Closed loop Seams Approximation for Video Compression

M. Décombas, E. Renan, F. Capman
Laboratoire MMP
Thales Communications & Security
92700 Colombes, France

F. Dufaux, B. Pesquet-Popescu
Dept. Traitement du signal et des Images
TelecomParisTech
75014 Paris France

Abstract—**Seam carving is a content-aware resizing method. Based on a seam carving approach to obtain semantic video compression, the non salient parts of the video are suppressed and the reduced sequence is encoded with H.264/AVC. To avoid a large overhead to encode the seams, the seams are regrouped together and approximated.**

In this paper, after the approximation, the seams are resynthesized similarly to the process carried out at the decoder side. The resulting information is used to define a new reduced sequence. This closed loop process leads to less geometric deformation when the initial seams are spatially scattered in the sequence.

Resizing; Seam carving; Content aware compression;

I. INTRODUCTION

Traditional video coding approaches like H.264/AVC [1] are not optimal from a psychovisual point of view as they do not explicitly consider the semantic content of the video and in particular the salient objects.

In this paper, we target defense and security application for which video transmission at low bitrate is needed due to limited infrastructures. In this context, the overall image quality is not a crucial criterion. It is preferable to maintain the semantic meaning of the scene, in such a way that users can correctly interpret the content and take decisions in critical conditions.

For this purpose, we propose a semantic content-aware video coding scheme based on seam carving, which concentrates salient information in a reduced resolution sequence. More specifically, the background is suppressed and the bitrate is greatly reduced. Some extra information is transmitted to rebuild seams at the decoder side. Therefore, the quality and position of the salient objects in the video are preserved.

Our previous work [2][3] is based on the content-aware image resizing seam carving algorithm in [4] which efficiently suppress non semantic content. The resulting coding scheme is shown to achieve significant bitrate savings compared to conventional H.264/AVC, while the salient objects are preserved and the scene geometry is well reconstructed. To avoid expensive extra information to rebuild the seams, we have proposed to approximate the seams by regrouping them together and encoding areas where the seams are concentrated. These areas are typically between salient objects. During the regrouping process, seams are associated to a representative one. However, in the presence of spatially scattered seams, some of them are not well represented. This can leads to geometric distortions as the resynthesize seams are not inserted at the same position as the original seams. Moreover, it leads to higher encoding cost of the reduced frame due to the high frequencies create during seams suppression.

In this paper, we improve upon [2][3] on a specific point. After having defined the seams to suppress, a resized image is obtained and the seams are approximated by the approach from [3]. Instead of transmitting this resized image, a new one is defined from the original image and the approximated seams. More precisely, the concept is to include the seams decoder inside the encoder. In this way, the seams at the encoder side are more similar to the seams at the decoder side than in the original approach.

The proposed semantic video coding scheme based on seam carving has been evaluated with the object-based quality metric in [5] on *Container*. The results show a significant performance gain for GEOMETRIC_SIFT, as well as a rate-distortion improvement for SSIM_SIFT.

II. SEAM CARVING REVIEW

Seam carving is an approach to resize images or video sequences while preserving the semantic content [4]. A seam is defined as an optimal – 8 connected path of pixels on a single image from top to bottom or left to right. Formally, let *I* be an *n* x *m* image, the term *vertical seam* is defined to be the set of points

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

with *x* the horizontal coordinate of the point. To define the seam, an energy function and a cumulative energy function are needed.

A. Energy function

The energy function defines the salient parts of an image to obtain an energy map. The first one based on a gradient on the luminance has been proposed by Avidan and Shamir [4]. The gradient is efficient to highlight the textured areas and the border of the objects. But the textured areas are not necessarily salient. To address this problem, improvement has been done by using more complex saliency algorithm like in [6] and [7].

