COARSE-TO-FINE TEMPORAL OPTIMIZATION FOR VIDEO RETARGETING BASED ON SEAM CARVING

Wei-Lun Chao[1], Hsiao-Hang Su[2], Shao-Yi Chien[2], Winston Hsu[3], and Jian-Jiun Ding[1]

[1]Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
[2]Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan
[3]Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan
E-mail: weilunchao760414@gmail.com, waterhangg@media.ee.ntu.edu.tw, sychien@cc.ee.ntu.edu.tw, winston@csie.ntu.edu.tw, djj@cc.ee.ntu.edu.tw

ABSTRACT

In this paper, a new video retargeting method based on temporal information and seam carving is presented. Two video energy functions, motion weight prediction and pixel-based optimization, are proposed to take the temporal information into account and make dynamic programming available during the process of retargeting. The motion weight prediction exploits both the block-based motion estimation and Gaussian masks to predict the coarse location of seams in the current frame and reduce the search range of dynamic programming. The pixel-based optimization then utilizes the concept of pixel-based optical flow to explore better temporal relations between the current frame and previous frames in the reduced search range. The experimental results show that combining these two video energy functions as well as dynamic programming, the proposed method could achieve content-aware and temporal smoothing retargeting results with less computational complexity.

Index Terms— retargeting, seam carving, motion estimation, optical flow

1. INTRODUCTION

With the fast development of technology, portable displayers have gradually become an indispensable part in human life, such as mobile phones, PDAs, e-book readers, and tablet computers. Different display units may have different sizes and aspect ratios, while each image or video clip is usually made with a single size, so how to resize these multimedia contents into the desired display size has been an important issue of multimedia researches in recent years. Among 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 features intact and minimizing the important information loss. Image retargeting is generally composed of two steps: importance analysis of image content (also called energy function), and image resizing. The first step exploits several concepts of image content analysis such as saliency map, gradients, and object recognitions 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. For video retargeting, how to take the temporal information (such as motion vectors) into consideration is still a hot topic in this area.

In this paper, we exploit the popular seam carving technique [1] for image and video retargeting. For image retargeting, the forward and backward energy terms are properly combined during the seam carving process, which determines the importance of image pixels not only based on their gradients and saliencies but also on the inserted energy after removing each pixel. For video retargeting, a new carving procedure considering temporal information is presented. We proposed two video energy functions, **motion weight prediction** and **pixel-based optimization**, to make dynamic programming available in video retargeting. The **motion weight prediction** utilizes the block-based motion vectors, which can be easily acquired from compressed video streams, to guide the position of seams in each frame, and the **pixel-based optimization** exploits the idea of optical flow to determine the accurate seam location during the process of dynamic programming. The new procedure could effectively utilize the motion vectors, speed up the original video carving procedure which is based on the graph-cut algorithm proposed by Rubinstein et al. [2], and release the restriction introduced in their works where the seam in the next frame can only move one pixel away from the seam location in the current frame. The experimental results showed that our modification improved both the image and video retargeting results.

This paper is organized as follows. In Section 2, related works about image and video retargeting are briefly reviewed, and the combination of forward and backward

978-1-61284-350-6/11/$26.00 ©2011 IEEE

energy terms for image retargeting is described in Section 3. In Section 4, the proposed video retargeting method and its detailed procedure is presented, and Sections 5 and 6 show our experimental results as well as the comparison of computational complexities. Finally, Section 7 concludes this paper and lists the future works.

2. RELATED WORKS

There have been several proposed works in the last decade toward image and video retargeting. The simplest and most intuitive ways for retargeting are scaling and cropping. Liu et al. [4] proposed an optimization process to minimize information loss by balancing the loss of details due to scaling with the loss of content and composition due to cropping for video retargeting. They also introduced virtual pans and cuts to ensure cinematic plausibility. Santella et al. [5] proposed a gaze-based image retargeting scheme which uses fixation data to identify important content and compute the best crop for any given aspect ratio or size.

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

Recently, Avidan et al. [1] proposed a new way of thinking towards image retargeting, the seam carving. After importance analysis, their algorithm finds the one-pixel width connected seam either horizontally or vertically with the minimum 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. [2] exploited the seam carving concept and used the graph-cut algorithm to perform video retargeting. They further attempted to combine several resizing techniques for better retargeting performance in [3]. In Fig. 1, image resizing results based on different retargeting methods are presented.

