Effective Video Retargeting With Jittery Assessment

Bo Yan, *Senior Member, IEEE*, Binhang Yuan, and Bo Yang

***Abstract*—This paper presents an effective video retargeting method with the assessment of jittery artifact. Our method firstly constructs a new energy function by including both spatial and temporal constrains. Then the retargeting processing can be performed effectively by minimizing our proposed energy function for each frame. We also propose a new objective measurement to assess temporal coherence after video retargeting, the accuracy of which is verified by psycho-visual tests. Experimental results show that our video retargeting method is able to provide comfortable resized videos in terms of subjective and objective measurements.**

***Index Terms*—Jittery artifact, spatial coherence, temporal coherence, video retargeting.**

I. Introduction

The escalating diversity of display devices has imposed new demands on digital image/video processing. The same image/video needs to be displayed with different resolutions and aspect ratios on variant devices. The common approaches, such as cropping or uniform scaling, may discard the important regions or distort the important objects in images/videos. Recently image/video retargeting has played a more and more significant role to support such requirements. Many video retargeting approaches have been proposed in order to convert the image/video to a new target resolution or aspect ratio while preserving the important content [1]–[3].

Among these approaches, seam carving is one kind of important approach (discrete approach) for content aware image/video retargeting [4]. It proposes to repeatedly search an 8-connected path of pixels, called seam, from top to bottom or from left to right; then insert/remove the optimal seam with the least total energy to change the resolution of image/video. Based on the idea of the original seam carving, many improved methods have been proposed in order to remove the possible artifacts [5]–[7].

Based on seam-carving, Matthias *et al.* proposed a new discontinuous seam-carving algorithm for video retargeting [8]. This proposed algorithm allows for frame-by-frame processing based on the temporal coherence formulation; thus results in temporally discontinuous seams, as opposed to geometrically smooth and continuous seams in [4]. Moreover, [8] also uses a novel automatically computed measurement of spatial-temporal saliency, because the existing per-frame saliency (feature-based or gradient-based) does not always produce desirable performance for video retargeting.

Warping is another kind of important approach (continuous approach) for image/video retargeting, which proposes to optimize the global energy and uniformly/non-uniformly scale pixels using the saliency map as well as some spatial and temporal constraints [9]–[12]. The limitation of such methods is that warping may reduce into blunt cropping or uniform scaling when salient objects are too concentrated or too dispersive in the frame [13], [14].

Manuscript received March 08, 2013; revised June 20, 2013; accepted August 07, 2013. Date of publication October 17, 2013; date of current version December 12, 2013. This work was supported by NSFC (Grant No.: 61073067 and 61370158). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shahram Shirani.

The authors are with the School of Computer Science, Fudan University, Shanghai, China (e-mail: byan@fudan.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2286112

Based on uniform warping, Panozzo *et al.* [12] propose a new robust image retargeting method via axis-aligned deformation. This method uses the 1D space of axis-aligned deformations for image retargeting; thus it is able to minimize the standard warping energies in 1D. As a result, it can retarget images with much less complexity than others. However, this method can only be used in image retargeting. Temporal constrain should be included in order to maintain the temporal coherence for video retargeting. We will introduce more details about this method in Section II.

When analyzing the recent video retargeting approaches, how to assess the performance in maintaining temporal coherence has become the prominent challenge for video retargeting. In this paper, we present an effective video retargeting method with jittery assessment. Firstly, we construct a new energy function by including both spatial and temporal constrains based on axis-aligned deformation [12]. Then each frame can be retargeted independently and effectively by minimizing our proposed new energy function instead of the old one. We also propose a new objective measurement to assess temporal coherence after video retargeting, the accuracy of which is verified by psycho-visual tests.

The rest of this paper is organized as follows. In Section II, we introduce the axis-aligned deformation based image retargeting [12]. Then, we propose our effective video retargeting algorithm in Section III. We introduce our new objective measurement to assess temporal coherence and evaluate its accuracy in Section IV. Following it, we evaluate the performance of the proposed video retargeting algorithm by resizing various test videos to show its effectiveness in Section V. We also conduct a psycho-visual test for user study in this section to support our advantages. Finally, we draw our conclusions in Section VI.

II. Related Work

In axis-aligned deformation based image retargeting [12], Panozzo *et al.* propose to divide the input image into uniform grids with N columns and M rows. Assume that W_o and H_o denote the width and height of the input image respectively. Thus, each initial grid is with the width W_o/N and height H_o/M. In order to resize the input image to the target image with size $W_r \times H_r$, this method assumes that the grids at the same row have the same height and the grids at the same column have the same width in the target image.