B. Cumulative energy function

To define the optimal seam to remove it is necessary to use dynamic programming in the cumulative energy function applied on the energy map. Avidan and Shamir first proposed backward energy [4]. However, this function is done design to measure the consequence of seams suppression and leads to some visual artifacts. Therefore, Rubinstein *et al.* proposed in [8] to use forward energy, defined as:

$$M(i,j) = e(i,j) + \begin{cases} M(i-1,j-1) + C_L(i,j) \\ M(i-1,j) + C_U(i,j) \\ M(i-1,j+1) + C_R(i,j), \end{cases}$$

where

$$C_L(i,j) = |I(i,j+1) - I(i,j-1)| + |I(i-1,j) - I(i,j-1)|$$
$$C_U(i,j) = |I(i,j+1) - I(i,j-1)|$$
$$C_R(i,j) = |I(i,j+1) - I(i,j-1)| + |I(i-1,j) - I(i,j+1)|$$

and $e(i,j)$ is an additional pixel based energy measure, for instance an energy function, $|.|$is the absolute value and I the image.

C. Compression application

Seam carving has been applied to image/video compression [2] [3][6][11][12][13][14][15].

Anh et al. note in [6] that existing spatial scalable coders only support dyadic resolutions and are not content-aware. So they introduce a content-aware multi-size image compression based on seam carving. The image is spatially reduced until touching the Region Of Interest (ROI). The reduced image is encoded and the position and the content of the seams are also encoded. This leads to an important cost to represent the seam information. Moreover, severe block-artifacts occur on the boundaries of the ROI and non-ROI regions.

To solve the problem of block artifacts in [6], Deng combine in [11] the advantages of seam carving and wavelet-based coding to obtain a novel content-based spatial-scalable compression scheme.

At the opposite, to avoid as much as possible the overhead information, a seam can be simplified by a straight line. This approach, called data pruning, has been used by Võ in [12] to spatially reduce the frames and encode them with H.264/AVC and has been used by Wang et al. in [13] to reduce the temporal dimension. The main problem of the data pruning is the strong constraint on the seams which may lead to visual distortions.

Tanaka et al. propose a compromise between the two approaches. The seam positions are encoded using piecewise straight lines [14] and the temporal aspect is added in [15].

All these techniques perform well while the bitrate is sufficiently high. But at very low bitrates, the overhead information for the seam positions becomes too significant and it is preferable to use a traditional coder.

To solve this problem that appears at very low bitrates, the works in [2] and [3] approximates the seams by regrouping them and only encoding some key points where the seams are concentrated. This idea is based on the observation that avoiding the salient objects, the seams are concentrated between them. In [2], the key points are detected using a concentration criterion and the seams on the lines having the highest concentration points are regrouped and encoded. This information is used during the reconstruction of the frame to control the position of the seams. This is done by modifying the cumulative energy map at the decoder. However, the method in [2] has some limitations. The energy map has a simple design concerning the temporal aspect, the synthesis needs enhancement and the seams encoding approach is limited in the case of a high concentration points and lots of isolated seams on the same line.

The works in [3] addresses the shortcomings of [2] by introducing several improved modules leading to better performances. A new energy map with superior temporal robustness, including a better combination of the saliency and gradient maps and an improved stopping criterion has been proposed. A new way to define groups of seams based on k-median clustering has been then defined. In this approach, the frames are reduced and the seams are then approximated. During the reinsertion, some geometric artifacts can appear.

The closed loop approach consists in including the seams decoder inside the coder to obtain the approximated seams. This will lead to less difference between the seams at the decoder and the seams at the coders. These approximated seams are used to redefine a new resized image, leading to better performances. Instead of transmitting the resized image defined at the beginning, a new one is defined from the original image and the approximated seams.

III. PROPOSED SEAM CARVING FOR SEMANTIC CODING

The proposed seam carving scheme for semantic video coding is based on our earlier work in [3].

Seam carving is used to reduce the dimension of the video sequence, while still preserving the semantic relevant objects. Then, the reduced video is encoded with H.264/AVC and the seams are encoded with our proposed scheme. After transmission, the video sequence is reconstructed at the decoder side. The information about the seams is used to recover the original dimension and to preserve the scene geometry. Fig. 1 shows the global approach.