3. COMBINATION OF FORWARD AND BACKWARD ENERGY FOR SEAM CARVING

The original seam carving technique proposed by Avidan et al. [1] used gradients as the measure of pixel importance (energy function) and then performed dynamic programming to achieve retargeting. This type of energy function preserves characteristic pixels such as corners and edges from carving-out, while doesn't ensure the plausibility after resizing. Rubinstein et al. [2] further proposed the forward

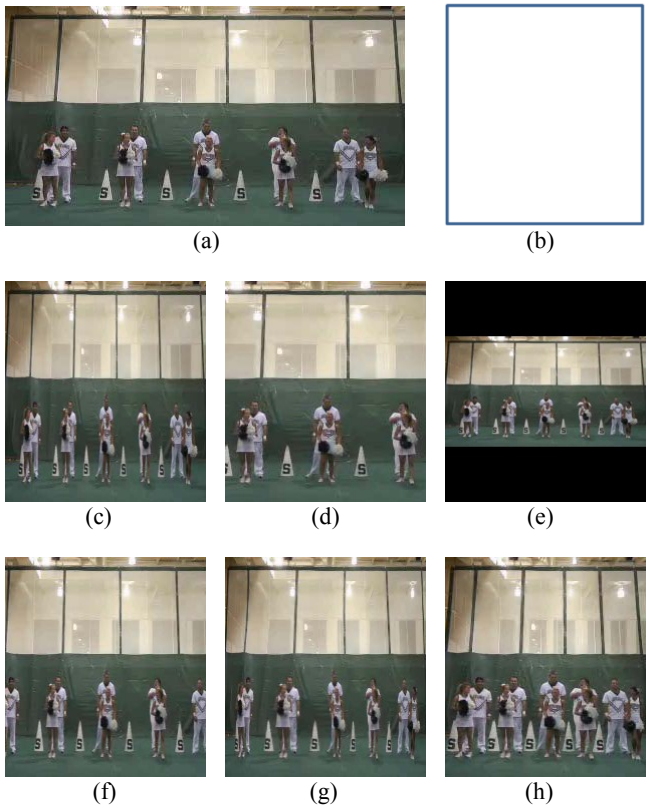


Fig. 1. A comparison among different image retargeting / resizing techniques. (a) The original image. (b) The desired image size with a half width. (c) Horizontal scaling. (d) Direct cropping. (e) Aspect-ratio preserving scaling with black box. (f) Manually cropping and scaling. (g) Warping based on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

energy concept which minimizes the energy after resizing (the inserted energy) rather than minimizing the energy before resizing (the carved-out energy). For convenience, the first proposed energy term is called the backward energy in this paper.

In our implementation, gradients, saliency map [8], and face detection [9] are considered in the backward energy term, and both the forward and backward energy terms are combined during resizing:

$$e(\mathrm{I}) = w_f \cdot face(x,y) + w_s \cdot saliency(x,y) + \left|\frac{\partial}{\partial x}\mathrm{I}(x,y)\right| + \left|\frac{\partial}{\partial y}\mathrm{I}(x,y)\right| + w_c \cdot foward\ energy(x,y), \quad (1)$$

where w_f is the weight of face information, w_s is the weight of saliency information, and w_c is the weight of forward energy. In our experiments, we set $w_f = 1$, $w_s = 1$, and $w_c = 2$. Fig. 2 shows the map of each composition energy function used in the backward energy term, and Fig. 3 compares the retargeting results based on forward, backward, and combined energy. This combined energy function is further used in our video retargeting algorithm.

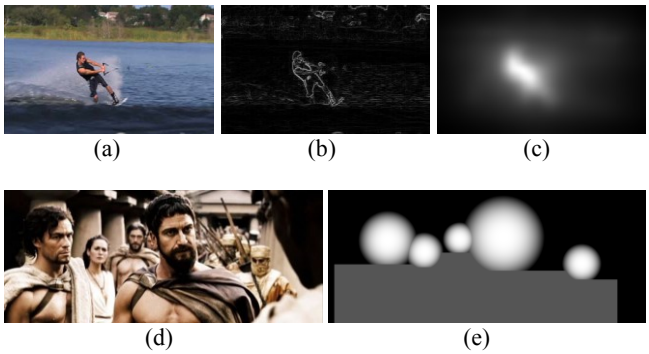


Fig. 2. Energy functions used in the backward energy term. (a) The original image. (b) The gradients map. (c) The saliency map. (d) The original image with human faces. (e) The face detection map. Each detected face is marked as a circular region with high energy, and the bottom part of each face is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

Fig. 3. The image retargeting results of different energy functions. (a) The original image. (b) Backward energy only. (c) Forward energy only. (d) The combined energy.

