Coarse-to-Fine Temporal Optimization for Video Retargeting Based on Seam Carving

Wei-Lun Chao¹, Student Member, IEEE, Hsiao-Hang Su²,
Jian-Jiun Ding*³, Member, IEEE, Shao-Yi Chien⁴, Member, IEEE, and
Winston Hsu⁵, Member, IEEE,
¹,³ Digital Image and Signal Processing Lab, Graduate Institute of Communication Engineering, National Taiwan University (NTU), Taipei, Taiwan, 10617
²,⁴ Graduate Institute of Electronic Engineering, NTU, Taipei, Taiwan, 10617
⁵ Graduate Institute of Networking and Multimedia, NTU, Taipei, Taiwan, 10617
E-mail: weilunchao760414@gmail.com¹, waterhangg@ media.ee.ntu.edu.tw²,
djj@cc.ee.ntu.edu.tw³, sychien@cc.ee.ntu.edu.tw⁴, winston@csie.ntu.edu.tw⁵

Abstract

Image and video retargeting have attracted significant attention in computer vision and multimedia researches in recent years. In this paper, a new video retargeting scheme based on seam carving and temporal information is presented. In order to make the efficient dynamic programming algorithm available for video retargeting and propagate the seam information between consecutive frames, two new video energy functions, motion weight prediction and pixel-based optimization, are proposed in our scheme. Motion weight prediction exploits the block-based motion estimation to coarsely predict the seam locations of the current frame based on the seams detected in the previous frame. To avoid the possible seam discontinuity and motion mismatch problems, the concept of Gaussian masks and pixel indexing are
introduced to modify the motion weight prediction function and reduce the search range of dynamic programming. Then, the technique of *pixel-based optimization* utilizes the concepts of the pixel-based optical flow and temporal coherence measurement to explore better spatial and temporal relationships among the frames in the shrunk search range. Simulation results show that the proposed video retargeting scheme could not only achieve content-aware and temporally smoothing retargeting results, but also drastically reduce the computational complexity.

*Index Terms*—Video retargeting, seam carving, motion estimation, dynamic programming
I. INTRODUCTION

With the fast development of technology, portable devices and displayers such as mobile phones, PDAs, e-book readers, and tabular computers have gradually become an indispensable part in human life. These display units usually have different screen sizes and aspect ratios, whereas images and videos are almost made in single sizes. Therefore, how to resize them to fit the screen size has become an important issue of multimedia researches in recent years. Among different kinds of resizing methods, the content-aware resizing, also called retargeting, has attracted the most attention nowadays.

The concept of retargeting is to change the size of images or video frames while maintaining the characteristic feature intact and minimizing the loss of important content. The algorithms of image retargeting generally consist of two steps: importance analysis of images (also called energy function) and resizing. The first step exploits the concepts of image content analysis like the saliency map, gradients, and object recognition to determine the importance of each pixel or region in the image. The second step then uses image resizing methods such as scaling, cropping, warping, and even seam carving to change the original image into the desired size. When considering video retargeting, how to take the temporal information (such as motion vectors) into consideration and maintain the perceptual coherence are still hot topics in this area.

In this paper, we exploit the popular seam carving technique [1] as the basic framework for image and video retargeting. For image retargeting, the forward and backward energy terms (proposed in [1-2]) are properly combined during the seam carving process; therefore, the importance of image pixels is determined not only by the gradients and saliency measures, but also based on the artifacts after removing each pixel. For the more challenging video retargeting problem, two new video energy functions, motion weight prediction and pixel-based optimization, are proposed to take the temporal information into consideration and make the dynamic programming algorithm available for video retargeting. The motion weight prediction exploits the block-based motion vectors, which can be easily acquired from the compressed video streams, to guide the coarse positions of seams in each frame based on the seams
detected in the preceding frame. To prevent the potential seam discontinuity and scale mismatch problems when applying the motion vectors, the idea of Gaussian masks and pixel indexing are adopted to diffuse the motion information and reduce the search range of dynamic programming.

The *pixel-based optimization* then utilizes the concept of the pixel-based optical flow and the temporal coherence measurement mentioned in [3] to further preserve the perceptual plausibility and determine the accurate seam location during the dynamic programming process. The new video carving procedure could effectively utilize the motion vectors, facilitate the original graph-cut-based video carving procedure proposed in [2], and release the restriction introduced in their work that the seams in one frame could move only one pixel away from the seams detected in the preceding frame. From the experimental results on several standard videos, the proposed retargeting scheme can not only preserve important content, but also achieve spatially and temporally smoothness in the retargeted results.