[12] uses the vectors S^{rows} and S^{cols} to denote the unknown heights of the rows and widths of the columns. S^{rows} and S^{cols} are defined as:

$$\begin{cases} S^{rows} = (S_1^{rows}, S_2^{row}, S_3^{rows}, \ldots, S_M^{rows}) \\ S^{cols} = \left(S_1^{cols}, S_2^{cols}, S_3^{cols}, \ldots, S_N^{cols}\right) \end{cases} \tag{1}$$

So the retargeting processing can be presented by the vector $S = (S^{rows}, S^{cols})^T \in \mathbb{R}^{M+N}$. The As-Rigid-As-Possible (ARAP) energy is defined with good performance as [12]:

$$E_{\mathrm{ARAP}} = \sum_{i=1}^{M}\sum_{j=1}^{N} \Omega_{i,j}^2 \left(\left(\frac{M}{H_o} S_i^{rows} - 1 \right)^2 + \left(\frac{N}{W_o} S_j^{cols} - 1 \right)^2 \right) \tag{2}$$

where $\Omega_{i,j}$ denotes the average saliency value of the grid at row i and column j.

Vector S can be obtained by minimizing (2) subject to:

$$\begin{cases} S_i^{rows} \geq L^h, \quad i = 1, \ldots . M, \\ S_j^{cols} \geq L^w, \quad j = 1, \ldots . N, \\ S_1^{rows} + S_2^{row} + S_3^{rows} + \ldots . + S_M^{rows} = H_r \\ S_1^{cols} + S_2^{cols} + S_3^{cols} + \ldots . + S_N^{cols} = W_r \end{cases} \tag{3}$$

where $L^h, L^w > 0$ are the minimum sizes allowed for rows and columns of the deformed grids [12].

(a) (b)

Fig. 1. (a) Original input image; (b) importance map.

With the help of vector S, each grid will be scaled to the target resolution. In this way, the target image can be obtained [12].

This method is quite effective and efficient for image retargeting. However, it cannot be directly used in video retargeting due to the lack of constrain of temporal coherence. In order to address this problem, we propose a new effective method for video retargeting in Section III.

III. Video Retargeting

In our method, firstly the importance map of each frame is calculated. The importance map is a 2-D matrix having the same dimensions as the original image. Its values range from 0 to 1, indicating the importance of each pixel. A large importance value implies the important pixel. In our method, importance map is represented by the saliency of every pixel. The saliency is computed by Itti's algorithm [15]. (We use a MATLAB implementation from http://www.saliency-toolbox.net.) Fig. 1 shows the original test image and its importance map. The bright regions are important regions. We can see that they can well capture our visual attention.

It should be noted that our method is not limited to a single specific saliency algorithm, and other algorithms could also be used as definitions of saliency. Other than saliency, any values that can enhance the accuracy of importance map could also be added, such as gradient and face detection.

After calculating the importance map of each frame, the first frame of the video will be processed by directly minimizing the ARAP energy in (2) to get good spatial coherence. For the following frames, temporal coherence should also be considered besides spatial coherence in order to avoid jittery artifact. In our method, motion vectors (MVs) are used to track the moving objects since the MVs can be obtained from the decompression process without additional complexity. Other complex tracking algorithms (such as optical flow) are neglected. For the kth frame ($k > 1$), the energy E_k for retargeting is composed of both spatial and temporal constrains, which are denoted by E_k^S and E_k^T respectively.

Similar to (2), E_k^S is defined as:

$$E_k^S = \sum_{i=1}^{M}\sum_{j=1}^{N} \Omega_{i,j}^2 \left(\left(\frac{M}{H_o} S_{i,k}^{rows} - 1 \right)^2 + \left(\frac{N}{W_o} S_{j,k}^{cols} - 1 \right)^2 \right) \quad (4)$$

where $S_{i,k}^{rows}$ denotes the height of the ith row in the kth target frame and $S_{j,k}^{cols}$ denotes the width of the jth column in the kth target frame.

The grid size for retargeting in our method is 4×4, which is the minimal block size in current video codec. If the block size of the decompressed video stream is larger than it, we can directly use its MV to set the MV value for each corresponding 4×4 grid. Let's use $G_k(i, j)$ to denote the grid at row i and column j in frame k. Thus, with the help of MV we may localize the matching region of $G_k(i, j)$ in frame $k-1$.

