mathbf{p}(i+1) + j, i+1). \quad (31)$$

In this sense, a set of minimum energy seams can be extracted $P = \{\mathbf{p}_i\}_{i=1}^n$, where n is the number of candidate seams. To avoid overlapping seams, the mean distance between two seams should satisfy

$$\frac{1}{H} |\mathbf{p}_i - \mathbf{p}_j| > T_s, \forall \mathbf{p}_i, \mathbf{p}_j \in P \quad (32)$$

where $|\cdot|$ stands for the sum of absolute difference, and T_s is the minimum mean distance between two seams. In the experiment, $T_s = 5$. The examples of the candidate seams are illustrated in Fig. 11. Hence, the visual saliency term for each candidate seam $\mathbf{p} \in P$ is

$$E_s(\mathbf{p}) = \frac{1}{H} \sum_{i=1}^H M_s(\mathbf{p}(i), i). \quad (33)$$

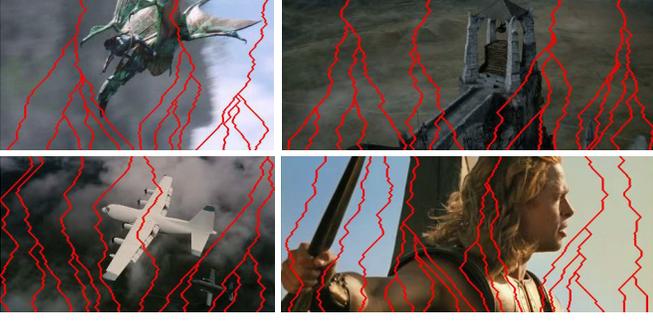


Fig. 11. Candidate seam extraction.

B. Deformation term

The deformation term aims at minimizing the matching cost of the salient curves in the retargeted frames and the original frames. The salient curves detected in the original frame are denoted as $C^W = \{c_i^W\}_{i=1}^n$, where W is the width of the original frames and n is the number of salient curves. After seams are removed from the frame in the retargeting operation, salient curves may be deformed, which are denoted as $C^w = \{c_i^w\}_{i=1}^n$, where w is current width of the frame and c_i^w is the deformed version of c_i^W . Naturally, we expect the shapes of the retargeted curves C^w to be as similar as possible to the original curves C^W . In other words, the matching costs of the retargeted curves and the original ones should be minimized.

If a vertical seam $\mathbf{p} = (p_1, \dots, p_H)$ is removed from the frame $I^w(x, y)$, a salient curve $c^w = (x_i^w)_{i=1}^m$ can be deformed to $c^{w-1}(\mathbf{p}) = (x_i^{w-1})_{i=1}^m$. Given $\mathbf{x}_i^w = (x_i^w, y_i^w)$, $\mathbf{x}_i^{w-1}(\mathbf{p}) = (x_i^{w-1}, y_i^{w-1})$ is a function of \mathbf{p} :

$$\begin{aligned} x_i^{w-1} &= \begin{cases} x_i^w, & \text{if } p_{y_i} > x_i \\ x_i^w - 1, & \text{if } p_{y_i} \leq x_i \end{cases} \\ y_i^{w-1} &= y_i^w \end{aligned} \quad (34)$$

The deformation of a curve after seam removal is demonstrated in Fig. 12.

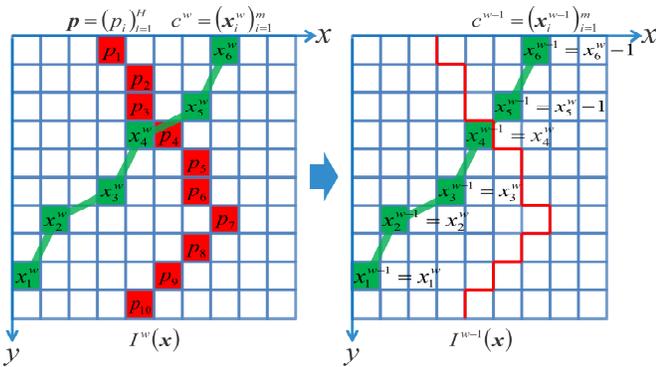


Fig. 12. A curve can be deformed after removing a seam from the frame. The left image shows a frame $I^w(\mathbf{x})$ of width w before seam removal. The seam $\mathbf{p} = (p_i)_{i=1}^H$ is marked with red blocks. The curve $c^w = (x_i^w)_{i=1}^m$ is marked with green blocks. The right image shows the frame $I^{w-1}(\mathbf{x})$ after seam \mathbf{p} is removed from $I^w(\mathbf{x})$. $c^w = (x_i^w)_{i=1}^m$ deforms to $c^{w-1} = (x_i^{w-1})_{i=1}^m$.