This paper is organized as follows: In Section II, related work about image and video retargeting are briefly reviewed, and the combination of forward and backward energy terms for image retargeting is presented in Section III. In Sections IV and V, the proposed video retargeting scheme, the two new video energy functions, and their detailed procedures are described. The comparison of computational complexities for video retargeting is shown in Section VI. Simulation results and discussions are presented in Section VII. Finally, Section VII concludes this paper and lists the future work.

### II. REVIEW OF RELATED WORK

There have been several published methods and algorithms for image and video retargeting in the last decade. The simplest and most intuitive ways for retargeting are scaling and cropping. For example, Liu et al. [4] proposed an optimization process to balance the loss of details due to scaling and the loss of contents caused by cropping for video retargeting. They also introduced the virtual pans and cuts to ensure cinematic plausibility. In [5], Santella et al. proposed a gaze-based image retargeting scheme, which uses the fixation data to identify the regions with important content and computes the best crop for
any given aspect ratio or size.

Warping is another way to achieve retargeting, which introduces non-homogeneous scaling to different regions of an image or video frame based on the region importance. In the work proposed by Wolf et al [6], the importance of each region in an image is first determined by local saliency, face detection, and motion detection. Then, a transformation is performed to shrink the less important regions more than the important ones. Wang et al. [7] proposed a scale-and-stretch warping method for image retargeting through iteratively computing the optimal local scaling factors for each region and updating a warped image to match these scaling factors as close as possible.

The seam carving algorithm, proposed by Avidan et al. [1], is also a popular method for image retargeting. After important analysis on each pixel, the dynamic programming algorithm is performed to search the one-pixel width connected seams (either horizontally or vertically) that pass through pixels with the minimum summed importance. By gradually carving these seams out or inserting them in, the original image could be reduced or enlarged into the desired size. Later, Rubinstein et al. exploited the seam carving concept and used the graph-cut algorithm to perform video retargeting [2], [8]. They further attempted to combine several resizing and retargeting methods for better retargeting performances in [9].

In this paper, we mainly discuss the carving-out case, while it can be easily generalized to all to the insertion case.

There are also other image or video retargeting techniques in the existing literature, such as the shift map [10], [11], motion-aware temporal coherence [12], crop-and-warp optimization [13], importance filtering [14], and scene-aware carving [15]. In Fig. 1, image resizing results based on different retargeting methods are presented and compared.

III. COMBINATION OF FORWARD AND BACKWARD ENERGIES FOR SEAM CARVING
The original seam carving technique proposed by Avidan et al. [1] uses gradients as the measure of pixel importance (energy function), and performed dynamic programming to determine the seams for achieving image retargeting. This type of energy function prevents characteristic pixels such as corners and edges from removing, while does not ensure the image plausibility after resizing. To overcome this problem, Rubinstein et al. [2] proposed the forward energy concept, which (in the dynamic programming process) minimizes the induced energy (artifacts) after resizing rather than the content loss computed on carved-out pixels. For convenience, the first type of energy is called the backward energy in this paper.

In our implementation, spatial gradients, saliency map [16], and face detection [17] are combined in the backward energy term, and the forward and backward energy terms are jointly considered to determine the pixel importance for image $I$: 

Fig. 1. The comparison among different image retargeting / resizing techniques: (a) The original image. (b) The desired image size with a half width. (c) Horizontal scaling. (d) Direct cropping. (e) Aspect-ratio preserving scaling with the black box. (f) Manually cropping and scaling. (g) Warping based on [6]. (h) Vertical seam carving [1].
Fig. 2. Energy functions used in the backward energy term: (a) The original image. (b) The gradients map. (c) The saliency map. (d) The original image with human faces. (e) The face detection map. Each detected face is marked as a circular region with high energy. The bottom part of each face is also given some energy to prevent face-body discontinuity.

Fig. 3. The image retargeting results using different energy functions: (a) The original image. (b) Backward energy only. (c) Forward energy only. (d) The combined energy ($w_f = 1, w_s = 1,$ and $w_a = 2$).

\[
E(x, y) = w_f \cdot \text{face}(x, y) + w_s \cdot \text{saliency}(x, y) + |\frac{\partial}{\partial x} I(x, y)| + |\frac{\partial}{\partial y} I(x, y)| + w_a \cdot \text{forward energy}(x, y)
\]

where the symbols $w_f, w_s,$ and $w_a$ denote the relative weight of each energy component corresponding to the gradients. The combined energy term $E(x, y)$ then could be fed into the dynamic programming algorithm for image retargeting. In Fig. 2, we show the map of each energy component in the backward
IV. VIDEO CARVING BASED ON TEMPORAL INFORMATION AND MOTION

WEIGHT PREDICTION

A. The Original Video Carving Methods

The original seam carving for image retargeting has been extended for video retargeting by Rubinstein et al. in [2], where three different schemes are proposed and compared. The first scheme is the frame-by-frame seam carving: That is, each frame is processed individually through dynamic programming to meet the desired size. The second one is called the static seam carving, which first computes a global energy map from the whole video clip, and then determines a set of static seams used for all video frames. In their claim, the frame-by-frame method does not consider the temporal information and results in serious artifacts (the seam shaking effect). The second method further regards this information and performs well in slow motion videos, whereas the static constraint may not suitable for the high motion cases (the distorting mirror effect).

