How to Measure the Relevance of a Retargeting Approach?

Christel Chamaret[1], Olivier Le Meur[2], Philippe Guillotel[1],
and Jean-Claude Chevet[1]

[1] Technicolor R&I, France
{philippe.guillotel,christel.chamaret,jean-claude.chevet}@technicolor.com
[2] University of Rennes 1
olemeur@irisa.fr

Abstract. Most cell phones today can receive and display video content. Nonetheless, we are still significantly behind the point where premium *made for mobile* content is mainstream, largely available, and affordable. Significant issues must be overcome. The small screen size is one of them. Indeed, the direct transfer of conventional contents (*i.e.* not specifically shot for mobile devices) will provide a video in which the main characters or objects of interest may become indistinguishable from the rest of the scene. Therefore, it is required to retarget the content. Different solutions exist, either based on distortion of the image, on removal of redundant areas, or cropping. The most efficient ones are based on dynamic adaptation of the cropping window. They significantly improve the viewing experience by zooming in the regions of interest. Currently, there is no common agreement on how to compare different solutions. A retargeting metric is proposed in order to gauge its quality. Eye-tracking experiments, zooming effect through coverage ratio and temporal consistency are introduced and discussed.

1 Introduction

Due to the proliferation of new cell phones having the capacity to play video, new video viewing experiences on small screen devices are expected. To reach this goal, conventional contents have to be retargeted in order to guarantee an acceptable viewing comfort. Today it is generally done manually: an operator defines a cropping area with its size and its location and also controls the cropping window location temporally. Retargeting the video content is thus expensive and time consuming. Live events require short delays that manual operations cannot provide. As most of video contents are not produced with small-screen viewing in mind, the direct transfer of video contents would provide a video in which the main characters or other objects of interest may become indistinguishable from the rest of the image. An automated way, delivering a compromise between the time consumption and the retargeting relevancy, would be a high economic differentiator.

In the past, three basic video format conversion techniques have been used to cope with such problem, e.g. anamorphic distortion, letter/pillar box and

centered cropping. The anamorphism consists in applying a non-linear filtering in one direction. Letter/pillar box technique adds black rows or columns to reach the target aspect ratio. Cropping a sequence consists in extracting a subarea of the picture. The centered cropping technique corresponds to the extraction of a centered sub-window assuming that the interesting areas are located at the center. All those techniques process all frames of the sequence in the same way. The drawback of these methods lies on the fact that they are not driven by the content. More recently, many new techniques have been published. A solution is to focus on the most visually interesting parts of the video. As simple as it appears, this solution brings a number of difficulties: the first concerns the detection (in an automatic manner) of the regions where an observer would look at (usually referred to region of interest or RoI). The principle of first studies [6,4,12] is based on the use of a visual attention model. This kind of model [7,9,8,2] is able to provide a map indicating the hot spots of a scene. Once the regions of interest have been identified, a cropping window enclosing the most visually interesting parts of the picture is computed. Rather than displaying the whole picture, the content of the cropping window is only displayed. One advantage of such approach is to keep the ratio of object as well as the distance between objects in the scene. One drawback concerns the loss of the context that can undermine the scene understanding. A different approach is the famous seam-carving approach [1]. Seam carving is a method for content-aware resizing that changes the size of an image according to its content. There exists a number of variant of such approach that deals with seam-carving's drawbacks. Indeed, the initial version selects the seam that has the lowest energy. Such seam can cross important contents. Since seam-carving approach removes seams having the lowest energy, significant distortions may occur on the shapes of object. To deal with this issue, Zhang et al. [20] added geometric constraints to preserve the original shape of the objects.