Fig. 2 shows an example. As shown in this figure, the matching region of $G_k(i, j)$ in frame $k-1$ is marked in blue, which may cross four grids in frame $k-1$, denoted by $G_{k-1}(x, y)$, $G_{k-1}(x+1, y)$, $G_{k-1}(x, y+1)$ and $G_{k-1}(x+1, y+1)$ respectively. Assume that the horizontal and vertical distances between the matching region and $G_{k-1}(x, y)$ are denoted by x_0 and y_0 respectively. Then the temporal constrained width and height of $G_k(i, j)$, which are denoted by $S_{i',k}^{rows}$ and $S_{j',k}^{cols}$ respectively, are obtained by combining those of the corresponding grids as:

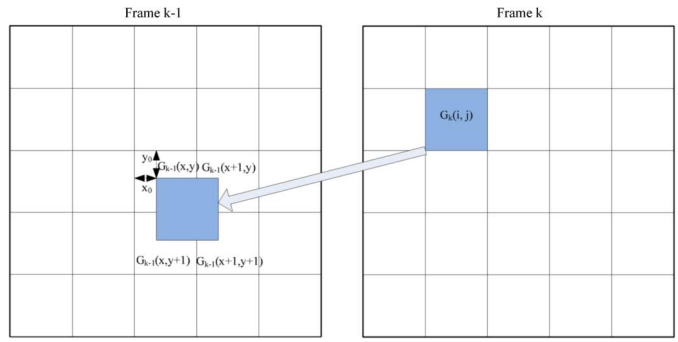

Fig. 2. Illustration for grid matching.

$$\begin{cases} S_{i',k}^{rows} = \left(1 - \frac{y_0 \cdot M}{H_o}\right) \cdot S_{y,k-1}^{rows} + \frac{y_0 \cdot M}{H_o} \cdot S_{y+1,k-1}^{rows} \\ S_{j',k}^{cols} = \left(1 - \frac{x_0 \cdot N}{W_o}\right) \cdot S_{x,k-1}^{cols} + \frac{x_0 \cdot N}{W_o} \cdot S_{x+1,k-1}^{cols} \end{cases} \quad (5)$$

As a result, temporal constrain of E_k is derived by:

$$E_k^T = \sum_{i=1}^{M}\sum_{j=1}^{N} \Omega_{i,j}^2 \left(\left(\frac{S_{i,k}^{rows}}{S_{i',k}^{rows}} - 1 \right)^2 + \left(\frac{S_{j,k}^{cols}}{S_{j',k}^{cols}} - 1 \right)^2 \right) \quad (6)$$

In this way, retargeting processing for the kth frame is performed by minimizing the energy E_k with the approach proposed in [12] as:

$$E_k = E_k^S + E_k^T \quad (7)$$

If there are no MVs for some grids, MV copy from the previous frame can be used [16], [17]. It should be noted that in our method vector median filter (VMF) [18] with 3×3 is applied to MVs in order to correct the erroneous MVs (such as the one caused by MV copy). It should be also noted that we only allow I and P frames in encoding the video sequences for simplicity. If B-frame is allowed, our work can be easily extended by updating (5).

IV. Jittery Assessment

In video retargeting, spatial coherence and temporal coherence are two factors to assess the retargeting performance. Spatial incoherency can be clearly demonstrated by comparing image quality. In contrast, temporal incoherency is difficult to be visually demonstrated. Unfortunately, there are no objective metrics to assess temporal coherence for video retargeting. In order to address this problem, in this section we will propose an objective measurement—jittery metric (JM) to assess the performance in maintaining temporal coherence. The results of psycho-visual tests can prove that the proposed metric is quite effective and credible.

A. Our Jittery Metric

In video retargeting, temporal incoherency is generally demonstrated with Jittery artifact, which is caused by independent scaling or seam-carving for each frame. Jittery artifact is quite harmful to video quality, which may cause users to be dizzy.