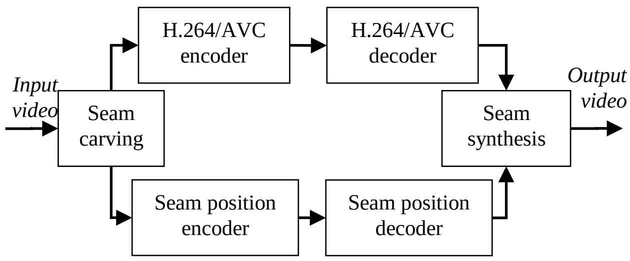

Fig. 1. Architecture of the proposed semantic video coding using seam carving

The process to define the list of seams to suppress is illustrated in Fig. 2. Taking a frame as input, an energy function is applied to define an energy map. Post-processing is applied to suppress small region and inflate the objects in the energy map. To define the seams, a cumulative energy function is then applied on the energy map. In parallel, the energy map is binarized to identify salient objects. The resulting control map is then used to govern the reduction process. More specifically, seam carving is iteratively carried out, as long as seams do not cross the salient objects in the control map.

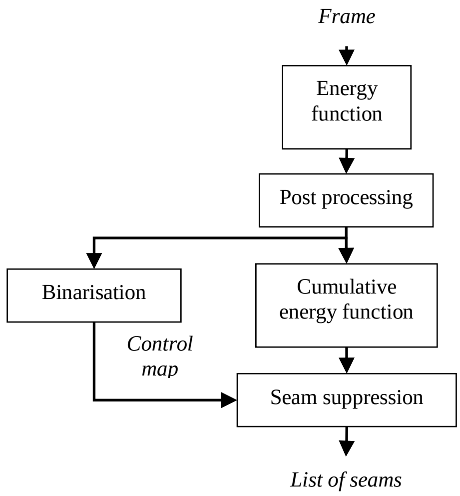

Fig. 2. List of seams process

In this paper, we introduced a closed loop approach consisting in defining a new resized image from approximated seams by reproducing the insertion process used at the decoder side. Fig. 3 illustrates the global process.

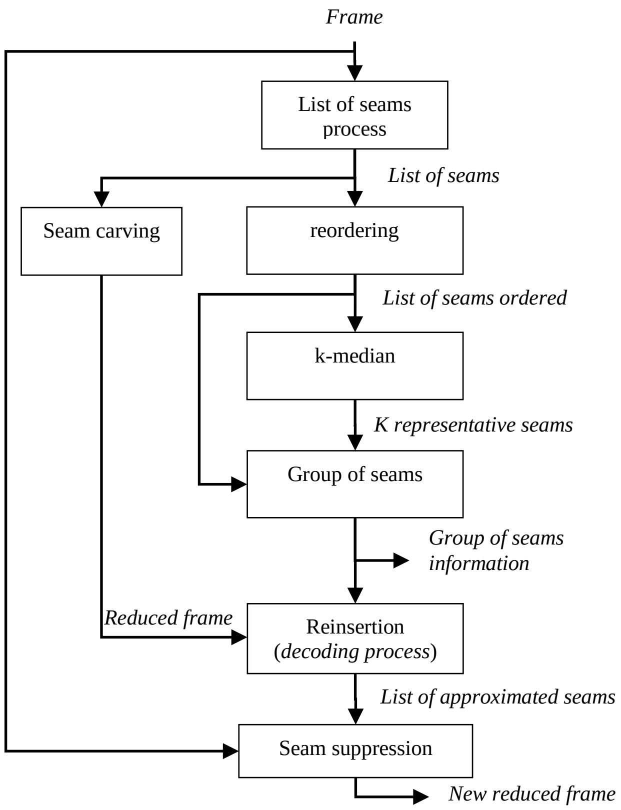

Fig. 3. Closed loop approach to redefine new resized image

A list of initial seams is first obtained by applying the process in Fig. 2. The list of seams is expressed on the original frame and reordered. In this way, K representative seams are defined from the initial list of seams. As the idea is to detect and encode the areas where the seams are concentrated, group of seams are defined by k-median clustering [3].

The next step is to simulate the seams reinsertion process taking place at the decoder side. The idea is that the seams suppressed at the encoder should closely match the seams reinserted at the decoder. For this purpose, the reduced frame is used to compute an energy map on which the groups of seams are inserted to control the position of the seams as in [3]. In this way, a list of approximated seams similar to the one obtained at the decoder is available to compute a new reduced frame. This new reduced frame is then encoded with H.264/AVC encoder and transmitted.