To deal with these problems, they proposed the third scheme, which models each pixel as a node and each energy measure between pixels as an edge, and uses the graph-cut algorithm to solve the optimization problem for finding a set of continuous seam surfaces in a video sequence. This method does allow seams to shift between consecutive frames along the seam surface, whereas the movement is restricted to at most one pixel and still not suitable for high motion videos (yet performs like the static method). Besides, the graph-cut algorithm is slow, especially on the domain of videos usually with more than a billion of nodes. Furthermore, this algorithm has to be repeatedly performed (each time for only one seam surface) to achieve the desired display size, which is very time consuming.
In order to speed up the optimization procedure, release the restriction of seam movement, and produce better retargeting results, a new video carving scheme based on temporal information and dynamic programming is proposed. The new scheme searches a seam sequentially in each frame by dynamic programming and propagates the detected seam to the next frame through pixel and object movement. Although this procedure needs repeatedly performed to meet the desired size, dynamic programming does require much less computational complexity than the graph-cut algorithm. The flowchart of the proposed method is depicted in Fig. 4.

B. The Concept of Motion Weight Prediction

The proposed scheme is based on an intuition, where pixels (or objects) removed in the previous frame should also be removed in the current frame. To achieve this goal, the motion vectors, which can be easily acquired from the compressed video streams, are exploited to estimate the pixel or object movement. Although the motion vectors are sometimes inaccurate, they do provide a coarse prediction for the seam locations in the current frame according to seams detected in the previous frame. To avoid confusion, the frame undergoing the seam carving process is denoted as Frame $t$ (the current frame).

C. The Seam Discontinuity Problem and the Idea of Gaussian Masks

The motion vectors acquired from the video streams are usually block-based. Directly using these vectors to determine the seam locations between frames will probably result in a discontinuous seam,
Fig. 5. The problem of seam discontinuity: The left-hand rectangle shows the detected seam and the block-based motion vectors in Frame $t-1$ and the right-hand one shows the discontinuous seam in Frame $t$ determined by the motion vectors.

which disobeys the basic concept of seam carving [1]. In addition, this process does not consider the image content in Frame $t$ and may cause serious image artifacts. The problem is illustrated in Fig. 5.

To solve this discontinuity problem, we propose the motion vector prediction function, which uses the motion vectors as a guideline, not a determination quantity for finding seams in Frame $t$. The motion weight prediction sets a unit Gaussian mask on each estimated pixel in Frame $t$ (the pixels marked by the green segments in Fig. 5) to define a range of seam locations, where the surrounding pixels also receive motion weights based on their distances to these estimated pixels (this can be accomplished through convolution). The use of Gaussian masks generates a motion weight map, where the weight on each pixel indicates the prediction accuracy of the seam location. To ensure the seam continuity and also consider the image content in Frame $t$ during seam carving, the image energy map in (1) is subtracted by this motion weight map, and dynamic programming is performed to find a continuous seam, which will probably go through pixels with high motion weights. To further speed up the seam carving process in Frame $t$, the dynamic programming algorithm is only performed in the range centered by the seams in Frame $t-1$ plus a motion offset.

D. The Motion Mismatch Problem

The video carving scheme just mentioned seems reasonable, whereas an important issue of motion vectors has not regarded. The motion vectors are block-based and the motion strengths are measured in the original frame size.
Fig. 6. The problem of motion mismatch: The left-hand rectangle shows the original frame size and the motion blocks. The right-hand one illustrates the frame with \(k\) seams removed and the corresponding shrunk blocks.

Fig. 7. The idea of the location index matrix: The top row denotes the original frame and the original index map. The bottom row then shows the resized ones. As marked by the blue rectangle, the resized index matrix records what pixels of the original frame have been remained in the resized frame.

While during the repeated video carving procedure, the frame size is iteratively changed, and the motion strengths cannot be directly exploited to guide the seam location. Moreover, since the carving procedure will unequally remove pixels in each motion block, after carving out several seams, determining the corresponding motion vector for a detected seam pixel in Frame \(t-1\) gradually becomes a problem. This motion mismatch issue is depicted in Fig. 6.
E. The Modified Motion Weight Prediction Function

To solve this crucial problem, a modified version of the motion weight prediction function is proposed. In the beginning of video retargeting, a 3-dimensional location index matrix is defined for each video frame. This matrix is initialized to have the same width and height of the original frame, while contains only two channels in the third dimension to record the location index (x and y) of the corresponding image pixel. After detecting a seam in one frame, both the frame and the accompanied index matrix will be resized through carving out pixels on the seam location, which means the index matrix will record what pixels of the original video frame are still remained in the retargeted video frame. This idea is illustrated in Fig. 7.