Concerning the video, existing methods are based on an extension of still images solutions. As there exist spatio-temporal models of visual attention indicating the positions of the salient areas of a video sequence, a natural extension of saliency-based retargeting approach has been proposed [11,19,17,5]. Those techniques can be classified into three categories depending on the strategy used to reframe the content: crop based, warp based or a mix. For the first category, Tao et al. [17] compute saliency clusters which are temporally tracked to estimate the position of the cropping window. Limitations are mainly due to wrong detections of RoIs. Wolf et al. [19] warp pixels from the original frame to the retargeted one depending on their visual importance. An extension of the seam-carving also exist [14]. They applied a graph cut technique to connect removed energy lines. These techniques have proved a high efficiency for some content, but still allows visual distortion which may be annoying. It is interesting to note that some works have mixed different techniques. Liu et al. [11] consider three different cases dealing with different kinds of content: a static cropping window, a horizontal pan and cutting the shot into two shots. The technique favors the original aspect ratio, but the selection is performed per shot which may be

inadequate if scene content is changing over time. When facing a sparse content, Deselaers et al. [5] allow the alteration of the image by enlarging the original aspect ratio to potentially enclose more columns in the pan-scan window; one additional strategy is to zoom out by adding black stripes/pixels when RoI is spatially sparse. Some approaches [11,17] intentionally prefer preserving the aspect ratio without distortion of the original frame, although others [19,14,5] have based their algorithm on introducing local distortion of the frame for a better rendering of original content. How are these approaches assessed? How to gauge their quality? Up to now, subjective approaches [11,17,15,19,13] are the most used. Some authors [11,17] have assessed their own algorithm by giving their visual opinion. However, most of the time, a subjective comparison is performed between a new algorithm and a baseline algorithm (the seam-carving algorithm is the most used as the baseline). Another approach [5] goes further in the validation by using annotated ground truth. A hand labelling was used to identify relevant regions from unrelevant ones. The percentage of those important pixels present in the cropping window is then computed and compared to a state-of-the-art implementation. In the same vein, Chamaret and Le Meur [3] proposed to assess the quality of a retargeting algorithm by using eye tracking data. The idea was to check whether fixation points were present in the retargeted result.

In this paper, we propose a metric to assess quantitatively the quality of a retargeting approach. Section II first dresses a list of important points that a retargeting approach should obey. From these features, a quality metric is proposed. Section III presents a video retargeting method. Its performance is measured in Section IV. Finally, some conclusions are drawn.

2   What Is a Good Retargeting Algorithm and How to Measure Its Quality?

Before describing the features that a retargeting approach should follow, it is important to define the context in which we are. The context is the TV broadcast for cell phones. Two solutions to retarget the video content exist. First, the retargeting approach is performed by the cell phones. The final users can switch from the original to the retargeted video. This is the most convenient approach for a number of reasons. The first one is the right over video. As this is the final user that chooses between both versions, the video content can be modified without problem. Object's shapes, aspect ratio and distance between objects can be significantly different from the original sequence. Seam-carving, warped-based approach can be used. The second solution consists in retargeting the video sequence just before its encoding and its broadcasting over the network. In this case, this is the responsibility of the broadcasters to provide a good quality of retargeting.

In this context the retargeting algorithm must obey a number of constraints:

- the object's shapes must be kept;
- the distance between objects must be kept.

These constraints are important since they significantly influence the choice of the retargeting algorithm. For instance, the seam-carving does not respect the distance between objects. An example for a soccer game is given in figure 1. The soccer game is a good example since the distance between players is fundamental to understand the action and the game.
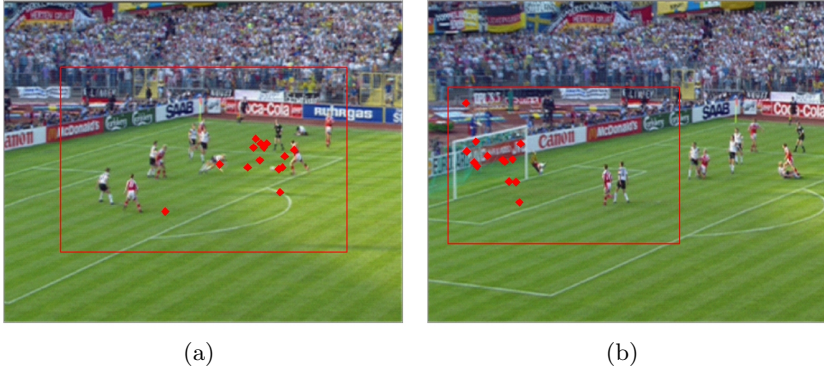

Fig. 1. Example of a retarget picture with the seam-carving approach. (a) Original picture; (b) retargeted picture.

