A new metric to assess temporal coherence for video retargeting

Ke Li, Bo Yan, Binhang Yuan

Ke Li, Bo Yan, Binhang Yuan, "A new metric to assess temporal coherence for video retargeting," Proc. SPIE 9273, Optoelectronic Imaging and Multimedia Technology III, 92732Z (31 October 2014); doi: 10.1117/12.2071637

Event: SPIE/COS Photonics Asia, 2014, Beijing, China

A New Metric to Assess Temporal Coherence for Video Retargeting

Ke Li, Bo Yan, and Binhang Yuan

School of Computer Science,
Shanghai Key Laboratory of Intelligent Information Processing,
Fudan University, Shanghai, China;

ABSTRACT

In video retargeting, how to assess the performance in maintaining temporal coherence has become the prominent challenge. In this paper, we will present a new objective measurement to assess temporal coherence after video retargeting. It's a general metric to assess jittery artifact for both discrete and continuous video retargeting methods, the accuracy of which is verified by psycho-visual tests. As a result, our proposed assessment method possesses huge practical significance.

Keywords: video retargeting, jittery artifact, temporal coherence.

1. INTRODUCTION

In the last few years, due to the escalating diversity of display devices, image/video retargeting has played a more and more significant role to support image/video display with various resolutions, formats, and aspect ratios. Traditional industry solutions don't perform well to meet this requirement. For example: uniform scaling will distort the salient object; cropping may cut the important part; letterboxing will waste the space on the screen; and partially warping will not properly handle the situation when the salient objects move to the margin of the screen.[1] In order to convert the video to a new target resolution or aspect ratio while preserving the salient content, video retargeting[2] has played a more and more significant role recently. Recent video retargeting approaches can be classified into two main branches in terms of the essential ideas: removing/inserting pixels from a video (discrete methods)[3–5] and merging pixels of a video (continuous methods).[6–8]

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 will present a new objective measurement to assess temporal coherence after video retargeting. It's a general metric to assess jittery artifact for both discrete and continuous video retargeting methods, the accuracy of which is verified by psycho-visual tests.

The rest of this paper is organized as follows. In Section 2, we introduce our general jittery metric to assess temporal coherence for video retargeting. Then, the accuracy of this proposed metric is evaluated by resizing various test videos with different state-of-art video retargeting methods in Section 3. We also conduct a psycho-visual test for user study in this section to support our advantages. Finally, we draw our conclusions in Section 4.

2. OUR PROPOSED METRIC

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 metric 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.

Further author information: (Send correspondence to Bo Yan)
Bo Yan: E-mail: byan@fudan.edu.cn

Optoelectronic Imaging and Multimedia Technology III, edited by Qionghai Dai, Tsutomu Shimura,
Proc. of SPIE Vol. 9273, 92732Z · © 2014 SPIE · CCC code: 0277-786X/14/$18
doi: 10.1117/12.2071637

Temporal incoherency is generally demonstrated with jittery artifact in video retargeting, 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.

Assume that we will resize the k^{th} original frame I_k^o with size $W_o \times H_o$ to a new retargeted frame I_k^r with size $W_r \times H_r$. For each $I_k^o(i,j)$, we may obtain its mapping pixel in the retargeted frame, the coordinate of which is denoted by $(X_k(i,j), Y_k(i,j))$.

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 X_k and Y_k for the k^{th} frame as:

$$X_k = \begin{pmatrix} X_k(0,0), & \dots & X_k(W_o-1,0) \\ \vdots & \ddots & \vdots \\ X_k(0,H_o-1) & \cdots & X_k(W_o-1,H_o-1) \end{pmatrix} \tag{1}$$

$$Y_k = \begin{pmatrix} Y_k(0,0), & \dots & Y_k(W_o-1,0) \\ \vdots & \ddots & \vdots \\ Y_k(0,H_o-1) & \cdots & Y_k(W_o-1,H_o-1) \end{pmatrix} \tag{2}$$

We can also define the coordinate differences of the corresponding pixels between two neighboring frames, denoted by $DX_k(i,j)$ and $DY_k(i,j)$ respectively, as:

$$\begin{cases} DX_k(i,j) = X_k(i,j) - X_{k-1}(i,j) \\ DY_k(i,j) = Y_k(i,j) - Y_{k-1}(i,j) \end{cases}, \qquad k > 0 \tag{3}$$

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

Thus, we use the Euclidean distances of X_k or Y_k in (1) and (2) to evaluate the jittery between two consecutive frames. The metric (ME_k) between the k^{th} and $(k-1)^{th}$ frames when retargeting in horizontal direction is defined as:

$$ME_k = \frac{\sqrt{\sum_{i=0}^{W_o-1}\sum_{j=0}^{H_o-1}(X_k(i,j) - X_{k-1}(i,j))^2}}{W_o \times H_o} \tag{4}$$

If retargeting the video in vertical direction, the ME_k is obtained as:

$$ME_k = \frac{\sqrt{\sum_{i=0}^{W_o-1}\sum_{j=0}^{H_o-1}(Y_k(i,j) - Y_{k-1}(i,j))^2}}{W_o \times H_o} \tag{5}$$

3. PERFORMANCE EVALUATION AND DISCUSSIONS

In order to evaluate the accuracy of our proposed metric, we tested it 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.