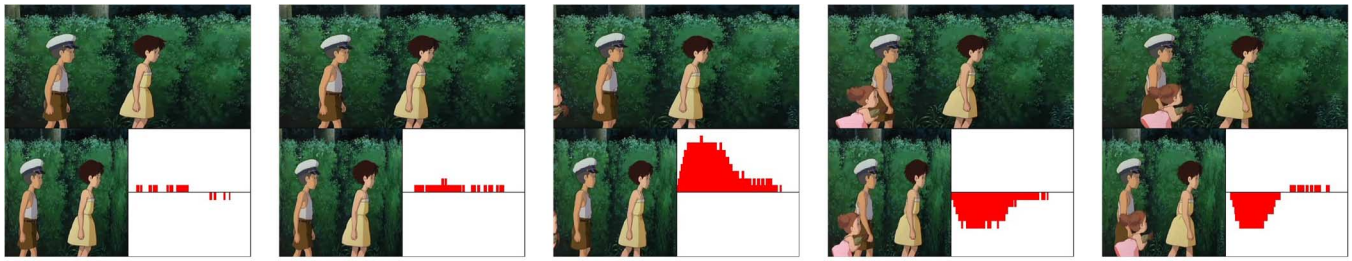

Fig. 3. The demo of jittery assessment. (a)–(e) Indicate several consecutive frames in chronological order. In the indicator, the x-axis indicates the column no. of the retargeted frame and y-axis denotes the value of $DS_{j,k}^{cols}$ in (9).

In order to detect jittery artifact caused by video retargeting, we firstly assume that the original input video is absolutely jittery-free. Then, we define the vectors S_k^{rows} and S_k^{cols} for the kth frame as:

$$\begin{cases} S_k^{rows} = \left(S_{1,k}^{rows}, S_{2,k}^{row}, S_{3,k}^{rows}, \ldots., S_{M,k}^{rows}\right) \\ S_k^{cols} = \left(S_{1,k}^{cols}, S_{2,k}^{cols}, S_{3,k}^{cols}, \ldots., S_{N,k}^{cols}\right) \end{cases} \tag{8}$$

where $S_{i,k}^{rows}$ and $S_{j,k}^{cols}$ are defined in (4).

We can also define the differences of heights or widths of the grids between two neighboring frames, denoted by $DS_{i,k}^{rows}$ and $DS_{j,k}^{cols}$ respectively, as:

$$\begin{cases} DS_{i,k}^{rows} = S_{i,k}^{rows} - S_{i,k-1}^{rows} \\ DS_{j,k}^{cols} = S_{j,k}^{cols} - S_{j,k-1}^{cols} \end{cases}, \qquad k > 1 \tag{9}$$

An intuitive assumption is that if $|DS_{j,k}^{cols}|$ is small, the horizontal jittery caused by retargeting at column j between the $(k-1)$th and kth frames should be slight. $DS_{j,k}^{cols} > 0$ presents that frame jitters from left to right at column j. Otherwise, it jitters from right to left. Similarly, the vertical jittery at row i can be presented by $DS_{i,k}^{rows}$.

Thus, we use the Euclidean distances of S_k^{rows} or S_k^{cols} in (8) to evaluate the jittery between two consecutive frames. The jittery metric (JM_k) between the kth and $(k-1)$th frames when retargeting in horizontal direction is defined as:

$$JM_k = \frac{\sqrt{\sum_{j=1}^{N}\left(S_{j,k}^{cols} - S_{j,k-1}^{cols}\right)^2}}{N} \tag{10}$$

If retargeting the video in vertical direction, JM_k is defined as:

$$JM_k = \frac{\sqrt{\sum_{i=1}^{M}\left(S_{i,k}^{rows} - S_{i,k-1}^{rows}\right)^2}}{M} \tag{11}$$

B. Psycho-Visual Tests

We implemented a demo for users to evaluate the accuracy of our proposed jittery metric by performing a psycho-visual testing. Fig. 3 shows an example of our demo. In this figure, the upper half part of each sub-figure shows the original frame. In the lower half part, the left one shows the retargeted frame and the right one is the indicator of our jittery assessment for each corresponding frame. In the indicator, the x-axis indicates the column No. of the retargeted frame and y-axis denotes the value of $DS_{j,k}^{cols}$ in (9). Fig. 3(a) to Fig. 3(e) indicate several consecutive frames in chronological order.

As we can observe from Fig. 3, in the first two frames jittery artifact is slight. Starting from the 3rd frame, a little girl enters the frame from the left. In order to keep the aspect ratio (spatial coherence) of the new object (the little girl), severe Jittery artifact will appear. In according with this phenomenon, the bars in Fig. 3(c) to Fig. 3(e) are much higher than those in the first two frames (Fig. 3(a) and Fig. 3(b)).