The two constraints listed above are required in a broadcasting system. However, they do not reflect at all the quality of the final result. In order to assess the quality of a retargeted video in the context of TV broadcasting, three features are examined:

- The preservation of the visually important areas, called p_f: This first constraint of a retargeting algorithm is to keep in the final result the most visually important areas. This first property is obvious. However, it is difficult to assess automatically the extent to which a retargeting algorithm succeeds in keeping the regions of interest. In a similar vein of [3], an elegant solution would use data coming from an eye tracking experiment. From the spatial positions of visual fixations, it is easy to count the number of visual fixations that falls inside the retargeted sequence. The value p_f is the percent of visual fixation inside the cropping window (see figure 2). A database of video sequence, for which eye fixations would be available[1], might be proposed to the community.
- Temporal consistency of the cropping window center, called $c = (x, t)^T$: The previous constraint is necessary but not sufficient to draw a conclusion on the quality of the retargeted video sequence. Indeed, it is also required that the cropping window moves coherently along the sequence. A second fundamental rule would be that displacements of the cropping window should be as smooth as possible. In practice, it is not so easy to obtain due to the

---

[1] Such database already exist for still images (see for instance, `http://www.irisa.fr/temics/staff/lemeur/visualAttention/` and `http://www-sop.inria.fr/members/Neil.Bruce/`)

(a)   (b)

Fig. 2. Pictures extracted from the Sports clip. Red points correspond to visual fixations from eye-tracking experiments. Red boxes are the cropping windows.

high number of particular cases. For instance, on a still shot, it might be necessary to track a person walking. In other case, a close-up of a person moving his head does not necessary imply a displacement of the cropping window.

– Temporal consistency of the zoom, called z: In the context of this study, the retargeting approach aims at providing a better visual experience. The solution is to dynamically adapt the amount of zoom over the sequence. No matter how this zooming factor is computed, what is important is to first respect the first constraint (to keep the RoI) and to be coherent over time. However, the more the zooming factor, the more the visual experience might be. It does not mean that the zoom factor have to be high whatever the visual content. The zooming factor has to be content-dependent. This rule must be taken into account in the metric. We use the coverage ratio (CR) to measure the zoom. This is the ratio between the pixel number of the cropping window and the total number of pixel. A high coverage ratio means a low zoom in. The coverage ratio may stand for the quantity of lost data during the cropping process.

Based on these three constraints, the overall quality Q of a retargeted video sequence can be computed. The quality score is between 0 (lowest quality) and 100 (best quality). This is given by:

$$Q = f\left(p_f(t) \times \frac{100}{100 + coh_c(t)^\gamma} \times \frac{100}{100 + coh_z(t)^\beta} \times \frac{100}{100 + g(z(t), z_{opt}(t))^\alpha}\right)$$

(1)

where, N is the number of frames of the video sequence. $coh_c(t) = \left\|\frac{\partial}{\partial t}\mathbf{c}(t)\right\|$ is the temporal coherency of the cropping center window. $coh_z(t) = \frac{\partial}{\partial t}z(t)$ is the temporal coherency of the coverage factor. $g()$ is a function that computes a distance between the current zoom factor and the optimal coverage factor $z_{opt}(t)$.

In our case, $g()$ is the absolute value function. The optimal coverage factor $z_{opt}(t)$ can be deduced from the eye tracking data or fixed to an average value. α and β are coefficients that could be used to favor one particular dimension. They are all set to 1, except γ. This coefficient is set to 3 in order to strengthen the weight of the temporal consistency of the cropping window. The function $f()$ is used to pool all the quality scores to an unique one. The most common is the average function. However, as it is performed to assess the quality of a video sequence, we can use a Minkowsky pooling or a percentile-based approach. In these last solutions, the lowest $t\%$ scores are used to compute the final score. Our hypothesis is that a bad retargeting even on few pictures can dominate the subjective perception.

