AN EVALUATION CRITERION OF SALIENCY MODELS FOR VIDEO SEAM CARVING

Marc Décombas[a], Pierre Marighetto[a*], Matei Mancas[a], Ioannis Cassagne[a], Nicolas Riche[a],*
Bernard.Gosselin[a], Thierry Dutoit[a], Robert Laganiere[b]
*Note: * indicates equal contribution*

[a] Numediart Institute, Faculty of Engineering (FPMs), University of Mons (UMONS) – Mons, Belgium
{Pierre.Marighetto, Matei.Mancas, Nicolas.Riche, Bernard.Gosselin, Thierry.Dutoit}@umons.ac.be
{marc.decombas, ioannis.cassagne}@gmail.com
[b] University of Ottawa (UOTTAWA) – Ottawa, Canada
laganier@uottawa.ca

ABSTRACT

Modeling human attention has been arousing a lot of interest due to its numerous applications. The process that allows us to focus on some more important stimuli is defined as the "attention'. Seam carving is an approach to resize images or video sequences while preserving the semantic content. To define what is important, gradient was first used but due to limitations of this approach, saliency models, which are able to predict whether regions in the images attract human attention, are now used. Most of them are optimized to conform a ground truth like eye tracking but there is no way to know the efficiency of the saliency models applied in a specific application like in seam carving. In this paper, we propose a criterion, based on the quantity of geometric deformation and the image's reduction, which evaluate the quality of a resizing by seam carving. This criterion is applied on evaluation of image (SCES) or video (SCED) resizing. We validate our criterion with subjective evaluation and used it to rank state of the art saliency models for seam carving. Evaluation of the image by SCES gives a Spearman correlation of -0.92196 and a Pearson correlation of -0.8812. For the video, the final SCED gives a Spearman correlation of -0.81351 and a Pearson correlation of -0.80581.

Index Terms— Subjective Evaluation, Criteria, Seam carving, Saliency Models

1. INTRODUCTION

For many years, modeling human attention has been arousing a lot of interest due to its numerous applications. The term "attention" can be defined as the process that allows one to focus on some more important stimuli and has been introduced in [1][2]. As a feature is not necessary important by itself, the idea of saliency models is to highlight rare, novel or surprising features in a given spatio-temporal context like in [3][4][5][6][7][8][9][10][11][12]. A popular application of the saliency models since several years are the content aware resizing tools due to the fact that they can be used in different use cases. Seam carving [13] is one that has been extended to do video retargeting [14], content aware compression [15] or video summary [16]. Originally based on the gradient to preserve what is important, this approach has shown quickly its limitations when the background has texture and the object has not. In [17] we already demonstrated that the saliency models ranking can be significantly different if using an eye-tracking reference or a manually segmented salient object reference. Therefore, we propose that for each application, an adapted reference and metric should be provided to rank the different saliency models. In the present paper, we propose a novel reference and evaluation criterion for the efficiency of video saliency models for seam carving. At the authors' best knowledge, there are not yet research papers proposing to evaluate saliency models based on an application-driven approach (especially video seam carving).

In the next section, the resizing approach using seam carving algorithm will first be presented. Then, different saliency models will be detailed. The novel application-driven saliency models evaluation criterion and the experimental setup will be then described and the influence of saliency models on seam carving will be presented and analyzed.

2. SEAM CARVING

Resizing tools have been designed to manage screens with different sizes or aspect ratios. Content-aware resizing technique like seam carving algorithm are popular due to the fact that it preserves semantically important content [13]. A seam is an optimal 8-connected path of pixels on a single image from top to bottom or left to right. Let *I* be an *n* x *m* image, the term *vertical seam* is defined to be the set of points s^X

$$\boldsymbol{s^X = \{s_i^x\}_{i=1}^n = \{x(i), i\}_{i=1}^n, s.t. \forall i, |x(i) - x(i-1)| \leq 1},$$

with *x* the horizontal coordinate of the point.