Therefore, the deformation term of frame $I_{w-1}(\mathbf{x})$ with respect to the original frame $I_W(\mathbf{x})$ is the sum of deformation

costs of corresponding curves:

$$E_d(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^n d_s(c_i^W, c_i^{w-1}(\mathbf{p})), \quad (35)$$

where d_s is the deformation cost in Eq. (14). Here, we do not use the matching cost in Eq. (22) to measure the dissimilarity of two curves, because c_i^{w-1} is the same but deformed version of c_i^W , so the centroid cost, scale cost and the orientation cost would not be necessary. Finally, the deformation term pursues the optimal seam \mathbf{p} to deform the curves $C^w = \{c_i^w\}_{i=1}^n$ to $C^{w-1} = \{c_i^{w-1}\}_{i=1}^n$ so that the deformation cost of C^{w-1} with respect to the original curves C^W can be minimized.

C. Temporal term

The temporal term aims at preserving the temporal consistency of the retargeted frames by keeping the relative positions of the curves in consecutive frames. Initially, we match the salient curves in consecutive frames $I_{i-1}(\mathbf{x})$ and $I_i(\mathbf{x})$, where the subscript denotes the index of the frame. The salient curves detected in $I_{i-1}(\mathbf{x})$ is denoted as $C^{i-1} = \{c_j^{i-1}\}_{j=1}^{n_{i-1}}$, and the salient curves detected in $I_i(\mathbf{x})$ is denoted as $C^i = \{c_j^i\}_{j=1}^{n_i}$, where n_{i-1} and n_i is the number of salient curves in $I_{i-1}(\mathbf{x})$ and $I_i(\mathbf{x})$, respectively. Curve $c_j^{i-1} \in C^{i-1}$ matches curve $c_l^i \in C^i$, if j and l minimize

$$\begin{aligned} \min_{1 \leq j \leq n_{i-1}, 1 \leq l \leq n_i} & d(c_j^{i-1}, c_l^i) \\ \text{s.t. } & d(c_j^{i-1}, c_l^i) < T_d \end{aligned} \quad (36)$$

where d is the matching cost defined in Eq. (22), and T_d is the threshold of the matching cost. Unlike the curve matching in the deformation term, the matching cost of two curves in consecutive frames contains the deformation cost, the centroid cost, the scale cost and the orientation cost altogether, because here we match two curves from different frames, where variations in centroids, scales and orientations may introduce many false matches.

Given two matched curves $c^{i-1} = (x_j^{i-1})_{j=1}^{L_{i-1}}$ in I_{i-1} and $c^i = (x_j^i)_{j=1}^{L_i}$ in I_i , the backward matching vector $\mathbf{m} = (m_1, \dots, m_{L_i})$ of c^i associates each point in c^i with the closest point in c^{i-1} ,

$$m_k = \underset{1 \leq m \leq L_{i-1}}{\operatorname{argmin}} |x_k^i - x_m^{i-1}|. \quad (37)$$

Thus, the relative location of c^i with respect to c^{i-1} can be denoted as $\mathbf{r} = (r_1, \dots, r_{L_i})$, where

$$r_j = \begin{cases} 1, & \text{if } x_{m_j}^{i-1} < x_j^i \\ 0, & \text{if } x_{m_j}^{i-1} = x_j^i \\ -1, & \text{if } x_{m_j}^{i-1} > x_j^i \end{cases} \quad (38)$$

for $1 \leq j \leq L_i$. Fig. 13 displays an example of the relative location of a curve with respect to the previous frame.