3   Application to a Video Retargeting Algorithm

The video retargeting algorithm used in this study is an extension to the temporal dimension of the algorithm published in [10]. We briefly describe it since the scope of this paper is to present a method to assess the quality of a retargeting approach rather than to propose a new method. Figure 3 gives the synoptic of the proposed algorithm. The starting point of the proposed method is based on the computation of a saliency map. The model proposed in [8] is used. This is a purely bottom-up model based on luminance, color and motion information. These visual information are merged to create a final/global saliency map per frame. This spatio-temporal saliency map is the first step of the reframing process. Once regions of interest have been identified, a cropping window which encloses the most important parts of the frame is deduced. This step is composed of three sequential operations:

– Window extraction: the goal of this step is to define a bounding box that encloses the most conspicuous parts of the picture. Based on the results coming from the attention model, a Winner-Take-All algorithm is applied. This algorithm allows the detection of the first N most important locations (having the highest saliency values). When the kth maximum location is selected and memorized, this location as well as its neighborhood is inhibited. Due to the inhibition process, a new salience peak will dominate and will be selected at the next iteration. The selection process is influenced by the center of the picture. Indeed, the bias of scene center has an important role: observers tend to fixate near the center of scenes, even if the salience is null. This tendency is due to a number of reasons notably detailed in [18]. Finally, it is important to underline that the value N is chosen in order to predict most of the salience of the saliency map. However, upper and lower bounds, called CR_{max} and CR_{min} respectively are used to control the amount of zoom. Note that the term zoom and coverage ratio (CR) have here the same meaning. A CR of 1 indicates that there is no zoom.
– Temporal consistency: as mentioned before, the temporal stability is likely the most important issue of a video retargeting process. The temporal stabilization acts here both on the position and the size of the bounding box. Two
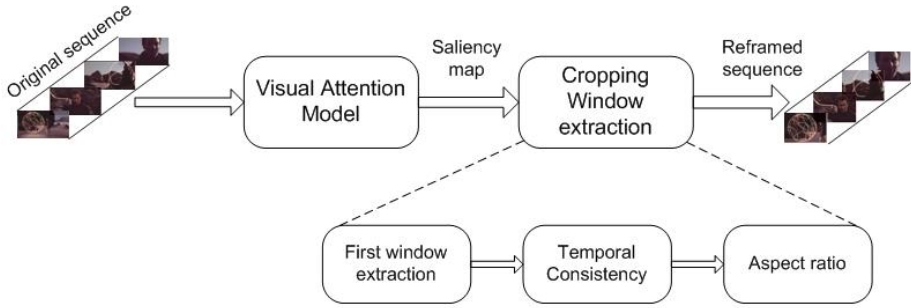
Fig. 3. General description of the proposed automatic retargeting process. Main operations are the visual attention model, the cropping window extraction and the temporal consistency.

> filters are used. A Kalman filter is first applied in order to better predict the spatial and the size of the cropping window. However, in order to deal with small displacement, a temporal median filter is used to lock the position as well as the size of the cropping window;

– Aspect ratio: as the first step (window extraction) does not guarantee the good aspect ratio, it is required to adapt the size of the window. This adaptation is arranged by extending the window size. The extension is either performed on the width or the height to reach the targeted aspect ratio.

Figure 4 gives some results of the proposed algorithm.

4   Quality Assessment

4.1   Database of Eye Tracking

Sixteen subjects have participated in the experiments. All observers had normal or corrected to normal visual acuity and normal color perception. All were inexperienced observers and naive to the experiment. Before each trial, the subject's head was correctly positioned so that their chin pressed on the chin-rest and their forehead lean against the head-strap. The height of the chin-rest and head-strap system was adjusted so that the subject sat comfortable and their eye level with the center of the presentation display.