From the original image, an energy function is used to define the important parts of the image by using gradient, background subtraction or saliency models leading to the energy map. This energy map is then used to define the seams' paths with a cumulative energy function. The seams' path position will be saved to evaluate the performance of the resizing and are also used to obtain the reduced image. The dimension of the reduced image is *n*-1 after the suppression of one seam. The same seam carving process is then iterated. On Figure 1, it is possible to see in orange the seams that are avoiding the main objects and the reduced image obtained after the suppression of the seams.

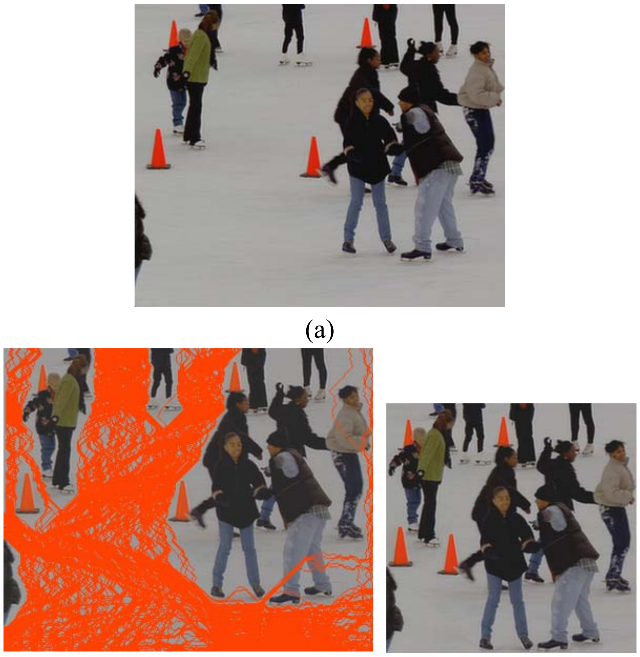


Figure 1 Illustration of the seams for ice sequence. (a) Original image, (b) Original image with the seams in orange, (c) Reduced image.

3. SALIENCY MODELS

Eight state-of-the-art video saliency models are tested in this paper to validate that the proposed criterion can measure differences between the models. The gradient-based energy map by Avidan and Shamir [13] is used as a baseline to show the interest of saliency methods. Seo and Milanfar (SEO) propose a framework for static and space-time saliency detection [3]. In [4], Culibrk *et al.* (CULIBRK) introduce a model based on motion and simple static cues. Harel *et al.* use a similar approach as Itti et al. [5] to create feature maps at multiple scales and propose a Graph-Based Visual Saliency model (GBVS) [6]. Zhang *et al.* propose a Bayesian framework for saliency detection called NMPT [7]. Mancas *et al.* (MANCAS) [8] only use dynamic features. In [9], Riche *et al.* proposes a model of rarity based on static features (colors and textures). Building upon on [8][9], a new model referred to as Spatio-Temporal RARE (STRARE) [10] is defined by using both temporal and spatial features. Following the idea from [10], a Spatio-Temporal RArity-based algorithm with prior information saliency model named STRAP [11] has been realized. Compared with [10], this model integrates several novelties like a temporal compensation of the movement on sliding windows allowing to better manage both static and moving cameras and to give more robust spatial and temporal features which are combined together using a rarity mechanism and low-level priors knowledge. The saliency model of Rahtu et al. [12] (RAHTU) has the advantage to be multiscale, does not require training and is computed in the CIE Lab perceptual color space.

4. SALIENCY-BASED SEAM CARVING

By using saliency as energy map for seam carving, we want to assess deformations made to the salient object by different models. Figure 2 illustrates the same vertical reduction with two different methods. In the case (b) and (d), 150 vertical seams are suppressed from the original image.