The index matrix conveys important information to overcome the scale mismatch problem. Assume now a seam has been detected in Frame t-1, the index matrix of Frame t-1 can determine where these seam pixels are located in the original frame size (original seam). Adding the corresponding motion vectors on the original seam location estimates the discontinuous seam location for Frame t, still in the original frame size. Then, we perform the Gaussian mask on these estimated seam pixels to produce the weight map, and resize this weight map into the current size of Frame t through the index matrix of Frame t (the index matrix can denotes what pixels as well as motion weights of the original Frame t are remained in the current frame size). Finally, subtracting the image energy map in (1) of the current-sized Frame t by this reduced weight map and performing the dynamic programming algorithm in the shrunk search range, the appropriate seam of Frame t in the current size can be determined. The detailed implementation of the modified motion weight prediction is described in Table I, and the flowchart is depicted in Fig. 8.
Fig. 8. The flowchart of the modified motion weight prediction function: Details of Steps (i)-(iv) are described in Table I.

### TABLE I

**Algorithm of the Modified Motion Weight Prediction.**

**Presetting and Concepts:**

- Assume that the video contains $T$ frames, where each frame is of height $U$ and width $V$, and there are totally $K$ seams should be carved out to achieve the desired video size.
- Perform width reduction.
- Denote the Frame $t$ and the corresponding location index map with $k$ seams removed by $F_k(x, y, t)$ and $L_k(x, y, t)$, respectively.
- Denote the motion vectors between Frame $t-1$ and Frame $t$ as $M(t)$. 
Denote the image energy map of Frame $t$ with $k$ seams removed as $E_k(x, y, t)$. The energy map is generated based on the combined fashion in (1) in Section III.

After $k$ seams are removed, $L_k(x, y, t)$ records the pixels of the original Frame $t$ that are still remained in $F_k(x, y, t)$.  

### Step (i)

**Procedure and Algorithm:**

**for $k = 1: K$**

--Do seam carving in $F_{k-1}(x, y, 1)$ based on $E_{k-1}(x, y, 1)$ to produce $F_k(x, y, 1)$ and denote the removed seam as $s_k(j, 1)$.

**for $t = 2: T$**

--Use the location map $L_k(x, y, t-1)$ of Frame $t-1$ to get the original seam location $S_k(j, t-1)$ based on $s_k(j, t-1)$ in the original frame.  

**Step (ii)**

--Use the motion vectors $M(t)$ and $S_k(j, t-1)$ to predict the estimated seam location $S_j(j, t)$ of Frame $t$ in the original frame size.

--Compute the horizontal variance $\sigma$ and mean $\mu$ of the motion vectors $M(t)$ along the pixels on $S_k(j, t-1)$ if the corresponding pixel on $S_k(j, t)$ is not outside the original frame size. Offset $s_k(j, t-1)$ by $(V-k+1) / V \cdot \mu$ and set the dynamic programming (DP) range centered around the offset seam location with a width based on $\sigma$. **Step (iii)**

**if (over 1/3 of pixels on $S_k(j, t)$ are outside the original frame size)**

--Do seam carving in $F_{k-1}(x, y, t)$ based on $E_{k-1}(x, y, t)$ within the shrunk DP range to produce $F_k(x, y, t)$ and denote the seam as $s_k(j, t)$. 

else

--Generate the weight map with the Gaussian mask in the original frame size and use
$L_k(x, y, t)$ to resize the weight map into the frame size of $F_k(x, y, t)$. Denote the
reduced weight map as $W_k(x, y, t)$. \textbf{Step (iv)}

--Do seam carving on $F_k(x, y, t)$ based on the new energy term: \textbf{Step (v)}

$$E_{k-1}(x, y, t) - w_m \cdot W_{k-1}(x, y, t)$$

within the shrunk DP range to produce $F_k(x, y, t)$, and denote the seam as $s_k(j, t)$.

end

end

end

Based on the modified motion weight prediction function and procedure, the motion vectors can be
exploited in the correct strength for estimating rough seam location. Although the estimated seam pixels
have the risk to be carved out in Step (iii) in Fig. 8, the Gaussian mask can diffuse their weights to nearby
pixels. Then after reducing the weight map, regions with high motion weight can still be found. In
addition, if the estimated seam pixels in Step (ii) are out of the original frame size, it possibly means that
the regions or objects containing these seam pixels are moving out in Frame $t$. For this special case, the
reduced weight map is not considered for seam carving in Frame $t$, but the dynamic programming is
stilled performed in the shrunk search range with an enlarged width.