4. VIDEO CARVING BASED ON COARSE-TO-FINE TEMPORAL OPTIMIZATION

The seam carving technique has been extended for video retargeting by Rubinstein et al. [2], where they denoted each pixel as a node, each energy term between pixels as an edge, and used the graph-cut algorithm to solve the optimization problem of finding a seam surface in a video sequence. This procedure is rather slow and required to perform several times to achieve the desired size. In addition, the graph-cut framework only allows the seam in the current frame to move by at most one pixel in the next frame, which is not suitable for high-motion video and will result in serious discontinuity. In order to speed up the optimization procedure and release the restriction of seam locations, a new video carving scheme based on the motion weight prediction and pixel-based optimization aiming to minimize the artifacts caused by motion between two consecutive frames is proposed.

4.1. Motion Weight Prediction

4.1.1. The idea of using motion vectors

Our work is based on an intuition, where pixels removed in the previous frame should also be removed in the current frame for video plausibility. Motion vectors can be easily acquired from compressed video stream. Although it is not accurate, it did provide a guideline for seam locations in the current frame based on seams in the previous frame. In the first frame of a video clip, the image seam carving is performed to find a seam for removing, and then the motion vectors are used to estimate where each pixel of the currently detected seam will be located in the next frame. Because the motion vectors are estimated block by block (usually based on 8-by-8 block size), the estimated seam will possibly be disconnected.

To solve this problem, the estimated seam based on the previous frame is not directly carved out in the current frame, but used as the guideline of the seam location. We give each pixel on the estimated seam a unit motion weight and a 2-D Gaussian mask, where the surrounding pixels also receive weights based on their distances to the estimated seam (which can be achieved through convolution). After this motion weight diffusion procedure, a weight map is generated, and the energy function map defined in (1) of the current frame is subtracted by this weight map. Then during the seam carving process at the current frame, the detected seam will probably go through pixels with high motion-estimated weight. In addition, dynamic programming in the current frame is only performed around the seam location of the previous frame plus a motion offset for speed-up.

The above procedure seems reasonable, but there is a serious problem of using motion vectors. The motion vectors are block-based and the strength of motion is computed based on the scale of the original frame size. While carving the k^{th} seam (totally k-1 seams have been removed in each frame before), the size of the current frame is different from the size of the original video, so the motion strength cannot be directly exploited because of scale mismatching. In order to deal with this problem, a modified procedure is proposed to take the number of removed seams into consideration. Fig. 4 shows the flowchart of our proposed procedure, and the details are explained in Table 1.

4.1.2. The modified procedure of using motion vectors

Based on the procedure described in Table 1, the motion vectors can be exploited in the correct strength scale. Although the estimated seam pixels in the original frame size have risk carving out in Step (iii), the Gaussian mask can diffuse their weights to the nearby pixels, so after reducing the weight map into $W(k\text{-}1, t)$, we can still find regions with high motion-estimated weights. Subtracting the combined energy map in (1) by the reduced weight map, the new energy function can be formulated as:

$$e'(\mathrm{I}) = e(\mathrm{I}) - \mathit{reduced\ weight\ map}(x, y). \qquad (2)$$

If the estimated seam pixels are out of the original frame size, it possibly means that the region containing the previous seam moves out in the current frame, so we just

Table 1. The modified procedure of using motion vectors

- Assume the video contains T frames, where each frame is of height h and width w, and totally K seams should be carved out to achieve the desired video size.
- Perform **width-reducing** in this case.
- Denote the frame t with k seams removed as $F(k, t)$.
- Denote the motion vectors between frame t-1 and t as $M(t)$.
- Denote the energy function map of the frame t with k seams removed as $E(k, t)$. The energy function map is generated based on the combined fashion (1) in Section 3.
- The removed seams are recorded for each frame. **(i)**
- A location map $L(k, t)$ is created for frame t to record the relationship of each pixel in the k-seam removed frame and its location in the original frame.

for k = 1: K
-Perform seam carving in $F(k\text{-}1, 1)$ based on $E(k\text{-}1, 1)$ to produce $F(k, 1)$ and denote the removed seam as $s(k, 1)$.

