Video Retargeting Using Non-homogeneous Scaling and Cropping

Wen-Yu Yo, Jin-Jang Leou, and Han-Hui Hsiao
Department of Computer Science and Information Engineering
National Chung Cheng University, Chiayi 621, Taiwan, R.O.C.
E-mail: {ywy99m, jjleou, hhh95p}@cs.ccu.edu.tw Tel: +886-5-2720411 ext. 33105

Abstract—**To display a video sequence in different display devices with different resolutions and aspect ratios, the video frames should be resized (retargeted) adaptively, whereas the important content of the video sequence should be retained. This is called the video retargeting problem.**

The proposed video retargeting approach consists of three main stages: preprocessing, optimization, and transformation. In the preprocessing stage, the initial scaling factor map, the saliency measure, and the temporal coherence measure of each video frame will be computed. In the optimization stage, an iterative optimization procedure is used to find the scaling factor maps of individual video frames via an optimization function involving spatial benefit and temporal cost. In the transformation stage, the scaling factor maps of individual video frames are improved by cropping. Then, the retargeted video frames are generated by pixel fusion using the scaling factor maps. Based on the experimented results obtained in this study, the performance of the proposed approach is better than those of five comparison approaches.

I. INTRODUCTION

In recent years, multimedia contents (videos) may be displayed in different devices with different resolutions and aspect ratios. To display a video sequence on different devices, the video frames should be resized (retargeted) adaptively, whereas the important content of the video sequence should be retained. This is called the video retargeting problem.

There are several simple video retargeting techniques, such as uniform scaling and direct cropping. Seam carving [1] is a content-aware image resizing technique, which will change the image size by iteratively removing or inserting a seam with the minimal cost. Content-aware video retargeting approaches, such as [2], should retain both spatial information and temporal coherence. Grundmann et al. [3] proposed a video retargeting approach, in which seams may be non-connected. To maintain temporal coherence, a cost function is used to select good seams in each video frame. Wolf et al. [4] proposed a non-homogeneous content-driven video retargeting approach. Wang et al. [5] proposed an image retargeting approach using warping. They use the mesh to get many quads, and compute the distortion energy for each quad by measuring the distance between the deformed quad and the uniformly scaled version of the original quad. The deformed mesh is obtained by global optimization iteratively. Wang et al. [6] proposed a video retargeting approach, extended from their image retargeting approach [5]. They align video frames by estimating motion firstly. Individual video frames are resized for preserving spatial content, whereas camera and object motions within video frames are aligned for retaining temporal coherence. Wang et al. [7] proposed a motion-based video retargeting approach with optimized crop-and warp. Wang et al. [8] proposed a video retargeting approach which handles spatial and temporal components sequentially. Kim et al. [9] determined the scaling factor of each column in the source image by using an importance map. In Yen et al. [10], after aligning video frames in a shot to a panoramic mosaic constructed for a shot, a global scaling map for these frames is derived for the panoramic mosaic so that effective temporal coherence can be retained. Hua et al. [11] proposed a distortion-free video retargeting approach by scale-space spatiotemporal saliency tracking. Li et al. [12] formulated video retargeting as the problem of finding an optimal trajectory for a cropping window to go through the video. Rubinstein et al. [13] defined a resizing space as a conceptual multi-dimensional space combining several resizing operators, and showed how a path in this space defines a sequence of operations to target media.

In this study, a video retarget approach using non-homogeneous scaling and cropping is proposed.

II. PROPOSED VIDEO RETARGETING APPROACH

As shown in Fig. 1, the proposed approach consists of three main stages: preprocessing, optimization, and transformation.

In the preprocessing stage, the initial scaling factor map, the saliency measure, and the temporal coherence measure of each video frame will be computed. In the optimization stage, an iterative optimization procedure is used to find the scaling maps of individual video frames via an optimization function involving spatial benefit and temporal cost. In the transformation stage, the scaling factor maps of individual video frames are improved by cropping. Then, the retargeted

+ This work was supported in part by National Science Council, Taiwan, Republic of China under Grants NSC 102-2221-E-194-028-MY2 and NSC 102-2221-E-194-041-MY3.

video frames are generated by pixel fusion [17] using the scaling factor maps.