IV. Results

Container CIF sequence with available binary segmentation mask [21] is used to evaluate the performance of the closed loop approach. The sequence is temporally subsampled to 5 fps and we make our tests on 5 GOPs of 3 frames. 6 median seams with 12 groups of seams are defined horizontally and vertically. H.264/AVC encoding is in full Intra mode. A spatial size reduction of 40 % is reached for *Container*.

Assessing the performance of a semantic object-based video coding scheme is challenging. Common quality metrics fail when geometric deformations occur. In [5], a metric especially designed to measure the compression artifacts and the geometric deformation has been proposed. Compression artifacts due to encoders like H.264/AVC are measured by SSIM windows around SIFT points. This metric is called SSIM_SIFT. The geometric artifacts are measured by the standard deviation between the matching of the SIFT points and is called GEOMETRIC_SIFT. The Table I illustrate the influence of the closed loop approach on the GEOMETRIC_SIFT metric, compared to [3].

TABLE I
GEOMETRIC_SIFT SCORE FOR THE SEAM CARVING APPROACH FROM [3] AND THE CLOSED LOOP APPROACH

	Geometric_Sift score for the Approach from [3]	Geometric_Sift score for the closed loop approach
Container	0,52	0,17

GEOMETRIC_SIFT has been divided by 3. The initial seams are defined in the sky and in the water. During the seams approximation, all the seams are approximated in the water creating a small translation. With the closed loop approach, the translation is avoided as the seams before and after the approximation are in the water.

The Fig. 4 illustrates the rate distortion curves for the *Container* ship considered as the salient objects. The visual distortion is now expressed using SSIM_SIFT. We compare conventional H.264/AVC coding, the seam carving approach from [3] and the closed loop approach proposed in this paper.

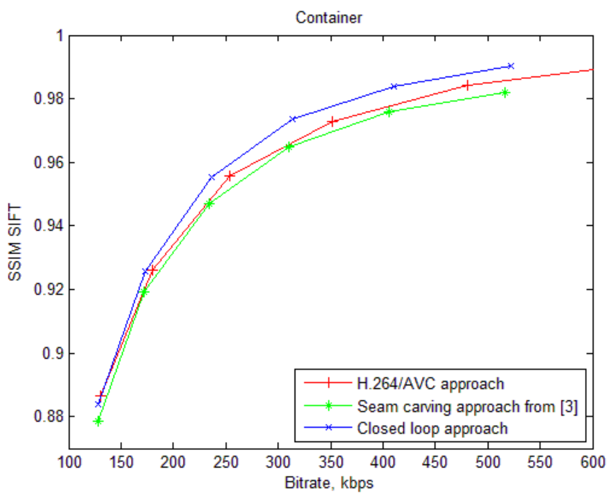

Fig. 4. RD curves for container

In Fig. 4, the closed loop approach leads to improved performance compared to the approach in [3]. Moreover, the proposed scheme also outperforms H.264/AVC. For example, at a SSIM_SIFT of 0.97, the bitrate of the H.264/AVC approach is about 352 bps. With the same SSIM_SIFT, the seam carving approach of [3] is about 405 bps and the closed loop approach reaches 314 bps. This represents a bitrate saving of 12% compared to the H.264/AVC approach and of 28% compared to the seam carving approach from [3].

Visually, the process is illustrated in Fig. 5. On the left, the original seams are visible in white. In the image in the center, the seams are approximated and reinserted in the original reduce frame. This is the result obtained at the decoder side with the approach from [2][3] and also at the encoder when the seams decoder is included. In the image on the right, the seams are reinserted in the new reduce frame and are visible at the decoder side. We remind that the decoder is included in the encoder and so the closed loop process is applied. The approximation between the first and second image is higher than the approximation between the second and the third frame. This leads to less geometric distortion.