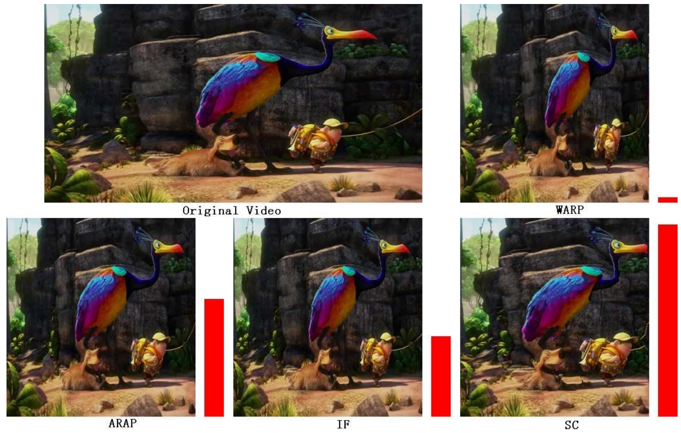

Figure 1. The demo of jittery assessment.

Table 1. Average metric comparison results between WARP,[6] ARAP,[10] IF[9] and SC.[3]

Sequences	Woman	Girls	Bird1	Walking-man	Bird2
WARP	9.81×10^{-5}	3.32×10^{-5}	1.70×10^{-5}	4.58×10^{-5}	1.40×10^{-5}
ARAP	3.63×10^{-3}	4.15×10^{-3}	9.15×10^{-3}	4.55×10^{-3}	2.89×10^{-3}
IF	2.25×10^{-3}	2.41×10^{-3}	3.44×10^{-3}	3.00×10^{-3}	1.65×10^{-3}
SC	1.80×10^{-2}	1.33×10^{-2}	1.67×10^{-2}	1.38×10^{-2}	1.21×10^{-2}

3.1 Testing platform design

We implemented a demo for users to evaluate the accuracy of our proposed jittery metric by performing a psycho-visual testing. In our demo, we selected several state-of-the-art techniques for video retargeting in testing, including seam carving (SC)[3] and importance filtering (IF) method,[9] nonhomogeneous warping (WARP),[6] and as-rigid-as-possible (ARAP) method.[10] The former two methods are two representative discrete video retargeting methods and others belong to continuous methods.

In our demo, in order to clearly show the differences of jittery artifact for different methods, we normalized the metric as:

$$H_k^{(i)} = \left(\frac{ME_k^{(i)}}{Max\{ME_m^{(j)}|1 \leq j \leq 4, 0 \leq m < N_f} \right)^{\alpha} \tag{6}$$

where $ME_k^{(i)}$ denotes the jittery metric in (4) for the k^{th} frame with the i^{th} method. N_f denotes the total frame number for the corresponding sequence. $\alpha = 0.5$ is selected in our demo for better visualization.

Fig. 1 shows a frame of our demo. As shown in this figure, the original video and four retargeted videos with different methods are shown simultaneously. A red bar is placed on the right of each retargeted video, the height of which is consistent with the value of our normalized metric ($H_k^{(i)}$ in (6)) for each corresponding frame.

3.2 Psycho-visual Tests

With this demo, we evaluate the accuracy of our metric by conducting a user study with 20 participants coming from diverse backgrounds and ages. Five sequences are selected in our testing. Fig. 2 to Fig. 3 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 and its jittery assessment for the same frame using different methods. The first column of each group of figures contains the frames from original video. Other columns are the retargeting videos generated by different methods.

With the help of our proposed metric (ME_k in (4)), we have compared the performances in maintaining temporal coherence between different methods. Fig. 4 to Fig. 5 show the comparison results with ME_k value

Figure 2. Comparison with the existing methods for "Woman" Sequence. From left to right: Original image, WARP,[6] ARAP,[10] IF[9] and SC.[3] All videos are resized to half width ($W'/W = 0.5$).