V. \textbf{PIXEL-BASED OPTIMIZATION}

A. \textit{From Coarse Prediction to Fine Detection}

The modified motion weight prediction function allows more freedom on seam shifting between
consecutive frames, and meanwhile attenuates the seam shaking artifact resulted from the frame-by-frame
seam carving through temporal information propagation. However, the inaccurate nature of motion vectors may lead to wrong motion estimation, and the error produced between two frames will propagate to frames after them. Besides, motion vectors can only provide coarse information about pixel and object movement, and a finer temporal energy term to accurately judge the motion and temporal coherence should be derived.

B. The Seam Difference Energy

As mentioned in Section III, the basic intuition of our work is to remove the same pixel or same objects in consecutive frames. To better estimate the motion, we utilize the concept of pixel-based optical flow to generate a new energy term called seam difference. The idea of optical flow is to approach the following equalities for accurate motion estimation:

\[
I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{2}
\]

\[
\frac{\partial I}{\partial x} \cdot dx + \frac{\partial I}{\partial y} \cdot dy + \frac{\partial I}{\partial t} \cdot dt = 0 \tag{3}
\]

where (3) is very similar to the spatial gradients used in (1) plus an additional temporal difference term and modulus operators. Based on this connection while not drastically increasing the computational complexity, a modified temporal difference term (seam difference) is proposed, which considers the color difference between pixels in the search range of Frame \(t\) and the detected seams in Frame \(t-1\). For width reduction, this term computes the color difference between image pixels and the seam pixels with the same vertical coordinate (x-axis); for the height reduction case, this difference is computed between pixels at the same horizontal coordinate (y-axis). Fig. 9 describes the concept of the seam difference term for the width reduction case. Denote the \(k^{th}\) detected seam in Frame \(t-1\) as \(s_{k}(j, t-1)\), the seam difference term for Frame \(t\) with \(k\)-1 seams removed is formulated as:

\[
\text{width: } \text{Seam}_\text{diff}_k(x, y, t) = |F_{k-1}(x, y, t) - s_k(x, t-1)|, \tag{4}
\]

\[
\text{height: } \text{Seam}_\text{diff}_k(x, y, t) = |F_{k-1}(x, y, t) - s_k(y, t-1)|. \tag{5}
\]
Fig. 9. The idea of the seam difference term based on the pixel-based optical flow: For the width reduction case, this seam difference is computed between the pixels in the search range of Frame $t$ and the detected seam pixels in Frame $t-1$ with the same vertical coordinate.

To further compensate the possible intensity and hue change of pixels and objects passed through the seams, a timing window, which contains at most $N$ frames (the current Frame $t$ and the preceding $N-1$ frames) can be considered. Taking the weighted sum over all the color differences computed between the current frame and the seams detected from the other frames in the timing window, a compensated seam difference term can be achieved. The weight set for each seam difference map is based on the timing distance between the corresponding frame $k$ and the current Frame $t$. Based on this idea, the seam information of the previous $N-1$ frames can be easily embedded into the seam carving process of the current Frame $t$ without exhaustively calculating the optical flow of the whole video, which significantly accelerates the retargeting process.

C. The Temporal Coherence Measurement

As denoted in Section I, when extending the image retargeting into video retargeting, not only the characteristic content of each frame is to be preserved, but the spatial coherence (removing the same pixels or objects) and the temporal coherence (avoiding the seam shaking effect) are required to achieve
for perceptual plausibility. Towards these goals, several energy terms are considered in this paper. The image energy function in (1) records the content importance of images, and the modified motion weight prediction function coarsely preserves the spatial and temporal coherence. The seam difference term just mentioned then aims to find the accurate seam location in Frame \( t \) to further improve the spatial coherence. Moreover, to better preserve temporal plausibility, the temporal coherence measurement proposed by Grundmann et al. [3] is considered in our work.

In their claim, the optimal seam in Frame \( t \) for temporal plausibility is just the seam detected in Frame \( t-1 \). While by directly doing so, the spatial coherence will be ignored. To achieve a balance, a temporal coherence measure was derived in [3]. Using the notation defined in Table I and assuming now the \( k^{th} \) seam is to be removed from \( F_{k-1}(x, y, t) \). Denote the retargeting result directly by \( s_k(j, t-1) \) as \( R_k(x, y, t) \), then for every pixel in \( F_{k-1}(x, y, t) \), a temporal artifact weight is given according to how much the retargeting result would differ from \( R_k(x, y, t) \) if that pixel were carved out. The formulation of this temporal coherence measure \( Temp_k(x, y, t) \) for both the width and height reduction cases are defined as:

**Width:**

\[
Temp_k(x, y, t) = \sum_{v=1}^{y-1} |F_{k-1}(x, v, t) - R_k(x, v, t)| + \sum_{v=y+1}^{y+k-1} |F_{k-1}(x, v, t) - R_k(x, v-1, t)|.
\]

**Height:**

\[
Temp_k(x, y, t) = \sum_{u=1}^{x-1} |F_{k-1}(u, y, t) - R_k(u, y, t)| + \sum_{u=x+1}^{x+k-1} |F_{k-1}(u, y, t) - R_k(u-1, y, t)|.
\]

where the sum-of-absolute error is used rather than the sum-of-square one mentioned in [3] to keep energy-scale matching.