Fig. 5. Seam visible in container. From left to right: Original seams at the encoder, Seams at the decoder with the approach from [2][3] and at the encoder with the closed loop approach, Seams at the decoder with the closed loop approach

V. Conclusion

This paper describes a new approach to limit geometric distortions in a seam carving based semantic video coding. More specifically, initial seams are approximated by simulating the seam reinsertion process. A new resized image is then determined from the approximated seams. With this approach, during synthesis at the decoder side, the reinserted seams are more similar to the seams suppressed at the encoder. Experimental results show that the proposed approach lead to better coding performance, using two object-based quality metrics, GEOMETRIC_SIFT and SSIM_SIFT.

References

[1] T. Wiegand, G.J. Sullivan, G. Bjøntegaard & A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003.

[2] M. Décombas, F. Capman, E. Renan, F. Dufaux and B. Pesquet-Popescu, "Seam carving for semantic video coding ," *SPIE Proc. Application of Digital Image Processing*, 2011.

[3] M. Décombas, F. Dufaux, E. Renan, B. Pesquet-Popescu, F. Capman, "Improved seam carving for semantic video coding", *IEEE. Proc. On MultiMedia Signal Processing*, 2012 (submitted).

[4] S. Avidan & A. Shamir, "Seam Carving for Content-Aware Image Resizing," *ACM Trans. Graphics*, Vol. 26, no. 10, 2007.

[5] M. Décombas, F. Dufaux, E. Renan, B. Pesquet-Popescu and F. Capman, "A new object based quality metric based on SIFT and SSIM," *IEEE Proc. Int. Conf. On Image Processing*, 2012 (submitted).

[6] N. Anh, W. Yang & J. Cai, "Seam carving extension: a compression perspective," *IEEE Proc. on International Conference on Multimedia*, pp. 825-828, 2009.

[7] D. Domingues, A. Alahi & P. Vandergheynst, "Stream carving: An adaptive seam carving algorithm," *IEEE Proc. International Conference on Image Processing*, pp. 901-904, 2010.

[8] M. Rubinstein, A. Shamir & S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graphics*, Vol. 27, 2008.

[9] D. Võ, J. Sole, P. Yin, C. Gomila & T. Nguyen, "Selective Data Pruning-Based Compression Using High-Order Edge-Directed Interpolation," *IEEE Transactions on Image Processing*, vol. 19, pp. 349-409, 2010.

[10] M. Bertalmio, A.L. Bertozzi, G. Sapiro, et al., "Navier-stokes, fluid dynamics, and image and video inpainting;" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, 2001.

[11] C. Deng, W. Lin & J. Cai, "Content-Based Image Compression for Arbitrary-Resolution Display Devices," *IEEE International Conference on Communications*, pp. 1-5, 2011.

[12] D. Võ, J. Sole, P. Yin, C. Gomila & T. Nguyen, "Selective Data Pruning-Based Compression Using High-Order Edge-Directed Interpolation," *IEEE Transactions on Image Processing*, vol. 19, pp. 349-409, 2010.

[13] T. Wang & K. Urahama, "Cartesian resizing of image and video for data compression," *IEEE Region 10 Conference*, pp. 1651 -1656, 2010.

[14] Y. Tanaka, M. Hasegawa & S. Kato, "Image coding using concentration and dilution based on seam carving with hierarchical search," *IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 1322-1325, 2010.

[15] Y. Tanaka, M. Hasegawa & S. Kato, "Generalized selective data pruning for video sequence," *IEEE Proc. International Conference on Image Processing*, 2011.

[16] Wang, A.C. Bovik, H.R. Sheikh & E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[17] D. Lowe, "Distinctive image features from scale invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[18] D. Azuma, Y. Tanaka, M. Hasegawa, & S. Kato, "SSIM based image quality assessment applicable to resized images," *IEICE Technical. Report*, vol. 110, no. 368, pp. 19-24, 2011.

[19] Y.J. Liu, X. Luo, Y.M. Xuan, W.F. Chen & X.L. Fu, "Image Retargeting Quality Assessment," *Eurographics*, vol. 30, no. 2, 2011.

[20] E. Rahtu & J.A. Heikkila, "Simple and efficient saliency detector for background subtraction," *IEEE 12th International Conference on Computer Vision Workshops*, pp. 1137-1144, 2009.

[21] ftp://ftp.tnt.uni-hannover.de/pub/MPEG/MPEG4_masks