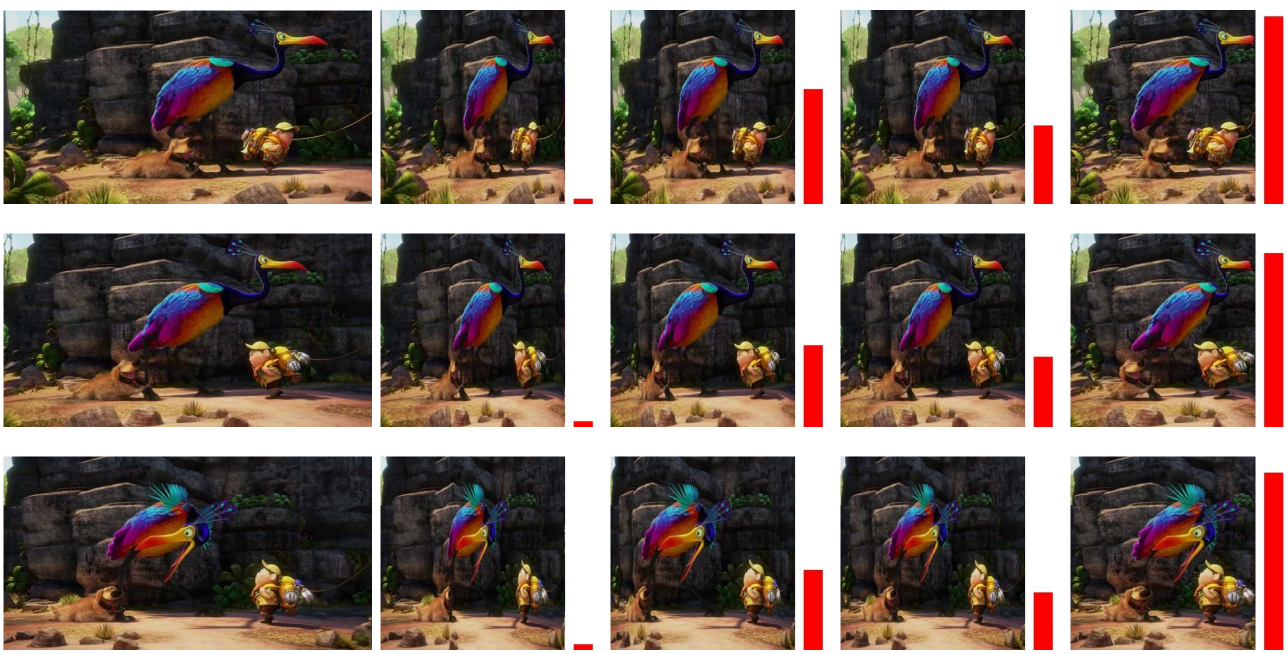

Figure 3. Comparison with the existing methods for "Bird2" Sequence. From left to right: Original image, WARP,[6] ARAP,[10] IF[9] and SC.[3] All videos are resized to half width ($W'/W = 0.5$).

versus frame number for these different test sequences with different methods. For clearer comparison of the diferences, Table 1 shows the comparison of the average metric over all frames between different methods. As mentioned before, WARP is very likely to reduce into blunt cropping or uniform scaling. As a result, its performance in maintaining temporal coherence is much better than others. This phenomenon is consistent with the much smaller metric compared with others as shown in our testing results.

In our psycho-visual testing, we use five levels mean opinion score (MOS) to evaluate our metric. The scores are between 1 and 5, in which 5 means that this metric is perfectly consistent with the user's subjective

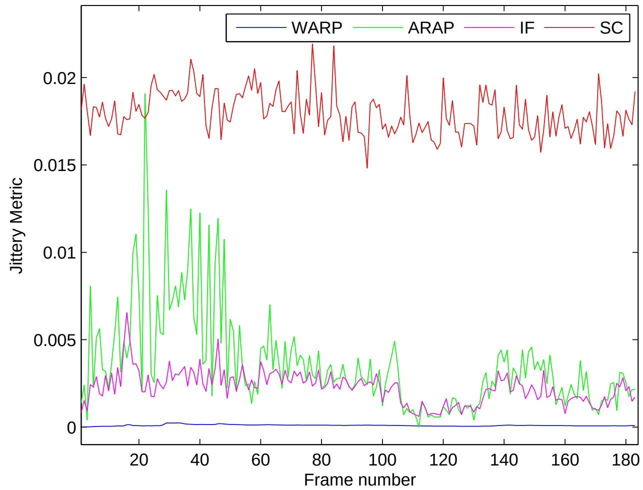

Figure 4. Artifact comparison for "Woman" Sequence between different methods.