Temporal consistency can be preserved by maintaining the relative position of corresponding curves in consecutive frames. The matched curves in I_{i-1}^{w-1} and I_i^{w-1} is, respectively, $C_{i-1}^{w-1} = \{c_{i-1,j}^{w-1}\}_{j=1}^n$ and $C_i^{w-1} = \{c_{i,j}^{w-1}\}_{j=1}^n$, where $c_{i-1,j}^{w-1}$ matches $c_{i,j}^{w-1}$. In retargeting, $c_{i,j}^{w-1}$ is determined by removing

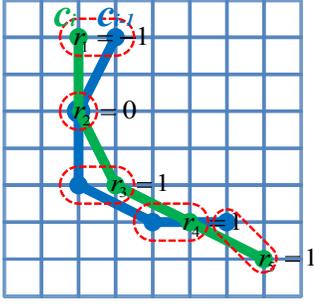


Fig. 13. The relative location of two curves in consecutive frames. The blue curve c^{i-1} is in the $(i-1)$ -th frame. The green curve is the corresponding curve in the i -th frame. The matched points are bound together with red circles. The locations of the points in c^i relative to the corresponding points in c^{i-1} is encoded with Eq. (38)

seam \mathbf{p} from I_i^w , so the relative position of $c_{i,j}^{w-1}$ with respect to $c_{i-1,j}^{w-1}$ is also parameterised by \mathbf{p} . The temporal term penalizes the difference of the relative location of curve c_j^{w-1} and the one in the original frame c_j^W ,

$$E_t(\mathbf{p}) = \frac{1}{n} \sum_{j=1}^n |\mathbf{r}_j^W - \mathbf{r}_j^{w-1}|, \quad (39)$$

where \mathbf{r}_j^W is the relative location of curve c_j in the original frame I^W .

D. Optimal seam detection

Finally, for each seam in the candidate seam set $P = \{\mathbf{p}_i\}_{i=1}^n$, its energy in Eq. (23) can be obtained by combining the saliency term, the deformation term, and the temporal term. The optimal seam for the i -th frame of width w is the one in the candidate seam set that has the minimum energy,

$$\mathbf{p}_i^w = \underset{\mathbf{p} \in P}{\operatorname{argmin}} E(\mathbf{p}) = E_s(\mathbf{p}) + \alpha E_d(\mathbf{p}) + \beta E_t(\mathbf{p}). \quad (40)$$

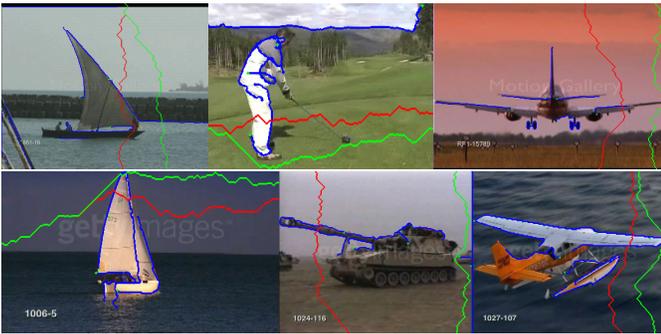


Fig. 14. The blue curves are the detected salient curves. The green seams are the optimal seams obtained by the proposed method. The red seams are the seams of minimal visual saliency. The proposed algorithm effectively prevents salient curves from deformation.

Fig. 14 displays the typical examples of the optimal seams in the test frames, where the blue curves denote the salient curves detected in the frames and the red curves denote the minimum saliency seams, i.e., the minimizer of Eq. (33), that would be selected by most seam carving based video

retargeting schemes. Obviously, the removal of these seams would deform the salient curves and cause severe artifacts. The green curves denote the optimal seams attained by the proposed algorithm, i.e., the minimizer of Eq. (23). It considers the preservation of visual saliency, the prevention of deformation and the temporal coherence, thus achieves the best visual effects. The procedure of the proposed algorithm is illustrated in Algorithm 1.

Algorithm 1: The proposed algorithm

Input: Original frames: $\{I_f^W\}_{f=1}^N$
Output: Retargeted frames: $\{I_f^{W'}\}_{f=1}^N$

```

for  $f = 1$  to  $N$  do
  |  $C_f^W = \{c_{f,i}^W\}_{i=1}^{n_f} = \text{salient curve extraction}(I_f^W)$ ;
end
for  $f = 2$  to  $N$  do
  |  $\{r_{f,i}^W\}_{i=1}^{n_f} = \text{inter-frame curve matching}(C_{f-1}^W, C_f^W)$ ;
end
for  $w = W$  to  $W'$  do
  for  $f = 1$  to  $N$  do
    |  $M_g = \text{gradient magnitude}(I_f^w)$ ;
    |  $M_r = \text{spectral residual}(I_f^w)$ ;
    |  $M_s = \theta M_r + (1 - \theta) M_g$ ;
    |  $P = \{\mathbf{p}_i\}_{i=1}^{N_P} = \text{seam extraction}(M_s)$ ;
    | saliency term  $E_s(\mathbf{p}) = M_s(\mathbf{p})$ ,  $\mathbf{p} \in P$ ;
    |  $E_d(\mathbf{p}) = \text{curve matching cost}(C_f^W, C_f^{w-1}(\mathbf{p}))$ ;
    | temporal cost  $E_t(\mathbf{p}) = \sum_{i=1}^{n_f} |r_{f,i}^W - r_{f,i}^{w-1}(\mathbf{p})|$ ;
    | seam energy  $E(\mathbf{p}) = E_s(\mathbf{p}) + \alpha E_d(\mathbf{p}) + \beta E_t(\mathbf{p})$ ;
    | optimal seam  $\mathbf{p}^* = \min E(\mathbf{p})$ ,  $\mathbf{p} \in P$ ;
    |  $I_f^{w-1} = \text{seam removal}(I_f^w, \mathbf{p}^*)$ ;
    |  $C_f^{w-1} = \text{deform curve}(C_f^w, \mathbf{p}^*)$ ;
  end
end