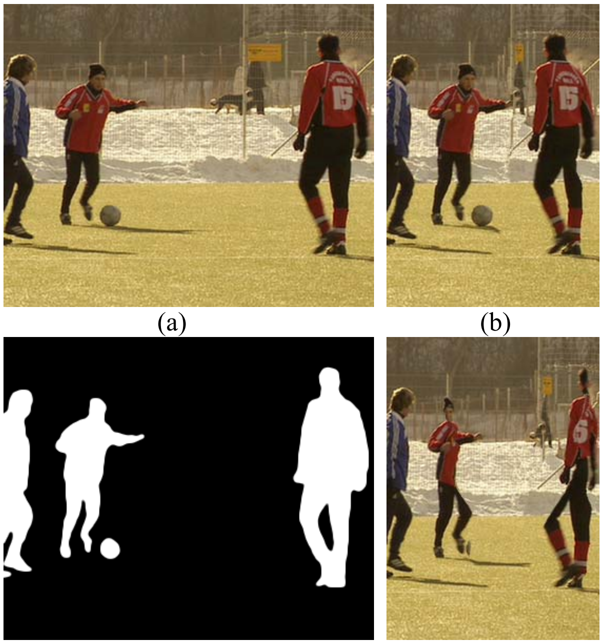
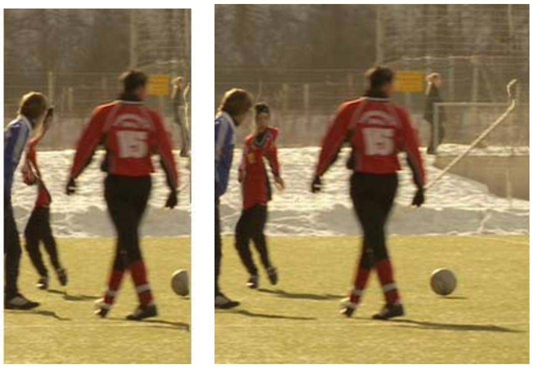


Figure 2 Result of different seam carving reductions on the soccer sequence. (a) original image, (b) vertical reduction of 150 seams with saliency map of SEO, (c) manual binary mask of the salient object, (d) vertical reduction of 150 seams with saliency map of MANCAS.

We also want, for a fixed quantity of salient object suppressed, to assess the number of vertical seam suppressed. Figure 3 illustrates the same quantity of salient object suppressed with two different methods. In the case (a) and (b), 23% of the salient object is suppressed from the original image.



(a) (b)

Figure 3 Result of seam carving reduction on the soccer sequence with SEO (a) and MANCAS (b) with a fixed quantity of salient object suppressed.

It can be seen that with SEO, less seams are passing through the salient object, leading to a better preservation of the salient object, compared to MANCAS, while more seams are suppressed to obtain the same quantity suppressed.
The goal of this criterion is to measure automatically, using these two observations, the efficiency of different methods and help to classify and optimize them.

5. SALIENCY EVALUATION CRITERIA FOR SEAM CARVING

A novel saliency evaluation criterion for seam carving is proposed. The Seam Carving Evaluation takes into account how much the salient objects are altered in each image or video frame and can be seen as a measure of the geometrical deformation of the object. This criterion's definition depends on inputs' type (image or video). For this purpose, we define the *Seam Carving Evaluation on Static input (SCES)*, as:

$$SCES = \frac{1}{2} * (ROS + RIS) \tag{1}$$

where *ROS is the Rate of salient Objects Suppressed* and *RIS is the Rate of Image Suppressed*, such as:

$$ROS = \frac{\sum_{x=1}^{X} S(x)}{\sum_{y=1}^{Y} I(y)} \tag{2}$$

$$RIS = \frac{Y}{X} \tag{3}$$

with I, the binary matrix of the original frame i (before seam carving) with X pixels and S, the binary matrix of the frame after seam carving with Y pixels. It is a ratio that represents a rate between the quantity of the salient objects suppressed and quantity of image suppressed.

We also define the *Seam Carving Evaluation on Dynamic input (SCED)*, which is the mean value of SCES computed on each frame:

$$\text{SCED} = \frac{1}{N} * \sum_{i=1}^{N} SCES(i) \tag{4}$$

with *N,* the number of frames. To define the salient objects, ground-truth manual binary segmentation masks are assumed to be available.