Eye movement recording has been performed with a dual-Purkinje eye tracker from Cambridge Research Corporation. The eye tracker is mounted on a rigid EyeLock headrest that incorporates an infrared camera, an infrared mirror and two infrared illumination sources. The camera recorded a close-up image of the eye. Video was processed in real-time to extract the spatial location of the eye position. Both Purkinje reflections are used to calculate the eye's location. The guaranteed sampling frequency is $50Hz$ and the accuracy is about 0.5 degree.

Four video sequences have been selected: *Movie*, *Cartoon1*, *Cartoon2* and *Sports*. The features of those clips are given in table 1.

Fig. 4. Visual comparison of still pictures for the seam carving and dynamic reframing schemes. Top row is the computed saliency heat maps (the reddish pixels are salient, the blue ones are not). Second row is the original picture with the cropping window in white. Third row is the resulting cropped picture.

Table 1. Features of the clips used during the eye-tracking experiments

Clip	Number of observers	Spatial resolution	Length (frames)	Type
Movie	16	720×480	1000	Trailer (action)
Cartoon1	16	720×480	1200	Trailer (cartoon)
Cartoon2	16	720×480	2000	Trailer (cartoon)
Sports	16	720×480	2000	basketball, soccer, cycling...

Each sequence was presented to subjects in a free-viewing task. Experiments were conducted in normalized conditions (ITU-R BT 500-10). The spatial resolution of video sequence is 720×480 with a frequency of $50Hz$ in a progressive mode. They are displayed at a viewing distance of four times the height of the picture ($66cm$). Subjects were instructed to look around the image. The objective is to encourage a visual bottom-up behavior and to lessen the top-down effects. Analysis of the eye movement record was carried out off-line after completion of the experiments. The raw eye data is segmented into saccades and fixations. The start- and end-points of each fixation were extracted as well as the spatial coordinates of visual fixation. A visual fixation must last at least $100ms$ with a maximum velocity of 25 degrees per second.

4.2 Preservation of the Visually Important Areas p_f

The loss of the region of interest is to be avoided, not only for the viewing experience but also in order to understand the content of the sequence. The idea is to compute the ratio of visual fixations that fall into the cropping window. A ratio of 1 would mean that all regions of interest are enclosed in the bounding box. As mentioned before, this is necessary but not sufficient.

Figure 5 gives the percentage of the human fixation points that fall into the cropping window for four video sequences. Two other information are given:

the minimum percentage as well as the average value of the lowest values (10% of the lowest values are taken into account). The former is about 20% for the *Movie* and *Sports* clips and greater than 60% for the other clips. These relatively low values are due to the temporal masking induced by a scene cut [8]. After a scene cut, the spatial coordinates of the visual fixation depend on the content displayed prior the cut. This temporal shifting is due to the inability of visual system to instantaneously adjust to changes. Previous studies demonstrated that the perception is reduced after a brutal changes and can last up to $100ms$ [16]. Therefore, just after a scene cut, the cropping window is well located whereas the position of the human's gaze is still locked on areas corresponding to the content prior the cut. Then, the use of the averaged 10% lowest value is more reliable that the raw value. Results are between 60% and 80% with an average value greater than 90%, suggesting that most important areas are preserved and that the accuracy of the proposed reframing solution is very high. The worst value (60%) is obtained by the *Sports* clip. It is not surprising since this kind of content contains numerous regions of interest and the consistency in visual fixation locations is not as high as those obtained by animated sequences or movie clips.

## 4.3   Temporal Consistency of the Cropping Window Center $c = (x, t)^T$

The third validation method deals with the temporal behavior of a reframing solution. The best solution is to observe the evolution of the cropping window, the more stable the position and size of the cropping window the better the subjective quality. Figure 6 (a) depicts this evolution for the position of the cropping window (horizontal only).