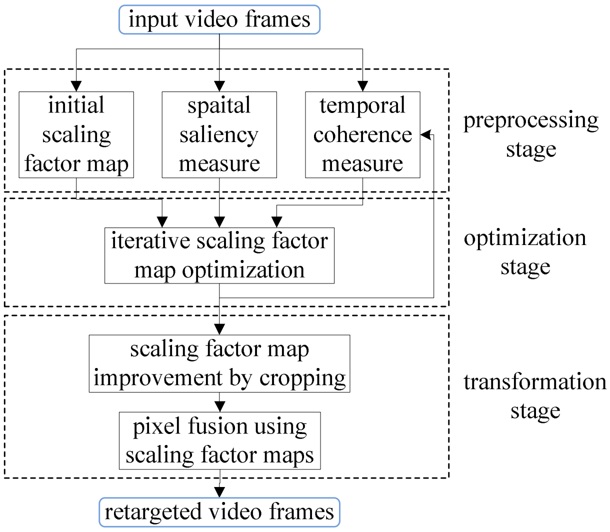

Fig. 1. The framework of the proposed approach.

Here, it is assumed that the input video frames are from a video shot with temporal coherence. F_t represents video frame t, T denotes the number of video frames, w and h represent the width and the height of each original video frame, respectively, and w' and h' represent the width and the height of each retargeted video frame, respectively. To simplify the presentation of the proposed approach, we only describe the case that the width of the original video frame is reduced (i.e., $h'=h$ and $w'<w$). The other cases can be similarly processed.

A. Preprocessing Stage

The scaling factor map of a video frame is used to transform the original video frame to the retargeted video frame by pixel-mapping. Each scaling factor is use to scale a pixel (or a patch) in the horizontal direction. Note that: (1) the size of the scaling factor map of a video frame is the same as that of the video frame, (2) the value of each scaling factor lies within the range $[0,1]$, (3) the sum of the scaling factors within each row should be equal to w', and (4) the difference between two neighboring scaling factors within each column should be very small. If $S_t(x,y)$ represents the scaling factor of pixel $F_t(x,y)$ in F_t, the above-mentioned constrains on scaling factor maps are as follows

$$\sum_x^w S_t(x,y) = w', \quad \forall y, \tag{1}$$

$$| S_t(x,y) - S_t(x,y+1) | \le TH_{SM}, \quad \forall x,y, \tag{2}$$

$$0 \le S_t(x,y) \le 1, \quad \forall x,y, \tag{3}$$

where TH_{SM} is a threshold (a small value) for spatial smoothness.

If $S_t^{(i)}(x,y)$ denotes the scaling factor of pixel $F_t(x,y)$ within the i-th iteration and $S_t^*(x,y)$ denotes the "optimal" scaling factor of pixel $F_t(x,y)$. Here the initial scaling factor $S_t^{(1)}(x,y)$ for $F_t(x,y)$ is

$$S_t^{(1)}(x,y) = \frac{w'}{w}, \quad \forall x,y. \tag{4}$$

The spatial saliency measure plays an important role in content-aware image/video retargeting, which reflects the importance of a pixel (or a region). The proposed energy function $E_t(x,y)$ for spatial saliency of pixel $F_t(x,y)$ combines the image saliency measure $IE_t(x,y)$ [14] and the video saliency measure $VE_t(x,y)$ [15] as follows

$$E_t(x,y) = 0.4 \cdot IE_t(x,y) + 0.6 \cdot VE_t(x,y), \forall x,y, \tag{5}$$

where both $IE_t(x,y)$ and $VE_t(x,y)$ are normalized to the range [0,1] using min-max normalization.