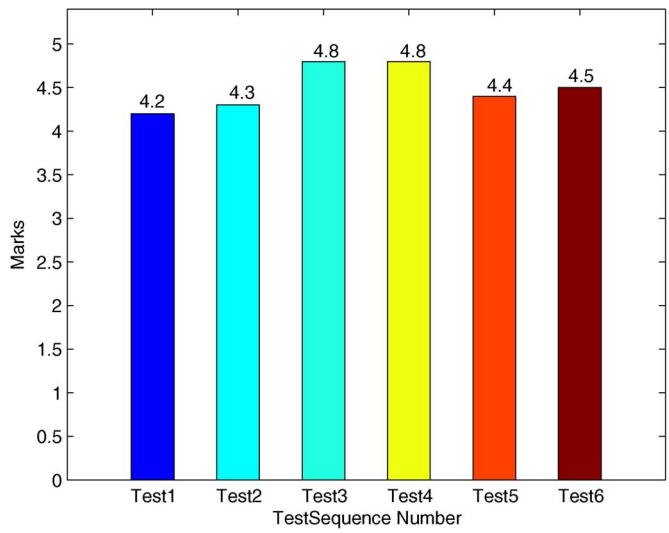

Fig. 4. MOS results of our psycho-visual tests.

With this demo, we evaluate the accuracy of our JM by conducting a user study with ten participants coming from diverse backgrounds and ages. Six sequences are selected and retargeted with Panozzo *et al.*'s ARAP method [12] to the half of the original width for testing. The chosen six videos range from low-motion scenes to high-motion scenes, from scenes containing one object movement to those involving multiple moving objects, and from movies to cartoons. Note that these sequences are different from those, which are used in performance evaluation in Section V. The metric (JM value) for each retargeted sequence can be calculated with (10).

Then we use five levels mean opinion score (MOS) to evaluate our metric in (10). The scores are between 1 and 5, in which 5 means that this objective metric value is perfectly consistent with the participant's subjective perception, while 1 means the metric value has totally no accordance with the user's visual sense. In the testing, video demo of each sequence was shown to each participant. Therefore, each participant was asked to provide six scores. The average score of each sequence is shown in Fig. 4. As shown in this figure, the subjective score is quite satisfying. Therefore, our proposed metric is credible to evaluate the jittery artifact. We will use this metric to assess temporal coherence performance in Section V.

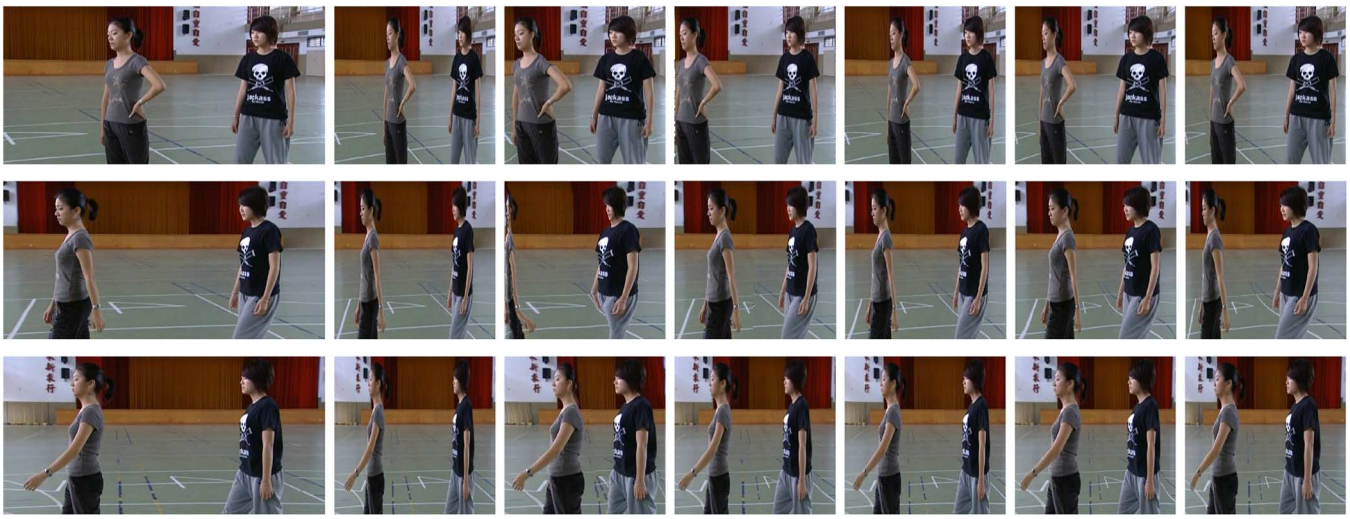

Fig. 5. Comparison with the existing methods for "girls" sequence. From left to right: original image, uniform scaling, DSC [8], WARP [9], IF [19], ARAP [12] and our method. All videos are resized to half width ($W'/W = 0.5$).