   In order to highlight the role of the temporal filtering in the proposed re-targeting scheme, the location of the cropping box center is drawn in figure 6
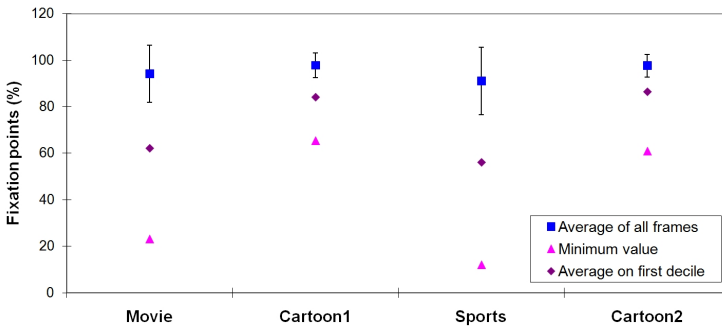
Fig. 5. Percentage of human fixation points in cropping window. Blue squares indicate the average values over time (the standard deviation is also given). The pink triangle and the purple diamond respectively correspond to the frame with the lowest percentage and the average of the 10% lowest values.
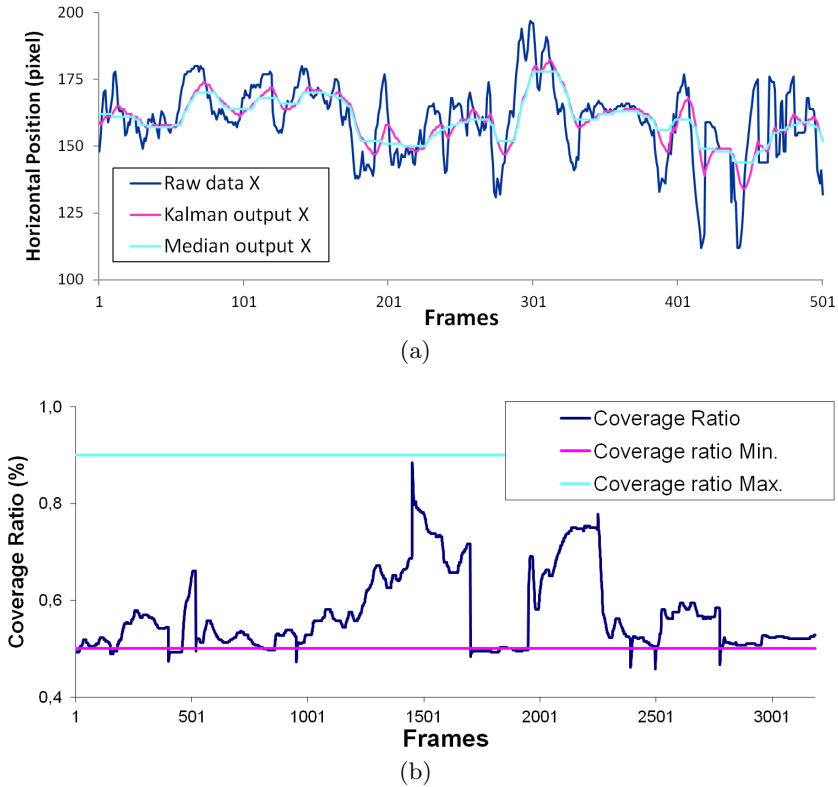
Fig. 6. Temporal evolution of the center of the cropping window (just the spatial coordinate X is presented) (a) and of the coverage value (b) for the *Sports* clip. Concerning the position of the cropping window, the temporal evolution is given after the cropping window extraction, the Kalman and the median filter.

(a) considering different processing steps. The dark blue, pink and light blue curves stand respectively for the raw data, the data after Kalman filtering and data after the median filtering. The curves clearly show the role of each filtering. The kalman filter attenuates the bound to the next sample and then creates smooth trajectories between strong gaps. However, when looking at the video, the cropping window location and size move still too much or too often compared to the few changes of content even if they are changing more smoothly. Too many changes of cropping window do not lead to a natural camera effect such as an operator would shoot. The median filter is used to cope with this issue. This filtering also fixes a visually disturbing problem: the backward and forward displacement of the window. Finally, the final curve of the cropping window location reaches the objective of both a smooth trajectory and a high adaptability to video content.