6. SUBJECTIVE EVALUATION PROTOCOL

Following the ITU-R BT.500-13 recommendation [18], the protocol DSIS is chosen for subjective evaluation.

During the session, the assessor is first presented with a binary mask defining the object, then an unimpaired reference, and finally with the same picture impaired. For the SCES, it is done with images and only one time. For the SCED, it is done twice with moving sequences.

At the beginning of each session, a training is given to the observers about the subjective assessment. In particular, assessors are specifically instructed to concentrate on the corresponding object. Afterwards, the assessor is asked to vote using the five-grade impairment scale: 5 imperceptible, 4 perceptible, but not annoying, 3 slightly annoying, 2 annoying, 1 very annoying.

Each assessor evaluates 30 images for the SCES and 19 sequences for the SCED, extracted from 12 test sequences in CIF format with the manually segmented binary mask ground-truth , that have been altered with different levels of artifacts spanning a large range of visual quality. The five first images are used for training and corresponding scores are discarded. Subjective scores are then processed and analyzed according to [18].

To assess our evaluation criterion, we use 8 different saliency models and the gradient. 9 different steps (26,50,100,126,150,176,200,250,300) of seam suppression are used to generate images or sequences presenting geometrical deformation. The evaluation was done with 15 non-expert assessors. Sequences were shown in a random order to each assessor. We have found no outliers among assessors when following this procedure. Test sequences and evaluation scores can be downloaded in [19].

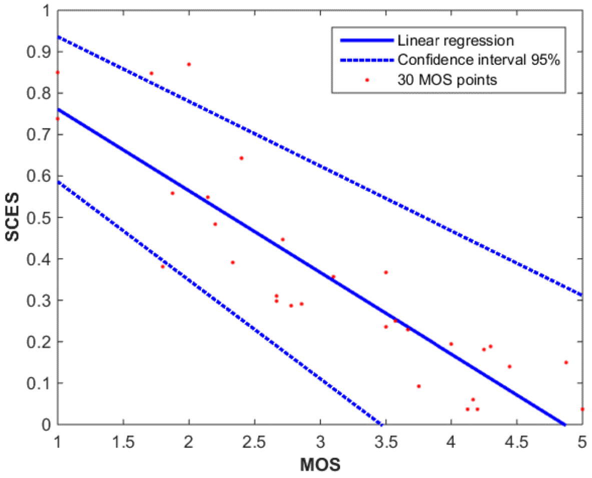


Figure 4 SCES as a function of MOS

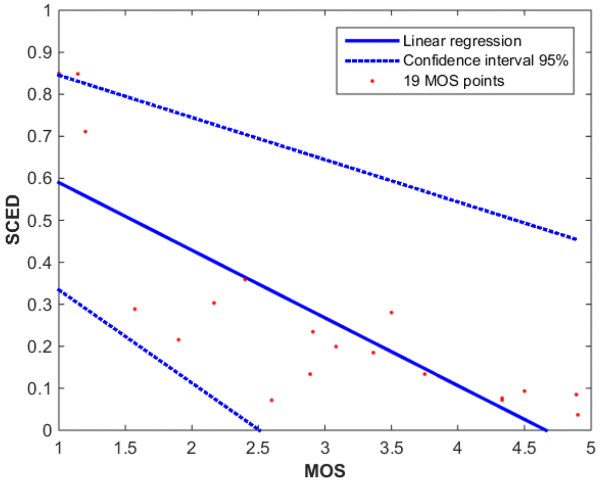


Figure 5 SCED as a function of MOS.

7. RESULTS

7.1. Performance of SCES

Figure 4 shows the proposed SCES as a function of the Mean Opinion Score (MOS). The Spearman correlation is -0.92196 and the Pearson correlation is -0.88127 showing a strong correlation.

An intuitive metric would have been to compute only the ROS on an image. In comparison, ROS gives a Spearman correlation score of -0.80227 and a Pearson correlation score of -0.80667, showing the importance of the combination of ROS and RIS.