To determine temporal coherence measure using motion vectors, the exhaustive block matching algorithm (EBMA) is used to determine block motion vectors between frames F_t and F_{t-1}. Here, the block size is set to 2×2 and the search range of block matching is set to ±8. Based on the optimized scaling factor map S^*_{t-1} and motion vectors between F_{t-1} and F_t, the predicted scaling factor map PS_t can be described as

$$PS_t = MV_{t-1\to t}(S^*_{t-1}), \tag{6}$$

where the motion prediction function $MV_{t-1\to t}(\cdot)$ can be computed based on the motion vectors between video frames F_{t-1} and F_t.

B. Optimization stage

Based on the following rules: (1) if the saliency measure of a pixel is large, the scaling factor of the pixel should close to 1; (2) the temporal coherence should be retained; and (3) the three constraints (Eqs. (1)-(3)) should be retained, the "optimal" scaling factor can be determined as

$$\begin{aligned} S_t^*(x,y) = \operatorname*{argmax}_{S_t^{(n)}(x,y)}(&W_{SB} \cdot SB_t(E_t(x,y), S_t^{(n)}(x,y)) \\ &- W_{TC} \cdot TC_t(PS_t(x,y), S_t^{(n)}(x,y))), \ \forall x,y, \end{aligned} \tag{7}$$

where $SB_t(E_t(x,y), S_t^{(n)}(x,y))$ and $TC_t(PS_t(x,y), S_t^{(n)}(x,y))$ represent the spatial benefit and temporal cost of F_t, respectively, and W_{SB} and W_{TC} are the weighting coefficients of the spatial benefit and temporal cost, respectively. Here,

$$SB_t(E_t(x,y), S_t^{(n)}(x,y)) = E_t(x,y) \cdot S_t^{(n)}(x,y), \tag{8}$$

$$TC_t(PS_t(x,y), S_t^{(n)}(x,y)) = \left| PS_t(x,y) - S_t^{(n)}(x,y) \right|. \tag{9}$$

The constraints described in Eqs. (1)-(3) will be embedded in Eq. (7) as

$$s.t. \sum_x^w S_t(x,y) = w', \quad \forall y, \tag{10}$$

$$|S_t^{(n)}(x,y) - S_t^{(n)}(x,y+1)| \leq TH_{SM}, \quad \forall x, y, \tag{11}$$

$$TH_{lb} \leq S_t^{(n)}(x,y) \leq TH_{ub}, \forall x, y, \tag{12}$$

where TH_{lb} and TH_{ub} are the lower and upper bounds of the scaling factor, respectively.

To determine the "optimal" scaling factor described in Eq. (7) with three constraints, the interior-point solver [16] is employed. To reduce the computation complexity, the "optimal" scaling factor map may be determined in a patch-wise manner rather than in a pixel-wise manner. In this study, the number of iterations is set to 100. Additionally, for F_1 (the first video frame), $W_{SB} = 1$ and $W_{TC} = 0$, while for each other video frame F_t $(t \neq 1)$, $W_{SB} = 1$ and $W_{TC} = 4$.

C. *Transformation Stage*

As an illustrated example shown in Fig. 2, there may be some problem in some part of an "optimal" scaling factor map S_t^* marked by a red rectangular in Fig. 2(d). This artifact is induced by the small scaling factors of the part marked by the red rectangular in Fig. 2(b). In this study, the scaling factors of the part marked by the rectangular in Fig. 2(f) will be set to zero (i.e., cropping) in Fig. 2(g). That is, if the area of some part having small scaling factors is larger than a threshold, the scaling factors of the part are identically set to zero (i.e., cropping).