```

V. EXPERIMENTAL RESULTS

In experiments, the proposed algorithm is compared with two state-of-the-art video retargeting methods: the improved seam carving [3] and the matching area based video retargeting [5]. The improved seam carving method computes the spatio-temporal L_1 norm as the energy map of the frame, and uses graph cuts to generate the optimal seam. Within the matching area based video retargeting method, a blend of gradient magnitude from the Sobel operator and the saliency map of [25] [27] is used as the energy map. The parameters are set according to [5]: $N_{KP} = 10$, match window = 3, match threshold = 0.2, search range = H/N_{KP} , where H is the height of the video and N_{KP} is the number of keypoints.

The test videos involve both movie clips and standard video sequences, which are further grouped into three categories in terms of the motion of the background, i.e., static background, mild motion background, and fast motion background. Experiments are conducted on a laptop with Intel Core 2 Duo CPU P8700 2.53GHz CPU and Ubuntu 13.10 operating system. The evaluation consists of four parts: frame comparison, subjective

evaluation, deformation analysis, and temporal consistency evaluation, respectively.

Some combinations of α and β in Eq. 23 are tested, and we found that $\alpha = 15$ and $\beta = 5$ gives relatively good performance. Thus, we use this parameter setting in the experiment.

A. Frame comparison

Fig. 15 (a) and Fig. 15 (b), respectively, illustrate the retargeted frames of the standard video sequences and the movie sequences, where the first column shows the original frames; the second column shows the frames retargeted by the improved seam carving method; the third shows the frames retargeted by the matching area based retargeting method; and the last column shows the frames retargeted by the proposed method. The resolutions of the original frame and the retarget frames and the run-time of the proposed method are provided below the frames. It can be observed that the proposed algorithm could preserve the salient curves and primary shapes in the videos than others. In addition, we enlarge the local patches in the retargeted frames to elaborately evaluate local detail in Fig. 16 and Fig. 17. Obviously, the proposed algorithm can significantly reduce the distortion caused by the deformation of salient shapes.

In addition, some frames from videos that have large camera motion are displayed in Fig. 18. It could be observed that the proposed method is robust with large camera motion.

B. Subjective evaluation

Since subjective evaluation is widely adopted for video retargeting [5] [10] [9], we also conducted a user study with 9 participants coming from diverse backgrounds. They were presented with an original video sequence and two retargeted videos side by side, and asked to answer which retargeted result was preferred. The users were kept naive about the purpose of the experiment and were not provided with any special technical instructions. We used 10 test videos in the user study, and resized each video to 60% to 75% width using fully automatic versions of the improved seam carving, the matching area based retargeting and the proposed method. For each video, there were three pairwise comparisons, and each participant was asked to make $3 \times 10 = 30$ comparisons. We obtained a total $30 \times 9 = 270$ answers, and each method was compared $2 \times 10 \times 10 = 200$ times.

TABLE I

PAIRWISE COMPARISON OF THE PROPOSED METHOD WITH IMPROVED SEAM CARVING (ICS) AND MATCHING AREA BASED RETARGETING (MA)

↓outperforms→	Ours	ISC	MA	Total	Prefer%
Ours	-	64	79	143	71.5%
ISC	79	-	26	105	52.5%
MA	11	11	-	22	11.0%

Table I shows the result of the pairwise comparison. The proposed method was preferred in 71.5% (143/200), and was favored over the improved seam carving in 64.0% (64/100) and over the matching area based retargeting in 79.0% (79/100).