Fig. 6. Comparison with the existing methods for "walking-man" sequence. From left to right: original image, uniform scaling, DSC [8], WARP [9], IF [19], ARAP [12] and our method. All videos are resized to half width ($W'/W = 0.5$).

V. Experiment Results and Discussions

In order to evaluate the performance of our proposed method, we will test the performance on videos with a wide range of resolutions and aspect ratios. The processing environment has an Intel Xeon CPU with 3.1 GHz operational frequency and 2G bytes RAM size under Windows Server 2003 operating system. All the algorithms are implemented by MATLAB. In our testing, all videos are resized to half of the original width. The results of several other state-of-the-art techniques for video retargeting are shown as comparison to our method, including uniform scaling (US), discontinuous video seam-carving (DSC) [8], nonhomogeneous warping (WARP) [9], Ding *et al.*'s importance filtering (IF) method [19] and Panozzo *et al.*'s ARAP method [12].

A. Video Comparison

Fig. 5 and Fig. 6 show some of the retargeting result comparisons for different videos. For each video, we select some frames in chronological order, and each row in the group of figures represents the retargeting result for the same frame using different methods. The first column of each group of figures contains the frames from original video. The last column is the retargeting result using our proposed algorithm. Other columns are the retargeting videos generated by other methods. As shown in these figures, our method successfully produces resized videos with good spatial coherence and faithfully preserves the visually salient objects in most cases.

Note that our proposed method is an extended work by including temporal coherence based on Panozzo *et al.*'s ARAP method [12]. Therefore, the performance in spatial coherence provided by ARAP [12] should be the upper bound of our proposed method. As we can observe from these figures, spatial coherence of our proposed method is very close to that of ARAP [12].

With the help of our proposed JM metric, we have compared the performances in maintaining temporal coherence between ARAP [12] and our method. Fig. 7 shows the comparison results with JM value in (10) versus frame number for these different test sequences with different methods. From these figures, we can observe that JM values of our proposed method are much smaller and more stable than those

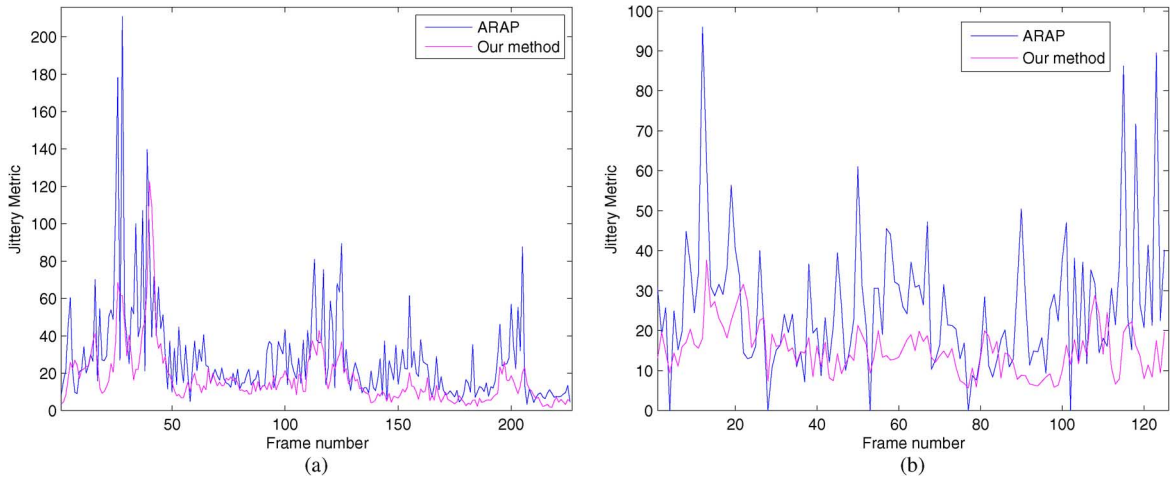

Fig. 7. Comparison of JM values in (10) for different sequences between ARAP [12] (the blue curve) and our proposed method (the magenta curve). (a) "Girls" sequence; (b) "walking-man" sequence.