for t= 2: T
-Use the location map $L(k, t\text{-}1)$ of frame t-1 to get the original locations of pixels on $s(k, t\text{-}1)$. Denote this original location sequence as $S(k, t\text{-}1)$. **(ii)**

-Use the motion vectors $M(t)$ and $S(k, t\text{-}1)$ to predict the estimated seam location of frame t in the original frame size.

if (more than 1/3 pixels of the estimated seams are out of the original frame size)
-Perform seam carving in $F(k\text{-}1, t)$ based on $E(k\text{-}1, t)$ to produce $F(k, t)$ and denote the removed seam as $s(k, t)$.

else
-Generate the weight map based on the Gaussian mask in the original frame size, and use the recorded seams of frame t to reduce the weight map into the frame size with k-1 seams removed. Denote the reduced weight map as $W(k\text{-}1, t)$. **(iii)**

-Compute the variance and mean of the motion vectors $M(t)$ along the horizontal direction for pixels on $S(k, t\text{-}1)$. Offset the location of $s(k, t\text{-}1)$ by $(w\text{-}k\text{+}1)/w\times$mean, and set the dynamic programming range around the offset seam location based on the variance (if the variance is large, the range is large). **(iv)**

-Perform seam carving on $F(k\text{-}1, t)$ based on $E(k\text{-}1, t)$-$W(k\text{-}1, t)$ defined in (2) within the dynamic programming range mentioned above to produce $F(k, t)$, and denote the removed seam as $s(k, t)$.

end
end
end

perform the original seam carving algorithm to find a new seam for carving out without using temporal information.

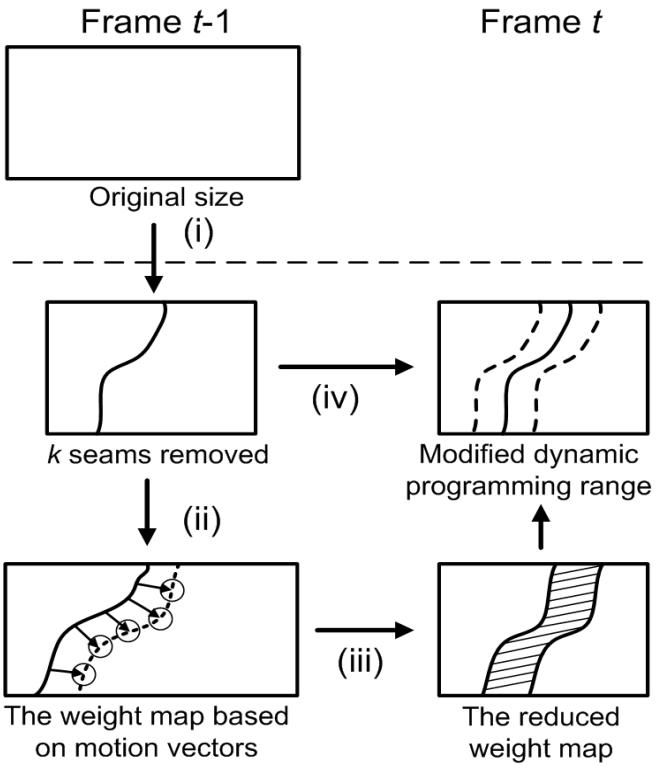


Fig. 4. The flowchart of the proposed video retargeting based on motion vectors. Details of Steps (i)-(iv) are described in Table 1.

4.2. Pixel-Based Optimization

The motion weight prediction procedure can individually used for video retargeting, while the motion vectors can only provide coarse motion estimation. In order to achieve finer temporal smoothness, the concept of optical flow is further considered in our work. The idea of optical flow is to perform accurate motion estimation by approaching the following equalities:

$$I(x, y, t) = I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t), \tag{3}$$

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt = 0, \tag{4}$$

where (4) is very similar to the gradients used in (1) with an additional temporal difference term (of digital videos). Based on this observation, the energy function in (2) could be further modified as:

$$\begin{aligned} e''(\mathrm{I}) &= w_f \cdot face(x, y, t) + w_s \cdot saliency(x, y, t) \\ &+ \left|\frac{\partial}{\partial x}\mathrm{I}(x, y, t)\right| + \left|\frac{\partial}{\partial y}\mathrm{I}(x, y, t)\right| + \left|\frac{\partial}{\partial t}\mathrm{I}(x, y, t)\right| \\ &+ w_c \cdot foward(x, y, t) - reduced\ weight\ map(x, y, t). \end{aligned} \tag{5}$$