Fig. 15. The visual frame of (a) standard video sequences (b) movie sequences retargeted, by the improved seam carving (second column), the matching area based retargeting (third column), and the proposed method (last column). The original frames are in the first column.

In contrast, the improved seam carving was favored only in 52.5% (105/200) and the matching area based retargeting in 11.0% (22/200).

C. Deformation analysis

Considering that there is no widely accepted objective metric to assess the quality of video retargeting, we introduce a matching cost metric to measure the deformation of salient curves in the retargeted videos with respect to the original videos. The matching cost of the retargeted frame $I_w(\mathbf{x})$ and the original frame $I_W(\mathbf{x})$ is computed as follows. First, salient curves are extracted from the two frames as described in Section II. Then, we match the curves in the retargeted frame and the original frame as described in Section III. Since the original frame and the retargeted frame have different sizes, the centroid cost and scale cost, which are related to the size of the image, should be taken into account. Here, the matching

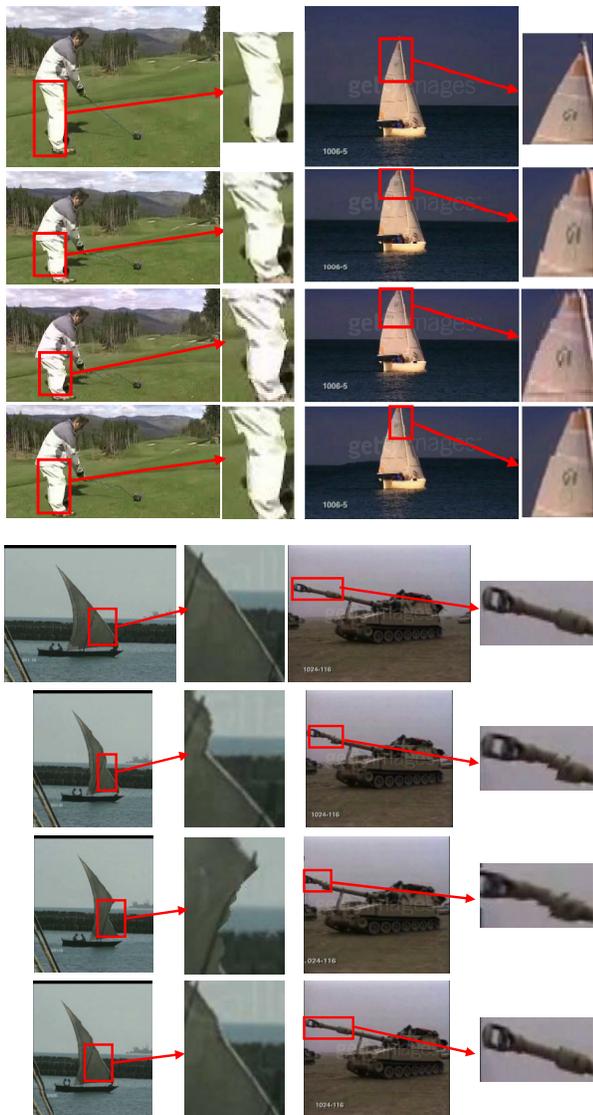


Fig. 16. Local effect of the standard sequences retargeted by the improved seam carving (second row), the matching area based retargeting (third row) and the proposed method (last row). The original frames are displayed in the first row.

cost consists of the deformation cost and the orientation cost.

$$d(c, c') = d_s(c, c') + d_o(c, c') \quad (41)$$

Thus, the matched curves in the original frames and the retargeted frames are, respectively, $\{c_i^W\}_{i=1}^n$ and $\{c_i^w\}_{i=1}^n$, where n is the number of matched curves, and c_i^w matches c_i^W . The matching cost of the retargeted video and the original video is defined as

$$D = \frac{1}{n} \sum_{i=1}^n d(c_i^w, c_i^W). \quad (42)$$

To ensure the accuracy of the matching cost, we further manually examine all of the matched curves and discard false matched ones.