TABLE I
PAIRWISE COMPARISON RESULTS BETWEEN OUR METHOD, WARP [9], IF [19] AND DSC [8]

↓ outperforms →	Ours	DSC	IF	WARP	Total	Prefer
Ours	-	55	78	71	204	68.0%
DSC [8]	45	-	59	61	165	55.0%
IF [19]	22	41	-	55	118	39.3%
WARP [9]	29	39	45	-	113	37.7%

of ARAP [12], which means that our proposed method has reduced lots of Jittery artifact compared with [12].

B. User Study

Because there is no widely-accepted objective metric to compare different retargeting methods so far, we perform a psycho-visual testing on these methods in this section. However, a thorough psychophysical testing of video retargeting is out of the scope of our work, so we decided to focus on a concise setting, comparing to the three most representative techniques, WARP of [9], DSC of [8] and IF of [19].

We evaluated our method by conducting a user study with 20 participants coming from diverse backgrounds and ages. Our psycho-visual tests followed the setting of [3], taking the paired comparisons approach: participants were presented with an original video sequence and two retargeting videos side by side, and they were asked to answer which retargeting result was preferred by them. The users were kept naive about the purpose of the experiment and were not provided with any special technical instructions. We used 5 different videos in the experiment and resized each video to 50% using WARP [9], DSC [8], IF [19] and our method. Therefore, for each video there were 6 pairwise comparisons, and each participant was asked to make $6 \times 5 = 30$ comparisons. The questions were presented in random order to avoid bias. We obtained a total of $30 \times 20 = 600$ answers, and each method was compared $3 \times 5 \times 20 = 300$ times.

Table I shows the summary of the obtained results, supporting significant preference of our method. Entry a_{ij} in the middle of the table means method i was preferred a_{ij} times over method j. Overall, our method was preferred in 68.00% (204/300) of the times it was compared. In contrast, WARP was favored only in 37.66% (113/300), DSC in 55.00% (165/300) and IF in 39.33% (118/300) of the comparisons. It should be noted that the results of our visual-psycho tests are indicative due to the conciseness of the chosen setting.

C. Limitations

Our proposed method proposes to use same scales for the pixels at same columns/rows in horizontal/vertical scaling. It excludes the localized rotations of images and has less freedom than general variational warping. As a result, some special structures such as the straight lines in an image cannot be preserved if they are not strictly horizontally or vertically aligned. However, our proposed method is able to preserve most structures as demonstrated in Fig. 5 and Fig. 6.

It should be noted that as stated in [12] our axis-aligned deformation based method is able to consume much less computational complexity than other state-of-art techniques. Therefore, our method possesses huge potential of practical applications. Our current proposed jittery metric is mainly designed for axis-aligned deformation based video retargeting algorithms. How to extend it to work with other algorithms is the topic of our future work.

VI. CONCLUSIONS

In this paper, we presented an effective video retargeting method with jittery assessment. The novelty of this paper lies in two parts. Firstly, it constructs a new energy function by including both spatial and temporal constrains based on axis-aligned deformation [12], and then processes each frame effectively by minimizing this proposed energy function. Secondly, it proposes a new objective measurement to assess temporal coherence after video retargeting, the accuracy of which is verified by psycho-visual tests. Our method outperforms other state-of-the-art retargeting systems, as demonstrated in the results and widely supported by the conducted user study.

REFERENCES

[1] Z. Yuan, T. Lu, Y. Huang, D. Wu, and H. Yu, "Addressing visual consistency in video retargeting: A refined homogeneous approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 890–903, Jun. 2012.
[2] B. Yan, K. Sun, and L. Liu, "Matching-area-based seam carving for video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 302–310, 2013.
[3] Y. Wang, H. Lin, O. Sorkine, and T. Lee, "Motion-based video retargeting with optimized crop-and-warp," *ACM Trans. Graph. (TOG)*, vol. 29, no. 4, p. 90, 2010.
[4] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph. (TOG)*, vol. 26, no. 3, p. 10, 2007.
[5] A. Shamir and O. Sorkine, "Visual media retargeting," in *ACM SIGGRAPH ASIA 2009 Courses*, 2009, p. 11.