**D. The Video Energy Function for Video Retargeting**

Combining all the energy terms introduced in this paper, the final video energy map used in Step (v) in Table I is denoted as:

\[
E_{video}(x, y, t) = E_k(x, y, t) - w_m \cdot W(x, y, t) + w_d \cdot Seam-diff_k(x, y, t) + w_t \cdot Temp_k(x, y, t).
\]
TABLE II

COMPARISON OF COMPUTATIONAL COMPLEXITIES

(K - NUMBER OF SEAMS, T – NUMBER OF FRAMES, N – NUMBER OF EDGES, |C| – MINIMAL CUT COST,

<table>
<thead>
<tr>
<th>Method</th>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed scheme</td>
<td>Dynamic programming</td>
<td>$O{ K \cdot T \cdot (R \cdot V) }$</td>
</tr>
<tr>
<td>Graph-based video retargeting</td>
<td>Dinic algorithm [18]</td>
<td>$O{ K \cdot N \cdot (T \cdot V^2)^2 }$</td>
</tr>
<tr>
<td></td>
<td>Push-Relabel algorithm [19]</td>
<td>$O{ K \cdot (T \cdot V^2)^2 }$</td>
</tr>
<tr>
<td></td>
<td>Algorithm proposed in [20]</td>
<td>$O{ K \cdot</td>
</tr>
</tbody>
</table>

VI. COMPUTATIONAL COMPLEXITY

We list the computational complexity of the proposed video retargeting scheme and the min-cut / max-flow algorithm used in [2] for graph cut in Table II. The symbol $K$ stands for the number of seams to remove, $T$ for the number of frames in a video clip, $N$ for the number of edges defined in the graph, and $|C|$ for the cost of the minimum cut. To derive a unified complexity for both the width- and height-reducing cases, we simply assume that width and height of the original video are equal to $V$. Then the number of pixel in each original video frame is $V^2$. The symbol $R$ shown in the complexity of the proposed method is the width of the shrunk search range, where $R < V$. Because the video carving method in [2] models each pixel as a node and the whole video as a graph, the total number of nodes in the graph is $T \cdot V^2$. From the view of computational complexity, the proposed scheme achieves significant speed-up against the use of the graph-cut algorithm. Although with a limited number of node (such as an image), the graph-cut algorithm can run fast. When the graph size is multiplied by the number of frames for video retargeting (a 10-second video clip contains 300-600 frames), the complexity will drastically increase and the retargeting process becomes very slow.

In addition, the graph-cut-based video retargeting has to perform the graph-cut algorithm for several times to reach the desired video size, where each time a seam surface is detected to remove one-pixel
width or height from the video. When on-line video retargeting is required, this repeated procedure cannot be used. The proposed video retargeting is also a repeated scheme as described in Table I. While with the order change on the two for-loop operations, our retargeting scheme can sequentially resizes each frame directly to the desired size and has the potential to support on-line video retargeting.

VII. SIMULATION RESULTS AND DISCUSSIONS

A. Motion-Guided Seam Propagation

The proposed video retargeting method is examined on several standard videos such as football, coastguard, Stefan, and harbour, which all contain object motion or background motion. At first, we check the seams detected in two different frames based on the frame-by-frame method [2], static method [2], and the proposed method. The results shown in Fig. 10 demonstrate that the seams between frames are guided by the motion information in the video, which releases the movement constraint from the static method (as well as the graph-cut-based method with only one-pixel shifting allowed) and attenuates the seam shaking effect from the frame-by-frame method.