In the experiment, we retargeted four videos for each motion category (static background, mild motion background and fast

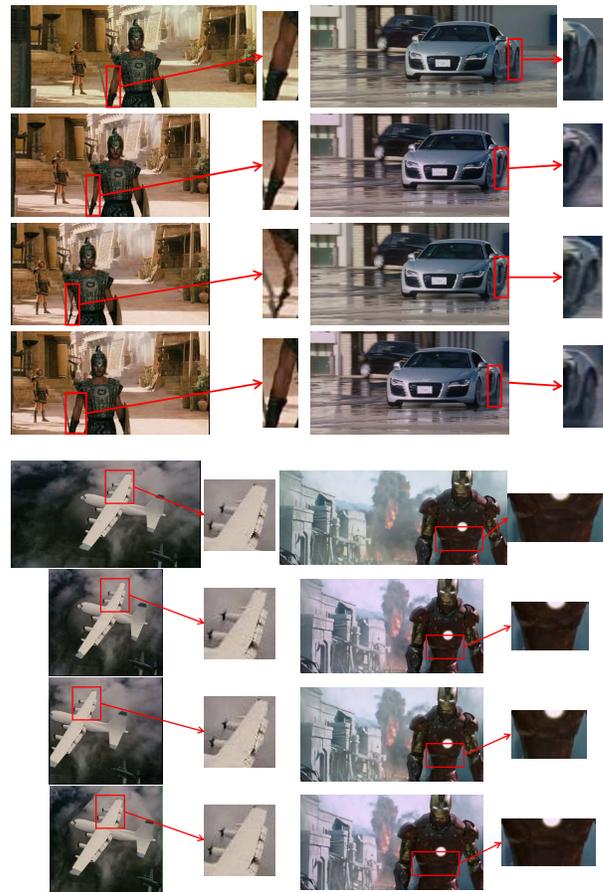


Fig. 17. Local effect of the movie clips retargeted by the improved seam carving (second row), the matching area based retargeting (third row) and the proposed method (last row). The original frames are displayed in the first row.



Fig. 18. Frames from videos that have large camera motion. For each video, the left column shows the original frames, and the right column shows the retargeted frames.

motion background) with the improved seam carving, the matching area based retargeting and the proposed method. To guarantee fairness, we only keep the matching pairs so that the curve c^W in the original video could be matched in at least two methods. If c^W can only find its matches in two methods, we manually label the corresponding curve in the remaining method. Finally, the matches curves in the original video and the videos retargeted by the improved seam carving, matching area based retargeting and the proposed method could, respectively, be denoted as $\{c_i^O\}_{i=1}^n$, $\{c_i^I\}_{i=1}^n$, $\{c_i^M\}_{i=1}^n$ and $\{c_i^P\}_{i=1}^n$, where c_i^I , c_i^M and c_i^P match c_i^O . Table II shows

the matching cost result.

TABLE II
THE DEFORMATION COST COMPARISON ($\times 10^{-1}$)

		video1	video2	video3	video4	Mean
Static	ISC	0.119	0.125	0.105	0.114	0.116
	MA	0.116	0.103	0.104	0.114	0.109
	Ours	0.117	0.094	0.102	0.111	0.106
Mild	ISC	0.108	0.138	0.110	0.116	0.118
	MA	0.116	0.152	0.098	0.114	0.120
	Ours	0.109	0.136	0.095	0.109	0.112
Fast	ISC	0.120	0.112	0.225	0.085	0.136
	MA	0.121	0.108	0.219	0.084	0.133
	Ours	0.118	0.106	0.218	0.084	0.132

In general, there are less severe deformations in the videos that has static background, and more severe deformations in the videos that has fast motion. The proposed algorithm has lower mean deformation costs than other two methods in most test videos. On average, it has the lowest mean deformation cost as well.

D. Temporal consistency evaluation

Temporal consistency is critical to video retargeting, because human visual system is very sensitive to the temporal artifacts, such as jittering and waving effects. To simplify the evaluation, we choose test videos with static backgrounds so that the misalignment of the salient curves caused by the temporal inconsistency can be obtained by the frame difference. Specifically, let $I_t(\mathbf{x})$ and $I_{t+\nabla t}(\mathbf{x})$ be the t -th frame and the $(t + \nabla t)$ -th frame either in the original video or in the retargeted video, the frame difference is defined as

$$\nabla_t I(\mathbf{x}) = |I_{t+\nabla t}(\mathbf{x}) - I_t(\mathbf{x})|. \quad (43)$$

The frame differences of the original frames (first column), the retargeted frames by the improved seam carving method (second column), the matching area based retargeting (third column) and the proposed method (last column) are displayed in Fig. 19.

As the original frames are temporally coherent, the frame differences of the original frames can be considered as the ground truth. Because the backgrounds of the test videos are static, the temporal inconsistent effects in the retargeted frames would produce noises in the frame differences. Therefore, the more similar to the original frame difference the frame difference of the retargeted frames is, the more temporally coherent the retargeted video will be.