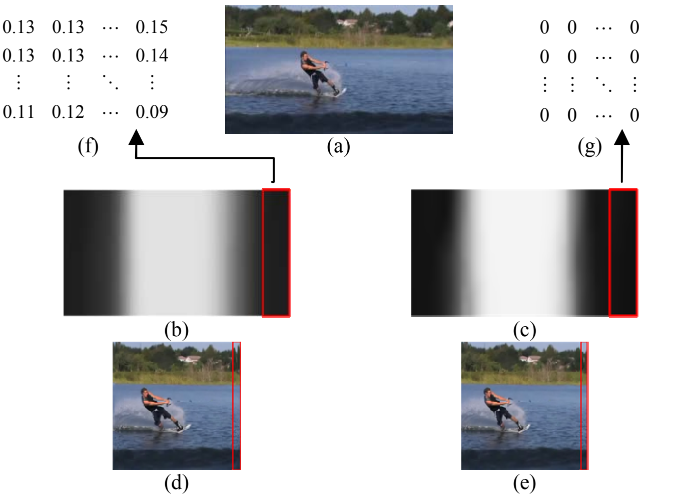

Fig. 2. (a) The original image, (b) the "optimal" scaling factor map, (c) the improved scaling factor map by cropping, (d) the retargeting result using the "optimal" scaling factor map, (e) the retargeting result using the improved scaling factor map, (f) the scaling factors of the part marked by a red rectangular in (b), and (g) the "improved" scaling factors of the part marked by a red rectangular in (c).

To maintain spatial coherence, cropping will be performed in a column-wise manner. If the average scaling factor of the pixels in a column is less than a threshold TH_{Crop} (0.2 in the proposed approach), this column will be cropped. To keep the constraint of Eq. (10) valid, the sum of the scaling factors of a cropped area will be uniformly distributed to the scaling factors of the non-cropped area.

Based on the final scaling factor maps S_t^F, t=1,...,T, the retargeting video frames RF_t, t=1,...,T, are generated by the pixel fusion method described in [17].

III. EXPERIMENTAL RESULTS

The proposed approach has been implemented on an Intel Core i7-2600 CPU 3.40 GHz PC using Matlab (version R2011b) software develop tool. To evaluate the effectiveness of the proposed approach, three video retargeting approaches are implemented, including uniform scaling, Grundmann et al.'s approach [3], and Yen et al.'s approach [10]. Additionally, Rubinstein et al.'s retargeting results [2] and Wang et al.'s retargeting results [6] will be compared. In this study, 11 test video sequences are employed. The characteristics of the 11 test video sequences are listed in Table 1.

Fig. 3 shows the original frames 1 and 60 (468×192) of "Car go forward," and the retargeted frames 1 and 60 (234×192) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach. Fig. 4 shows the original frames 1 and 60 (688×286) of "Cape No. 7 (1)," and the retargeted frames 1 and 60 (344×286) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach. Fig. 5 shows the original frames 1 and 150 (688×286) of "Cape No. 7 (2)," and the retargeted frames 1 and 150 (344×286) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach. Based on the experimental results obtained in this study, the retargeted results of the proposed approach are better than those of five comparison approaches.

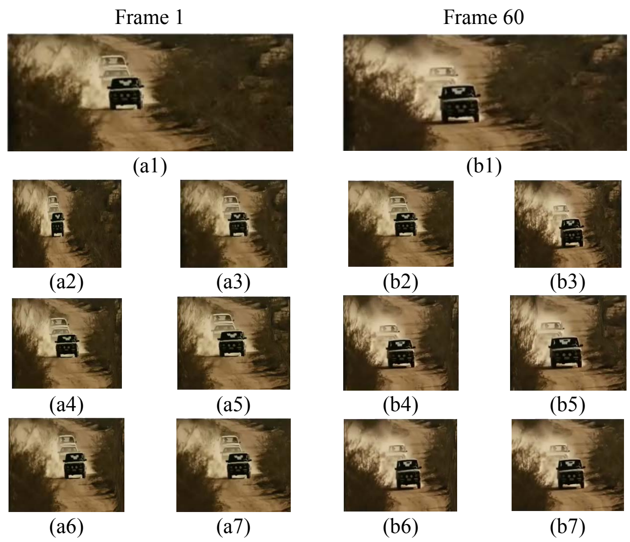

Fig. 3. The original frame 1 (a1) and frame 60 (b1) (468×192) of "Car go forward;" (a2)-(a7) the retargeted frames 1 and (b2)-(b7) the retargeted frames 60 (234×192) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach, respectively.