7.2. Performance of SCED

The result of the experiment on video's evaluation is given in Figure 5. The Spearman correlation is -0.81351 and the Pearson correlation is -0.80581. Once we prove that our proposed criterion is correlated to experiment results, we use it in the following section to rank the different saliency models in the case of seam carving.

In comparison, using ROS as a metric gives a Spearman correlation score of -0.67398 and a Pearson correlation score of -0.72251.

7.3. Saliency models ranking for seam carving

rank	**SCED**	**AUC (mask)**	**AUC (eye)**
1	STRAP	STRAP	GBVS
2	STRARE	MANCAS	STRAP
3	GBVS	STRARE	RAHTU
4	CULIBRK	GBVS	MANCAS
5	MANCAS	NMPT	STRARE
6	SEO	RHATU	SEO
7	RAHTU	SEO	NMPT
8	NMPT	CULIBRK	CULIBRK

Table 1 Comparison between our application-driven and the classical eye-tracking mean ranking

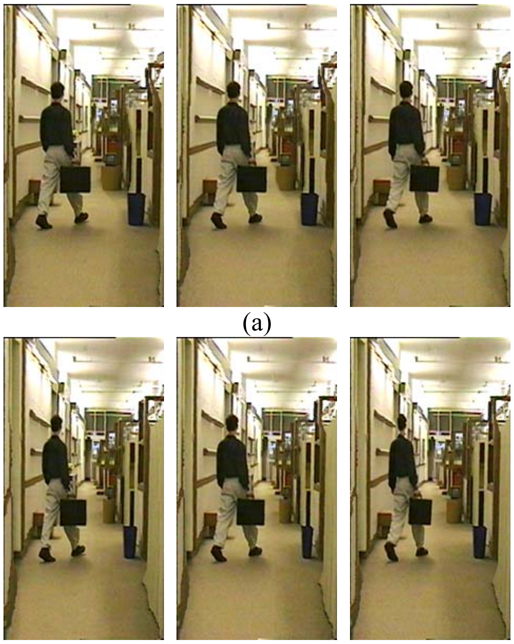
With our criterion, we can rank the saliency models for different sequences. However, it can be seen that in function of the sequences and the number of seams, the best model is not the same. To overcome this issue, SCED is computed for each model, each number of seams suppressed and each sequence. Thus, a model rank corresponds with its mean rank for each combination of sequence and number of seams suppressed.

Therefore, we performed the ranking of the 8 saliency model with our criterion, the ground-truth manual binary segmentation masks using AUC (Area Under the ROC Curve) and the eye-tracking data using the AUC.

The results show a significant difference between the three rankings and the different models. There is a Kendall's correlation of 0.35714 between SCED and AUC for the mask, and a Kendall's correlation of 0.28571 between SCED and AUC for the eye maps. It highlights the fact that a good saliency model following the eye tracking or binary segmentation masks is not a good one for seam carving.

7.4. Discussion

Figure 6 illustrates the problem of temporal instability. On (a), it can be seen that the object has on the 3 frames the same quantity of geometric deformation and seems stable



(a)

(b)

Figure 6 3 consecutive frames of the hall sequence (200 seams suppressed). (a) STRAP, (b) RHATU.

compared to (b) where on the first image, the seams in the head are suppressed on the first and third frame and not on the second which leads to a strong instability. These instability leads to a bad rating of the image or video.

A flickering effect on video can also lead to the same results. Experience shows that a strong flickering effect on a video with no geometric deformation on the salient objects will end in a worst rating than video with less flicker but more deformation.

These two aspects are an explanation to the difference between SCES and SCED results. They might also explain the Kendall's correlation found in section 6.3. Further work should take this instability and flickering into account to improve SCED and correlation between this criteria and ground-truth measures.

8. CONCLUSION

In this paper a criteria that evaluate the quality of a seam carving resizing for image (SCES) and video (SCED) was developed. It can be also used for reduced images by different content aware resizing that modify the shape of the object. The proposed component has been validated by a subjective evaluation following DSIS. Evaluation of the image by SCES gives a Spearman correlation of -0.92196 and a Pearson correlation of -0.8812. For the video, the final SCED gives a Spearman correlation of -0.81351 and a Pearson correlation of -0.80581. The MOS and the test sequences can be downloaded in [19].