It could be observed that the improved seam carving method produces the worst results, because the frame differences of the improved seam carving method contain a lot of background noise and the misalignment of salient curves and boundaries. On the other hand, the proposed algorithm obtains the best results, because its frame differences is the most similar to the ones in the original frames.

Also, temporal inconsistency could accumulate along the frames, which distorts objects frame-by-frame. Therefore, another experiment is conducted to compare the last frames of the test videos retargeted by the three methods. The result is displayed in Fig. 20. The improved seam carving and the

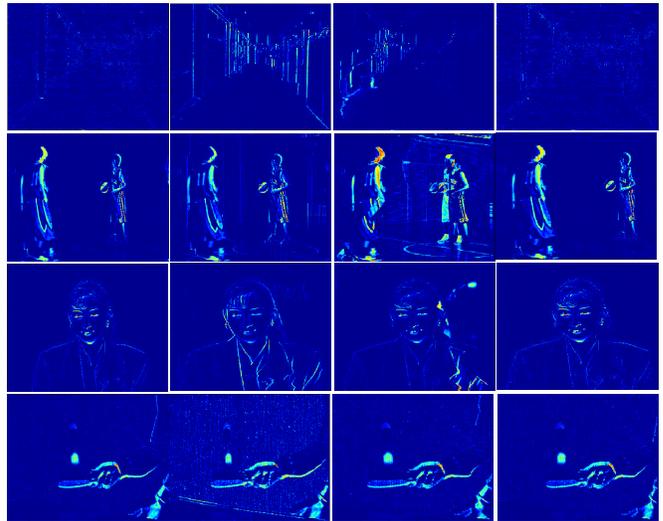


Fig. 19. The frame differences of the original frames (first column) and the retargeted frames by the improved seam carving method (second column), the matching area based retargeting method (third column) and the proposed method (last column).

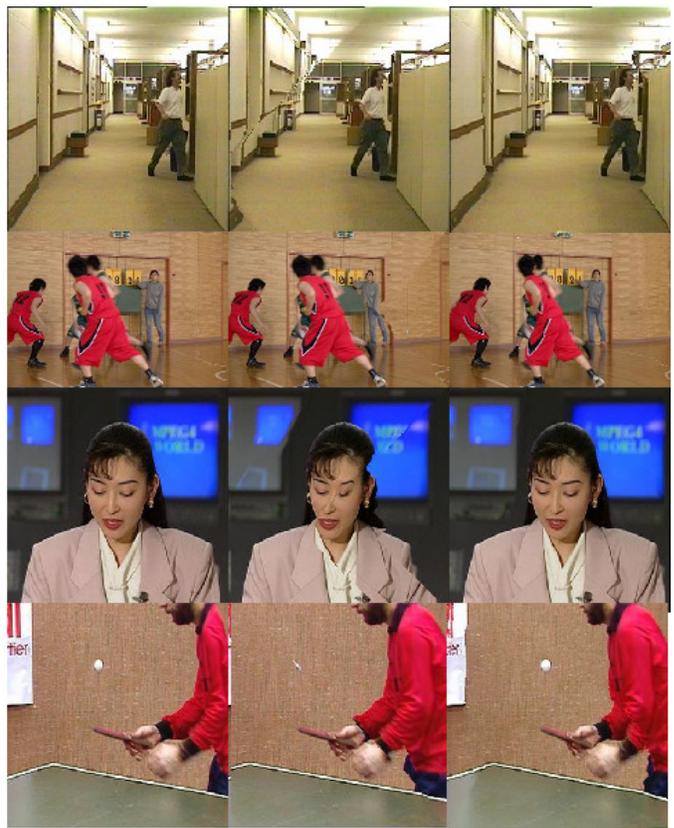


Fig. 20. The last frame of the test videos. Left: improved seam carving. Middle: matching-area based retargeting. Right: proposed method.

proposed method have similar performance, which is much better than matching-area based retargeting.

VI. CONCLUSIONS

In this paper, a deformable shape preserving video retargeting algorithm is proposed from the perspective of salient

curves. A salient curve detection and a curve matching algorithm are designed where a deformation term and a temporal term are introduced into the energy map of the seam carving framework. To be invariant to translation, rotation and scaling, the curve matching is operated by transforming the curves in the Bookstein coordinate. The temporal consistency of video is maintained by keeping the relative position of the matched curves in consecutive retargeted frames with respect to the original frames. The saliency preservation, salient shape protection and temporal consistency are jointly considered to generate the optimal content of the retargeted video. Extensive experiments are validated by the visual comparison, user evaluation, deformation analysis and temporal consistency evaluation, which prove that the proposed scheme outperforms state-of-the-art video retargeting methods.