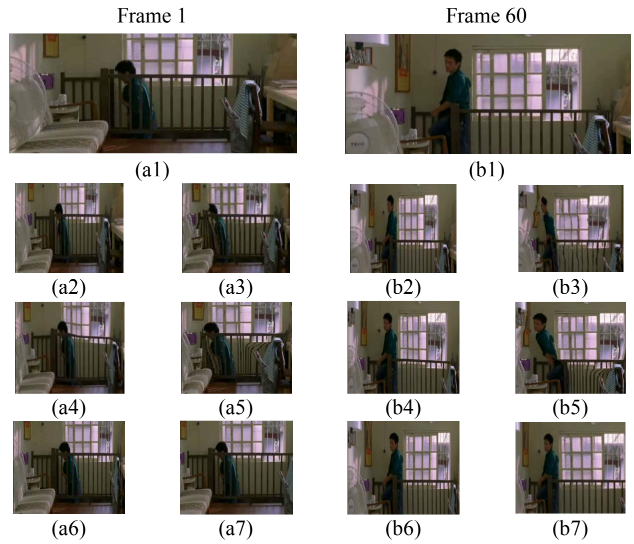

Fig. 4. The original frame 1 (a1) and frame 60 (b1) (688×286) of "Cape No. 7 (1);" (a2)-(a7) the retargeted frames 1 and (b2)-(b7) the retargeted frames 60 (344×286) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach, respectively.

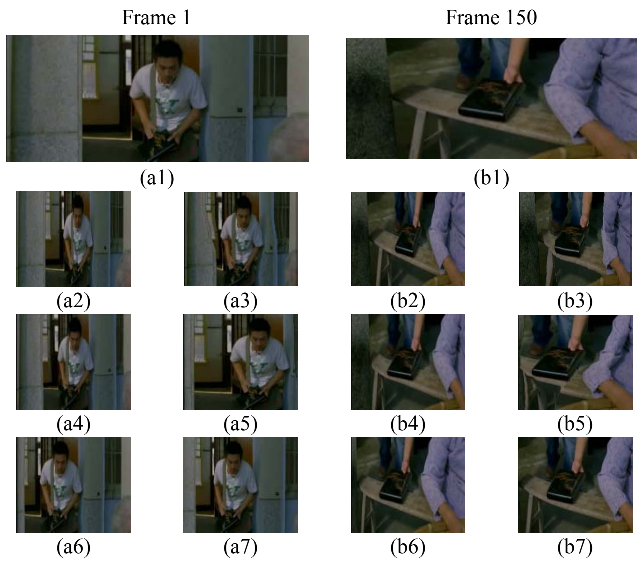

Fig. 5. The original frame 1 (a1) and frame 150 (b1) (688×286) of "Cape No. 7 (2);" (a2)-(a7) the retargeted frames 1 and (b2)-(b7) the retargeted frames 150 (344×286) of uniform scaling, Rubinstein et al. [2], Wang et al. [6], Grundmann et al. [3], Yen et al. [10], and the proposed approach, respectively.

IV. CONCLUDING REMARKS

The proposed video retargeting approach consists of three main stages: preprocessing, optimization, and transformation. In the preprocessing stage, the initial scaling factor map, the saliency measure, and the temporal coherence measure of each video frame will be computed. In the optimization sage, an iterative optimization procedure is used to find the scaling factor maps of individual video frames via an optimization function involving spatial benefit and temporal cost. In the transformation stage, the scaling factor maps of individual video frames are improved by cropping. Then, the retargeted video frames are generated by pixel fusion [17] using the scaling factor maps. Based on the experimented results obtained in this study, the performance of the proposed approach is better than those of five comparison approaches.