4.4   Temporal Consistency of the Zoom, z

The coverage ratio, called CR, is used to measure the amount of zoom. It may stand for the quantity of lost data during the cropping process. Figure 6 (b) depictes the coverage ratio over time for the *Sports* sequence. Upper and lower bounds are used to control the amount of zoom. It is interesting to note that the coverage value depends on the scene. For instance, a classical sky view is shot for a cycling race at the frame 1000. The coverage ratio has a low value because the region of interest (typically the cyclists) covers few pixels and is not spatially spread out.

The average, minimum and maximum coverage ratios per clip are presented in Table 2. All clips have the same tendency for average and minimum statistics. The CR average is low (close to the minimum boundary), while the CR minimum is a bit inferior to the minimum boundary. Regarding the CR maximum, results are different: although the *Sports* sequence reaches a maximum of 0.88, the *Cartoon1* sequence comes up to 0.66. This difference is clearly due to the sequence content.

Table 2. Coverage ratio data for different sequences

Clip	$Avg \pm std$	Minimum	Maximum
Movie	0.54 ±0.05	0.44	0.69
Cartoon1	0.54 ±0.052	0.49	0.66
Cartoon2	0.51 ±0.04	0.44	0.70
Sports	0.57 ±0.08	0.46	0.88

4.5   Final Quality

The formula 1 is used to compute the final quality score. The setting are: $f()$ is the average function, $g()$ is the absolute value function. α and β are set to 1 and γ is equal to 3. The optimal coverage is arbitrary set to 0.65 for the sequences *Cartoon1*, *Cartoon2* and *Sports*. The optimal coverage for the sequence *Movie* is set to a smaller value (0.5) due to the presence of black stripes.

Table 3 gives the average quality scores over these sequences. These scores are given after the different filters used in the proposed algorithm. Results indicate that the quality increases when the temporal filters are used. These results are consistent with our subjective perception. Table 4 gives the distribution of the quality scores per quartile. The quality scores of the final retargeted sequence are much more uniformly distributed than the two other distributions. This is again consistent with our own perception. However, there still exist a number of problem since the quality scores of the first quartile is dramatically weak. Several reasons can explain it: first, the performance is strongly tied to the ability of the computational model of visual attention to predict the RoI. Second, we did not handle the scene cut in the computation of the quality scores. Finally, the proposed metric does not handle a smooth and coherent displacement of the cropping window.

Table 3. Quality score for the sequences (average with the standard error of the mean. A value of 100 indicates the best quality.

Clip	Optimal Coverage	Avg ± sem		
		Original data	Filtered data (Kalman)	Filtered data (Kalman+median)
Movie	0.5	50.31 ± 0.181	71.85 ± 0.157	81.7 ± 0.127
Cartoon1	0.65	59.99 ± 0.177	78.48 ± 0.139	84.64 ± 0.109
Cartoon2	0.65	50.34 ± 0.126	67.86 ± 0.102	73.25 ± 0.08
Sports	0.65	43.69 ± 0.135	73.96 ± 0.107	77.27 ± 0.101

Table 4. Distribution of the quality per quartile for the sequence *Sports*. The first quartile represents the lowest quality scores.

Clip	Original data	Filtered data (Kalman)	Filtered data (Kalman+median)
First	2.12×10^{-7}	4.92×10^{-5}	8.43×10^{-6}
Second	1.67	69.01	74.02
Third	42.32	82.95	86.16
Fourth	82.09	88.25	88.94

5   Conclusion and Future Work

This paper proposes a metric to assess the quality of a video retargeting algorithm. This metric is based on four fundamental factors: the capacity to keep the visually interesting areas in the retargeted sequence, the temporal coherence of the cropping window, the temporal coherence of its size and the ability to be close to an optimal zoom factor.