REFERENCES

- [1] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Trans. Graphics*, vol. 29, no. 6, Dec. 2010.
- [2] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graphics*, vol. 26, no. 3, pp. 1-10, July 2007.
- [3] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graphics*, vol. 27, no. 3, Aug. 2008.
- [4] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 569-576, June 2010.
- [5] Y. Bo and S. Kairan, and L. Liu, "Matching-area-based seam carving for video retargeting," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 302-310, Feb. 2013.
- [6] R. Gal, O. Sorkine, and D. Cohen-Or, "Feature-aware texturing," in *Proc. Eurographics Symp. on Rendering Techniques*, Switzerland, pp. 297-303, June 2006.
- [7] Y. Wang, C. Tai, O. Sorkine, and T. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graphics*, vol. 27, no. 5, Dec. 2008.
- [8] L. Wolf, M. Guttman, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. Int'l Conf. Computer Vision*, Rio de Janeiro, Brazil, pp. 1-6, Oct. 2007.
- [9] Y. Wang, H. Fu, O. Sorkine, T. Lee, and H. Seidel, "Motion-aware temporal coherence for video resizing," *ACM Trans. Graphics*, vol. 28, no. 5, Dec. 2009.
- [10] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A System for Retargeting of Streaming Video," *ACM Trans. Graphics*, vol. 28, no. 5, Dec. 2009.
- [11] Y. Niu, F. Liu, X. Li, and M. Gleicher, "Warp propagation for video resizing," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 537-544, June 2010.
- [12] Z. Yuan, T. Lu, Y. Huang, D. Wu, and H. Yu, "Addressing visual consistency in video retargeting: a refined homogeneous approach," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 22, no. 6, pp. 890-903, June 2012.
- [13] X. Fan, X. Xie, H. Zhou, and W. Ma, "Looking into Video Frames on Small Displays," in *Proc. ACM Int'l Conf. Multimedia*, pp. 247-250, 2003.
- [14] F. Liu and M. Gleicher, "Video retargeting: automating pan and scan," in *Proc. ACM Int'l Conf. Multimedia*, New York, NY, USA, pp. 241-250, Oct. 2006.
- [15] Y. Wang, H. Lin, O. Sorkine, and T. Lee, "Motion-based Video Retargeting with Optimized Crop-and-Warp," *ACM Trans. Graphics*, vol. 29, no. 4, Dec. 2010.
- [16] T. Deselaers, P. Dreuw, and H. Ney, "Pan, zoom, scan – Time-coherent, trained automatic video cropping," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, AK, USA, pp. 1-8, June 2008.
- [17] P. Felzenszwalb and J. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, pp. 1-8, June 2007.
- [18] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graphics*, vol. 28, no. 3, July 2009.
- [19] T. Alter, and R. Basri, "Extracting Salient Curves from Images: An Analysis of the Saliency Network," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 13-20, 1996.
- [20] S. Wang, T. Kubota, J. Siskind, and J. Wang, "Salient closed boundary extraction with ratio contour," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 546-561, April 2005.
- [21] P. Felzenszwalb and D. McAllester, "A min-cover approach for finding salient curves," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshop*, pp. 185, June 2006.
- [22] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530-549, May 2004.
- [23] J. Shi and C. Tomasi, "Good Features to Track," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 593-600, June 1994.
- [24] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, pp. 1-8, June 2007.
- [25] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254C1259, Nov. 1998.
- [26] I. Dryden and K. Mardia, "Statistical shape analysis," *John Wiley and Sons*, 1998.
- [27] J. Harel, "Graph-Based Visual Saliency (GBVS)," July. 2012. [Online]. Available: <http://www.klab.caltech.edu/~harel/share/gbvs.php>
- [28] A. Mansfield, P. Gehler, L. Gool, and C. Rother, "Scene Carving: Scene Consistent Image Retargeting," in *European Conf. Computer vision*, pp. 143-156, 2011.
- [29] Y. Wang, J. Hsiaio, O. Sorkine, and T. Lee, "Scalable and Coherent Video Resizing with Per-Frame Optimization," *ACM Trans. Graphics*, vol. 30, no. 4, July 2011.