[6] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph. (TOG)*, vol. 27, no. 3, pp. 1–9, 2008.
[7] C.-K. Chiang, S.-F. Wang, Y.-L. Chen, and S.-H. Lai, "Fast JND-based video carving with GPU acceleration for real-time video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 11, pp. 1588–1597, Nov. 2009.
[8] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 569–576.
[9] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE 11th Int. Conf. Computer Vision, 2007 (ICCV 2007)*, Oct. 2007, pp. 1–6.
[10] T.-C. Yen, C.-M. Tsai, and C.-W. Lin, "Maintaining temporal coherence in video retargeting using mosaic-guided scaling," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2339–2351, Aug. 2011.
[11] K. Sun, B. Yan, and Y. Gao, "Low complexity content-aware image retargeting," in *Proc. 2012 19th IEEE Int. Conf. Image Processing (ICIP)*, 2012, pp. 2105–2108.
[12] D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," *Comput. Graph. Forum*, vol. 31, no. 2, pp. 229–236, 2012.
[13] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph. (TOG)*, vol. 28, no. 5, p. 126, 2009.
[14] L. Shi, J. Wang, L. Duan, and H. Lu, "Consumer video retargeting: Context assisted spatial-temporal grid optimization," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 301–310.
[15] 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.
[16] B. Yan and H. Gharavi, "A hybrid frame concealment algorithm for H.264/AVC," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 98–107, Jan. 2010.
[17] B. Yan and J. Zhou, "Efficient frame concealment for depth image-based 3-D video transmission," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 936–941, 2012.
[18] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, 1990.
[19] Y. Ding, J. Xiao, and J. Yu, "Importance filtering for image retargeting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2011*, Jun. 2011, pp. 89–96.

Random Network Coding for Multimedia Delivery Services in LTE/LTE-Advanced

Dejan Vukobratović, Chadi Khirallah, Vladimir Stanković, and John S. Thompson

***Abstract*—Random Network Coding (RNC) has recently been investigated as a promising solution for reliable multimedia delivery over wireless networks. RNC possess the potential for flexible and adaptive matching of packet-level error resilience to both video content importance and variable wireless channel conditions. As the demand for massive multimedia delivery over fourth generation wireless cellular standards such as Long-Term Evolution (LTE)/LTE-Advanced (LTE-A) increases, novel video-aware transmission techniques are needed. In this paper, we investigate RNC as one such promising technique, building upon our recent work on RNC integration within the LTE/LTE-A Radio Access Network at the Multiple Access Control (MAC) layer (MAC-RNC). The paper argues that the proposed MAC-RNC solution provides fundamentally new set of opportunities for dynamic collaborative transmission, content awareness, resource allocation and unequal error protection (UEP) necessary for efficient wireless multimedia delivery in LTE/LTE-A.**

***Index Terms*—LTE/LTE-A, random network coding, wireless video delivery.**

I. Introduction

Recent technological advances both at the user and the network side are bringing advanced multimedia delivery services over mobile cellular networks closer to reality. Indeed, today's smartphones are powerful handheld computers with large screen sizes/resolutions while user data rates offered by fourth generation cellular interfaces such as 3GPP Long-Term Evolution (LTE)/LTE-Advanced (LTE-A) satisfy high-quality multimedia streaming demands. In recent years, several mobile multimedia delivery services have been standardized, evolved and commercially used; the most notable examples include the 3GPP Packet-Switched Service (PSS) [1], the 3GPP Multimedia Broadcast/Multicast Service (MBMS) [2] and adaptive HTTP streaming (AHS) [3]. However, apart from these upper layer efforts, LTE/LTE-A radio access network (RAN) protocols remain largely oblivious to multimedia traffic, with rare exceptions in the domain of priority-based

Manuscript received December 12, 2012; revised March 28, 2013 and June 23, 2013; accepted June 27, 2013. Date of publication September 16, 2013; date of current version December 12, 2013. This work was supported in part by the EPSRC Sensor Signal Processing Platform Grant EP/J015180/1. The work of D. Vukobratović was supported by a Marie Curie European Reintegration Grant FP7-PEOPLE-ERG-2010 MMCODESTREAM within the 7th EU Framework Programme. This paper is an extended version of the original paper which appeared in the Proceedings of IEEE ICME 2012 Conference and was among the top-rated 4% of ICME'12 submissions. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Enrico Magli.

D. Vukobratović is with the Department of Power, Electronics and Communication Engineering, University of Novi Sad, Novi Sad, Serbia (e-mail: dejanv@uns.ac.rs).

C. Khirallah and J. Thompson are with the School of Engineering and Electronics, University of Edinburgh, Edinburgh, U.K. (e-mail: c.khirallah@ed.ac.uk; john.thompson@ed.ac.uk).

V. Stanković is with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K. (e-mail: vladimir.stankovic@eee.strath.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2282129