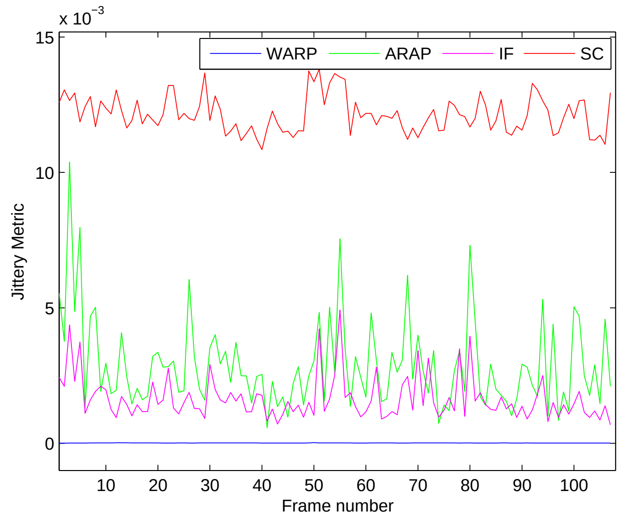

Figure 5. Artifact comparison for "Bird2" Sequence between different methods).

perception, while 1 means the metric 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 5 scores. The average score of each sequence is shown in Fig. 6. As shown in this figure, the subjective score is quite satisfying. Therefore, our proposed metric is credible to evaluate the jittery artifact.

It should be noted that our proposed metric is mainly used to assess the performance in maintaining the temporal coherence for video retargeting. The assessment of spatial coherence should use other metrics.[11] Therefore, readers may ignore the possible spatial incoherent artifacts during watching the video demos.

All the video demos in our testing can be downloaded from the supplementary multimedia material.

4. CONCLUSIONS

In this paper, we presented a general objective jittery metric to assess the performance in maintaining temporal coherence for video retargeting. This metric can be widely used in all kinds of video retargeting methods, including discrete and continuous methods. The accuracy of this metric is verified by psycho-visual tests. It possesses huge potential of practical applications.

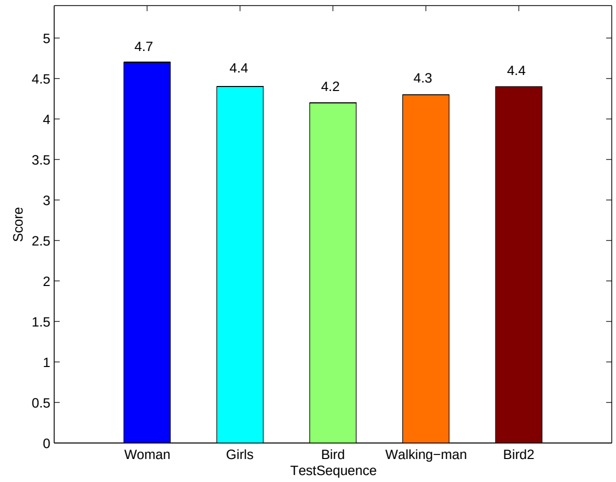

Figure 6. MOS results of our psycho-visual tests.

REFERENCES

[1] Yan, B., Sun, K., and Liu, L., "Matching-area-based seam carving for video retargeting," *IEEE Transactions on Circuits and Systems for Video Technology* **23**(2), 302–310 (2013).

[2] Shamir, A. and Sorkine, O., "Visual media retargeting," in [*ACM SIGGRAPH ASIA 2009 Courses*], 11 (2009).

[3] Rubinstein, M., Shamir, A., and Avidan, S., "Improved seam carving for video retargeting," *ACM Transactions on Graphics (TOG)* **27**(3), 1–9 (2008).

[4] Grundmann, M., Kwatra, V., Han, M., and Essa, I., "Discontinuous seam-carving for video retargeting," in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 569–576 (2010).

[5] Avidan, S. and Shamir, A., "Seam carving for content-aware image resizing," in [*ACM Transactions on Graphics (TOG)*], **26**(3), 10 (2007).

[6] Wolf, L., Guttmann, M., and Cohen-Or, D., "Non-homogeneous content-driven video-retargeting," in [*IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007.*], 1 –6 (oct. 2007).

[7] Yen, T.-C., Tsai, C.-M., and Lin, C.-W., "Maintaining temporal coherence in video retargeting using mosaic-guided scaling," *IEEE Transactions on Image Processing* **20**, 2339–2351 (Aug. 2011).

[8] Wang, Y., Fu, H., Sorkine, O., Lee, T., and Seidel, H., "Motion-aware temporal coherence for video resizing," in [*ACM Transactions on Graphics (TOG)*], **28**(5), 127 (2009).

[9] Ding, Y., Xiao, J., and Yu, J., "Importance filtering for image retargeting," in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011*], 89 –96 (june 2011).

[10] Panozzo, D., Weber, O., and Sorkine, O., "Robust image retargeting via axis-aligned deformation," in [*Computer Graphics Forum*], **31**(2), 229–236 (2012).

[11] Liu, Y., Luo, X., Xuan, Y., Chen, W., and Fu, X., "Image retargeting quality assessment," in [*Computer Graphics Forum*], **30**(2), 583–592 (2011).