Such metric requires to collect the human visual fixations. At first sight, it might seem too complex and time-consuming. However, as it was done for the video/image quality assessment, we believe that such databases are necessary to benchmark the different retargeting algorithms. It will be required to make in a near future a comparison between the proposed metric and user studies. Indeed, it will be important to check whether the proposed metric match the user's preferences. Moreover, the different setting used here might be learned to reflect our perception.

In the future, we will endeavor to provide to the community such databases and to make a benchmark between different approaches by using the proposed quality metric.

Supplement materials are available at `http://www.thlab.net/~guillotelp/`.

References

1. Avidan, S., Shamir, A.: eam carving for content-aware image resizing. ACM Transactions on Graphics, SIGGRAPH 26 (2007)

2. Aziz, M.Z., Mertsching, B.: Fast and robust generation of feature maps for region-based visual attention. Image Processing 17(5), 633–644 (2008)
3. Chamaret, C., Le Meur, O.: Attention-based video reframing: validation using eye-tracking. In: ICPR (2008)
4. Chen, L., Xie, X., Fan, X., Ma, W., Zhang, H., Zhou, H.: A visual attention model for adapting images on small displays. ACM Multimedia Systems Journal 9(4) (2003)
5. Deselaers, T., Dreuw, P., Ney, H.: Pan, zoom, scan – time-coherent, trained automatic video cropping. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2008)
6. Fan, X., Xie, X., Ma, W., Zhang, H., Zhou, H.: Visual attention based image browsing on mobile devices. In: ICME 2003, vol. 1, pp. 53–56 (2003)
7. Itti, L., Koch, C.: Model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
8. Le Meur, O., Le Callet, P., Barba, D.: Predicting visual fixations on video based on low-level visual features. Vision Research 47(19), 2483–2498 (2007)
9. Le Meur, O., Le Callet, P., Barba, D., Thoreau, D.: A coherent computational approach to model the bottom-up visual attention. IEEE Trans. on Pattern Analysis and Machine Intelligence 28(5), 802–817 (2006)
10. Le Meur, O., Castellan, X., Le Callet, P., Barba, D.: Efficient Saliency-Based Repurposing Method. In: IEEE International Conference on Image Processing, pp. 421–424 (2006)
11. Liu, F., Gleicher, M.: Video retargeting: automating pan and scan. In: MULTIMEDIA 2006: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 241–250. ACM Press, New York (2006)
12. Liu, H., Xie, X., Ma, W., Zhang, H.: Automatic browsing of large pictures on mobile devices. In: ACM Multimedia Conference, pp. 148–155 (2003)
13. Kraehenbuehl, P., Manuel Lang, A.H., Gross, M.: A system for retargeting of streaming video. In: ACM Transactions on Graphics (Proc. of SIGGRAPH Asia) (2009)
14. Rubinstein, M., Shamir, A., Avidan, S.: Improved seam carving for video retargeting. ACM Transactions on Graphics (SIGGRAPH) 27(3), 1–9 (2008)
15. Santella, A., Agrawala, M., Decarlo, D., Salesin, D., Cohen, M.: Gaze-based interaction for semi-automatic photo cropping. In: Proceedings of ACM's CHI 2006, pp. 771–780 (2006)
16. Seyler, A.J., Budrikis, Z.: Details perception after scene changes in television image presentations. IEEE Trans. Inform. Theory 11(1), 31–43 (1965)
17. Tao, C., Jia, J., Sun, H.: Active window oriented dynamic video retargeting. In: International Conference Computer Vision (2007)
18. Tatler, B.W., Baddeley, R.J., Gichrist, I.D.: Visual correlates of eye movements: Effects of scale and time. Vision Research 45(5), 643–659 (2005)
19. Wolf, L., Guttmann, M., Cohen-Or, D.: Non-homogeneous content-driven video-retargeting. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–6 (October 2007)
20. Zhang, G., Cheng, M., Hu, S., Martin, R.R.: A shape-preserving approach to image resizing. Pacific Graphics 28 (2009)