B. Video Retargeting Results

The video retargeting results of the proposed scheme and other existing methods are presented and compared in Fig. 11. In our implementation, the weight set for each image energy term in (1) is $w_f = 1$, $w_v = 1$, and $w_a = 2$; the weight set for each video energy function in (8) is: $w_m = 0.5$, $w_d = 1$, and $w_c = 0.5$. From these experiments, the proposed video retargeting scheme preserves more important content and compositions against the scaling and cropping methods. Moreover, when compared with the static method, which is also based on the dynamic programming algorithm, the proposed method is more content-aware while also achieves temporally smoothing. Especially, for the harbour video in the third row of Fig. 11, there are a lot of straight lines and other algorithms usually perform poorly. However, the proposed video retargeting scheme does preserve the characteristic objects and the straight lines in this video.
Fig. 10. The comparison of seam locations detected in consecutive frames. The left-hand column is Frame 105 from the *stefan* video and the right-hand column is Frame 130. The first 20 detected vertical seams are shown and marked by red lines. Results in (a) and (b) are from the frame-by-frame method (with serious seam shaking). Results in (c) and (d) are from the static method where seam movement is not allowed. Results in (e) and (f) are from the **proposed method** where the seams are guided by the motion vectors, the background, and object movement tracking.
Fig. 11. The comparison of the video retargeting results based on the standard coastguard (row 1 on top), stefan (row 2), harbour (row 3), and football (row 4) videos. The first column (the left-hand column) shows the original frame (all with size $352 \times 288$), and right-hand four columns depict the retargeting results with size $240 \times 288$ based on cropping (column 2), scaling (column 3), static seam carving [2] (column 4), and the proposed video carving scheme (column 5). The red circles mark the important content preserved by the proposed scheme against other retargeting methods.

C. The Distorting Mirror Effect

Although the retargeting results based on the static method sometimes looks plausible from the view of still images, the distorting mirror effect occurring when objects or parts of the background shift through these static seams results in serious spatial artifacts. Fig. 12 illustrates this effect, where the width of the rectangle bounding the same object has significant changed between consecutive frames.
Fig. 12. The distorting mirror effect caused by the static seam carving: The left-hand column is Frame 80 in the stefan video; the right-hand column is Frame 90, both with 64 seams carved out. The red rectangles bound the same object (a person) in each frame. As shown, the retargeting results based on the static method (at the top row) are with serious spatial artifacts between consecutive frames in (a) and (b), while the proposed method (at the bottom row) significantly attenuates this problem.

The graph-cut-based method will also cause this artifact due to the one-pixel shifting constraint. As shown, the proposed method can definitely reduce this problem and equally maintain the size of objects in frames.

D. The Seam Carving Strategy in the First Frame

The performance of the proposed method heavily depends on the seams detected in the first frame, where the seams carved out in the coming frames are based on the locations of these first-frame seams. In some cases where the main objects of a video does not appear in the first frame while suddenly move in
Fig. 13. The seams removed in the first video frame will propagate to the following frame based on the proposed scheme. The left-hand column shows the first 20 vertical seams detected in Frame 1 in the coastguard video though only considering the content in Frame 1 (top row) and through finding the static seams in the first 150 frames (bottom row). The right-hand column then shows the seams detected in Frame 30, where a white boat drives in from the left frame boundary and the background (grass and the river) moves to the right (so the seams shift to the right). As shown, the first-frame seams determined by the static method can avoid passing through important objects in the coming few frames.

In the following few frames, seams detected in the first frame has risks passing through these objects if they are located around the frame boundary. Fig. 13 demonstrates this potential case. To solve this problem, the idea of the static method is utilized in our scheme, where the seams in the first frame are determined by the static seams found from the first $P$ frames ($P$ is set to 150 in our experiments). Through this modified procedure, the content information of the first $P$ frames is jointly considered, and the seams detected in the first frame are more likely to keep away from the path of the moving-in objects (as illustrated in Fig. 13). This strategy is exploited in the retargeting results shown in Fig. 11 for the coastguard and stefan videos.
Fig. 14. The video retargeting result on a simple background video with strong motion (the baseball video, 360 × 240). The original frame is shown in (a), and the retargeting results (240 × 240) based on the static seam carving method [2] and the proposed scheme are presented in (b) and (c).

E. Limitation

The proposed method faces some limitations when the frame background is fairly simple while the motion is highly strong in the video, which drastically decreases the accuracy of motion vectors. Fig. 14 shows an example with the line distortion artifact.

VIII. CONCLUSION AND FUTURE WORK

In this paper, a new video retargeting scheme is proposed by combining the content and temporal information with the seam carving technique. This scheme is based on two new video energy functions: motion weight prediction and pixel-based optimization. Motion weight prediction utilizes the motion vectors to coarsely guide the seam shifting among frames and makes the dynamic programming algorithm available for video retargeting. To solve the possible seam discontinuity and motion mismatch problems, the Gaussian masks and the location index map are introduced to modify the motion weight prediction function and reduce the search range of dynamic programming. Pixel-based optimization further exploits the concept of pixel-based optical flow and temporal coherence measurement to refine the video energy term and preserve the spatial and temporal coherence for seam carving.
Compared to the existing video retargeting schemes, the proposed scheme could
(a) effectively use the motion vectors contained in the compressed video stream,
(b) release the restriction of seam movement to attain more flexible retargeting results,
(c) reduce the search range of dynamic programming, and
(d) achieve lower computational complexity through dynamic programming.