Eq. (5) considers both the content energy in the current frame (frame t) and the motion energy of the current frame

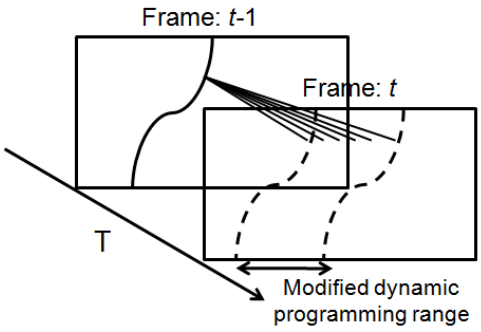
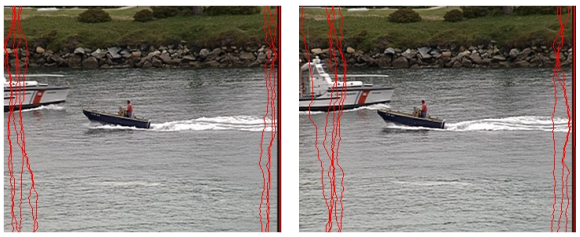


Fig. 5. The idea of exploiting the pixel-based optical flow for video retargeting, where the temporal difference term of each pixel in the current frame is the intensity difference between this pixel and the seam pixel at the same vertical coordinate (for width-reducing case).

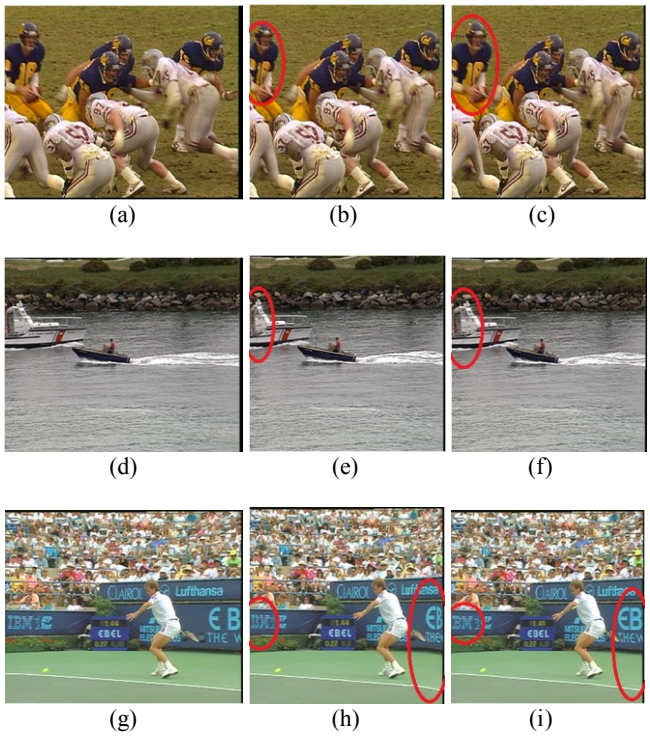


(a) (b)

Fig. 6. The use of motion vectors and pixel-based optimization could guide the detection of seams in consecutive frames to achieve temporal plausibility. (a) Frame *t*. (b) Frame *t*+10.

with respect to the detected seam in the previous frame (frame *t*-1). Still taking the width-reducing case as an example, the second row in (5) are just the spatial gradients in frame *t* plus a temporal difference term, which is modified as the intensity difference between each pixel in the search range of frame *t* with respect to the seam pixel at the same vertical coordinate in frame *t*-1. For height-reducing case, this temporal term denotes the difference of intensities between each pixel and the seam pixel at the same horizontal coordinate. To perform pixel-based optimization, the energy function (2) used in Table 1 is replaced by the energy function in (5). Fig.5 shows the concept of using the pixel-based optical flow for width-reducing case.