In future work, we will aim to take instability on images and flickering on videos into account to improve SCED and its correlation to ground-truth measures.

9. REFERENCES

[1] C. Koch, S. Ullman, “Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry,” Human Neurobiology, vol. 4, pp. 219-227, 1985.

[2] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y.H. Lai, N. Davis, F. Nuflo, “Modelling Visual Attention via Selective Tuning,” Artificial Intelligence, vol. 78, no. 1-2, pp. 507-545, Oct. 1995.

[3] H. Seo, P. Milanfar, “Static and space time visual saliency detection by self-resemblance,” Journal of vision, vol. 9, no. 12, pp. 1-12, 2009.

[4] D. Culibrk, M. Mirkovic, V. Zlokolica, P. Pokric, V. Crnojevic, D. Kukolj, “Salient motion features for video quality assessment,” in Proc. IEEE International Conference on Image Processing (ICIP), Hong Kong, Hong Kong, Oct. 2010

[5] L. Itti, C. Koch, & E. Niebur, “A model of saliency-based visual attention for rapid scene analysis”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254 - 1259, 1998.

[6] J. Harel, C. Koch, P. Perona, “Graph-based visual saliency”, Advances in Neural Information Processing Systems, pp. 545 - 552, 2006.

[7] L. Zhang, M. Tong, T. Marks, H. Shan, G. Cottrell, “SUN : A Bayesian framework for saliency using natural statistics,” Journal of vision, vol. 9, no. 7, pp. 1-20, 2008

[8] M. Mancas, N. Riche, J. Leroy, B. Gosselin, “ Abnormal motion selection in crowds using bottom-up saliency,” in Proc. IEEE International Conference on Image Processing (ICIP), Bruxelles, Belgium, 2011

[9] N. Riche, M. Mancas, B. Gosselin, T. Dutoit, “ Rare: A new bottom-up saliency model” in Proc. IEEE International Conference on Image Processing (ICIP), Orlando, FL, Oct. 2012

[10] M. Décombas, N. Riche, F. Dufaux, B. Pesquet-Popescu, M. Mancas, B. Gosselin, T. Dutoit, “Spatio-temporal saliency based on rare model”, IEEE Proc. on International Conference on Image Processing, 2013.

[11] N. Riche, M. Décombas, M. Mancas, Y. Fellah, F. Dufaux, B. Pesquet-Popescu, B Gosselin, T. Dutoit , “STRAP: A Spatio-Temporal RArity saliency model using Priors for eye fixations prediction and objects detection”, submitted to Journal of Image and Vision Computing, Elsevier.

[12] E. Rahtu, J. Heikkila, “A simple and efficient saliency detector for background subtraction”. In Proc. in IEEE

International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1137-1144, 2009

[13] S. Avidan, A. Shamir, “Seam Carving for Content-Aware Image Resizing”, ACM Trans. on Graphics, vol. 26, no. 10, 2007.

[14] M. Rubinstein, A. Shamir, S. Avidan, “Improved seam carving for video retargeting,” ACM Trans. Graphics, vol. 27, no. 3 , pp. 1-16, 2008

[15] M. Décombas, F. Dufaux, E. Renan, B. Pesquet-Popescu, F. Capman, “Improved seam carving for semantic video coding,”, in Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP 2012), Banff, Canada, Sept. 2012

[16] M. Décombas, F. Dufaux, B. Pesquet-Popescu, “Spatio-temporal grouping with constraint for seam carving in video summary application”, IEEE Proc. on International Conference on Digital Signal Processing, 2013.

[17] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, T. Dutoit, “A study of parameters affecting visual saliency assessment”, Proceedings of the 6th International Symposium on Attention in Cognitive Systems (ISACS'13) , Beijing, China, 2013.

[18] Recommendation ITU-R BT.500-13, Methodology for the subjective of the quality of television pictures,2012

[19] Attention Web Site: http://tcts.fpms.ac.be/attention