Moreover, from the simulation results on several standard videos, the proposed method can achieve
not only content-aware, but also spatial and temporal smoothing retargeting results against other existing
method such as scaling, cropping, and static seam carving, etc.

For future work, we aim to solve the artifacts resulted from inaccurate motion vectors in
simple-background videos, combine several video retargeting algorithms for performance improvements,
and generate an automatic mechanism for determining the weight set for each energy term.

**Reference**


Figure and Table Captions

Fig. 1. The comparison among different image retargeting / resizing techniques: (a) The original image. (b) The desired image size with a half width. (c) Horizontal scaling. (d) Direct cropping. (e) Aspect-ratio preserving scaling with the black box. (f) Manually cropping and scaling. (g) Warping based on [6]. (h) Vertical seam carving [1].

Fig. 2. Energy functions used in the backward energy term: (a) The original image. (b) The gradients map. (c) The saliency map. (d) The original image with human faces. (e) The face detection map. Each detected face is marked as a circular region with high energy. The bottom part of each face is also given some energy to prevent face-body discontinuity.

Fig. 3. The image retargeting results using different energy functions: (a) The original image. (b) Backward energy only. (c) Forward energy only. (d) The combined energy ($w_b = 1$, $w_s = 1$, and $w_a = 2$).

Fig. 4. The flowchart of the proposed video carving procedure.

Fig. 5. The problem of seam discontinuity: The left-hand rectangle shows the detected seam and the block-based motion vectors in Frame $t-1$ and the right-hand one shows the discontinuous seam in Frame $t$ determined by the motion vectors.

Fig. 6. The problem of motion mismatch: The left-hand rectangle shows the original frame size and the motion blocks. The right-hand one illustrates the frame with $k$ seams removed and the corresponding shrunk blocks.

Fig. 7. The idea of the location index matrix: The top row denotes the original frame and the original index map. The bottom row then shows the resized ones. As marked by the blue rectangle, the resized index matrix records what pixels of the original frame have been remained in the resized frame.

Fig. 8. The flowchart of the modified motion weight prediction function: Details of Steps (i)-(iv) are described in Table I.
Fig. 9. The idea of the seam difference term based on the pixel-based optical flow: For the width reduction case, this seam difference is computed between the pixels in the search range of Frame $t$ and the detected seam pixels in Frame $t-1$ with the same vertical coordinate.

Fig. 10. The comparison of seam locations detected in consecutive frames. The left-hand column is Frame 105 from the *stefan* video and the right-hand column is Frame 130. The first 20 detected vertical seams are shown and marked by red lines. Results in (a) and (b) are from the frame-by-frame method (with serious seam shaking). Results in (c) and (d) are from the static method where seam movement is not allowed. Results in (e) and (f) are from the proposed method where the seams are guided by the motion vectors, the background, and object movement tracking.

Fig. 11. The comparison of the video retargeting results based on the standard *coastguard* (row 1 on top), *stefan* (row 2), *harbour* (row 3), and *football* (row 4) videos. The first column (the left-hand column) shows the original frame (all with size $352 \times 288$), and right-hand four columns depict the retargeting results with size $240 \times 288$ based on cropping (column 2), scaling (column 3), static seam carving [2] (column 4), and the proposed video carving scheme (column 5). The red circles mark the important content preserved by the proposed scheme against other retargeting methods.

Fig. 12. The distorting mirror effect caused by the static seam carving: The left-hand column is Frame 80 in the *stefan* video; the right-hand column is Frame 90, both with 64 seams carved out. The red rectangles bound the same object (a person) in each frame. As shown, the retargeting results based on the static method (at the top row) are with serious spatial artifacts between consecutive frames in (a) and (b), while the proposed method (at the bottom row) significantly attenuates this problem.

Fig. 13. The seams removed in the first video frame will propagate to the following frame based on the proposed scheme. The left-hand column shows the first 20 vertical seams detected in Frame 1 in the *coastguard* video though only considering the content in Frame 1 (top row) and through finding the static seams in the first 150 frames (bottom row). The right-hand column then shows the seams detected in Frame 30, where a white boat drives in from the left frame boundary and the background (grass and the river) moves to the right (so the seams shift to the right). As shown, the
first-frame seams determined by the static method can avoid passing through important objects in the coming few frames.

Fig. 14. The video retargeting result on a simple background video with strong motion (the *baseball* video, 360 × 240). The original frame is shown in (a), and the retargeting results (240 × 240) based on the static seam carving method [2] and the proposed scheme are presented in (b) and (c).

**TABLE I** Algorithm of the modified motion weight prediction.


The proposed scheme has the least complexity.