In addition, we can take a timing window which contains *N* frames (the current frame and *N*-1 frames before it) into account. The weighted sum is performed over all the temporal differences of the current frame with respect to the detected seams in the previous *N*-1 frames as the new temporal difference term to be considered, where each temporal difference map (between frame *k* and the current frame) receives weight according to the timing distance between frame *k* and the current frame. Based on this idea, the timing constraints between previous frames and the current frame can be easily merged into the dynamic programming process without exhaustively calculating the optical flow of the whole video, which significantly accelerates the retargeting process.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 7. The width-reducing retargeting results on the *football*, *coastguard*, and *tennis tournament* videos. (a)(d)(g) The original frames (size = 352×288). (b)(e)(h) The results based on motion weight prediction only (size = 288×288). (c)(f)(i) The results based on motion weight prediction and pixel-based optimization (size = 288×288). The red circles show the differences, where the usage of pixel-based optimization preserves better content quality.

5. EXPERIMENTAL RESULTS

The proposed video retargeting method is examined on several standard videos and some other videos with high motion. Fig. 6 shows that the motion vectors and pixel-based optimization could guide the detection of seams in consecutive frames to achieve temporal plausibility.

Several retargeting results with width reduction are shown in Fig. 7 to compare the performances with and without the use of pixel-based optimization. As marked with the red circles, the concept of pixel-based optical flow preserves more important content and achieves less distortion in each video frame. From these experimental results, the proposed video retargeting method does achieve not only the content-aware but also the timing smoothing video resizing. Limitations happen when the background of each frame is fairly simple while the video is of high motion, which makes the motion vectors and pixel-based optimization fail to estimate the correct motion directions. Fig. 8 shows an example with discontinuity artifacts.

(a) (b) (c)

Fig. 8. The retargeting results on the *baseball game* video. (a) The original frame (size = 360×240). (b) The result based on motion weight prediction only (size = 280×240). (c) The result based on motion weight prediction and pixel-based optimization (size = 280×240). The red circles show the artifacts.

6. COMPUTATIONAL COMPLEXITY

In Table 2, we list the **computation complexities** of the proposed method and the min-cut / max-flow algorithm used in [2] for graph cut, where V is corresponding to the product of the frame size and the number of fames in a video clip (V also stands for the number of nodes in the graph), K denotes the number of seams to be carved out, E is the number of edges in the graph, and $|C|$ stands for the cost of the minimum cut. From the view of computational complexity, our method achieves significant speed-up.

7. CONCLUSIONS

In this paper, a new video retargeting scheme is proposed by combining content and temporal information with the seam carving technique. The block-based motion estimation and Gaussian masks are integrated to achieve the motion weight prediction, which can guide the coarse seam locations in the current frame and reduce the search range of dynamic programming for seam carving. The pixel-based optimization is further exploited to find the optimal seam in the current frame with respect to the seams in the previous frames considering the temporal constraints. Through combining the motion weight prediction and the pixel-based optimization schemes, the temporal constraints can easily be introduced in the dynamic programming process so as to reduce the artifacts caused by temporal discontinuities.

Compared to the graph-cut based video retargeting method proposed in [2], the proposed method can (a) effectively use the motion vectors contained in compressed video streams, (b) release the restriction of seam movement to attain more flexible retargeting results, (c) achieve lower computational complexity through dynamic programming, and finally (d) reduce the dynamic programming search range. The experimental results show that our method can achieve not only the content-aware but also the timing smoothing video retargeting results. For our future works, we'll try to solve the artifacts on high-motion and simple background videos mentioned above, and aim at combining several video retargeting techniques for performance improvement.

Table 2. The comparison of **computational complexities**

Method	Algorithm	Complexity
Proposed method	Dynamic programming	$\boldsymbol{O(V\times K)}$
Graph cut	Dinic algorithm	$O(K\times E\times V^2)$
	Push-Relabel algorithm	$O(K\times V^3)$
	Algorithm proposed in [10]	$O(\|C\|\times K\times E\times V^2)$

8. REFERENCES

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

[2] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, article 16, Aug., 2008.

[3] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, article 23, Aug. 2009.

[4] F. Liu and M. Gleicher, "Video retargeting: automating pan and scan," *ACM Multimedia*, pp. 241-250, 2006.

[5] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," *ACM Human Factors in Computing Systems*, pp. 771-780, 2006.

[6] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video retargeting," *Proceedings of IEEE ICCV,* pp. 1-6, 2007.

[7] Y. Wang, C. Tai, O. Sorkine, and T. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, article 128, Dec. 2008.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[9] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[10] Y. Boykov and V. Kolmogorov, "An experimental Comparison of min-cut / max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124-1137, 2004.