TABLE I

CHARACTERISTICS OF 11 TEST VIDEO SEQUENCES.

Sequences	Original size	Target size	Frame number	Motion	Objects	Background
Car forward	468×192	234×192	124	low	one	Natural
Coastguard	352×288	240×288	151	low	two	Natural
Two children	468×192	234×192	155	medium	two	Natural
Cape No. 7 (1)	688×286	344×286	120	low	two	artificial
Cape No. 7 (2)	688×286	344×286	365	low	two	artificial
Dark glass woman	640×320	320×320	179	low	one	artificial
Two girls	720×350	360×350	222	medium	two	artificial
Cheer leaders	384×288	232×288	165	low	many	artificial
Highway	448×192	224×192	194	medium	many	artificial
Water ski	540×280	270×280	96	high	one	Natural
Basketball	232×176	160×176	228	high	many	artificial

REFERENCES

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

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

[3] M. Grundmann, V. Kwatra, Han Mei, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. of 2010 IEEE Conf. on Computer Vision and Pattern Recognition* , June 2010, pp. 569-576.

[4] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content driven video retargeting," in *Proc. of 2007 IEEE Int. on Computer Vision*, Oct. 2007, pp. 1-6.

[5] Y. S. Wang, C. L. Tai, O. Sorkine, and T. Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. on Graphics*, vol. 27, no. 5, Dec. 2008.

[6] Y. S. Wang, H. Fu, O. Sorkine, T. Y. Lee, "Motion-aware temporal coherence for video resizing," *ACM Trans. on Graphics*, vol. 28, no. 5, Dec. 2009.

[7] Y. S. Wang, H. C. Lin, O. Sorkine, and T. Y. Lee, "Motion-based video retargeting with optimized crop-and-warp," *ACM Trans. on Graphics*, vol. 29, no. 54, July 2010.

[8] Y. S. Wang, J. H. Hsiao, O. Sorkine, and T. Y. Lee, "Scalable and coherent video resizing with per-frame optimization," *ACM Trans. on Graphics*, vol. 30, no. 4, July 2011.

[9] J. H. Kim, J. S. Kim, and C. S. Kim, "Image and video retargeting using adaptive scaling function," in *Proc. of 2009 European Conf. on Signal Processing*, Aug. 2009, pp. 819-823.

[10] T. C. Yen, C. M. Tsai, and C. W. Lin, "Maintaining temporal coherence in video retargeting using mosaic-guided scaling," *IEEE Trans. On Image Processing*, vol. 20, no. 8, pp. 2339-2351, Aug. 2011.

[11] G. Hua, C. Zhang, Z. Liu, Z. Zhang, and Y. Shan, "Efficient scale-space spatiotemporal saliency tracking for distortion-free video retargeting," in *Proc. of 2009 Asian Conf. on Computer Vision*, Sept. 2009, pp. 182-192.

[12] Y. Li, Y. Yian, J. Yang, L.-Y. Duan, and W. Gao, “Video retargeting with multi-scale trajectory optimization,” in *Proc. of 2010 ACM Int. Conf. on Multimedia Information Retrieval*, Mar. 2010, pp. 45-54.

[13] M. Rubinstein, A. Shamir, and S. Avidan, “Multi-operator media retargeting,” *ACM Trans. on Graphics*, vol. 28, no. 3, Aug. 2009.

[14] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, “Global contrast based salient region detection,” in *Proc. of 2011 IEEE Conf. on Computer Vision and Pattern Recognition*, June 2011, pp. 409-416.

[15] J. Harel, C. Koch, and P. Perona, "Graph-Based visual saliency," in *Proc. of 2006 Neural Information Processing Systems*.

[16] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 877-900, 1999.

[17] T. Ren, Y. Liu, and G. Wu, “Image retargeting based on global energy optimization,” in *Proc. of 2009 IEEE Int. Conf. on Multimedia and Expo*, June 2009, pp. 406-409.