# Improved Seam Carving for Semantic Video Coding

M. Décombas[*#1], F. Dufaux[*2], E. Renan[#3], B. Pesquet-Popescu[*4], F. Capman[#5]

[#] *Thales Communications & Security, Laboratoire MMP,*

*160 Boulevard de Valmy, 92700 Colombes France*

[1] marc.decombas@thalesgroup.com

[3] erwann.renan@thalesgroup.com

[5] francois.capman@thalesgroup.com

[*] *Telecom ParisTech, Dept. Traitement du Signal et des Images,*

*37-39 rue Dareau, 75014 Paris, France*

[2] frederic.dufaux@telecom-paristech.fr

[4] beatrice.pesquet@telecom-paristech.fr

*Abstract*—**Traditional video codecs like H.264/AVC encode video sequences to minimize the Mean Squared Error (MSE) at a given bitrate. Seam carving is a content-aware resizing method. In this paper, we propose a semantic video compression scheme based on seam carving. Its principle is to suppress non salient parts of the video by seam carving. The reduced sequence is then encoded with H.264/AVC and the seams are represented and encoded with our proposed approach. The main idea is to encode the seams by regrouping them. Compared to our earlier work, the main contributions of this paper are: a new energy map with better temporal robustness, a new way to define groups of seams using k-median clustering, and an improved background synthesis. Experiments show that, compared to a traditional H.264/AVC encoding, we reach a bitrate saving between 10% and 24% % with the same quality of the salient objects.**

## I. Introduction

Traditional video coding approaches like H.264/AVC [1] or HEVC aim at minimizing the Mean Squared Error (MSE). Moreover, they do not explicitly consider visually salient regions. As a result, they may not be optimal from a psycho-visual point of view.

We target defense and security applications with limited infrastructures, constraining video transmission at low data rates. In this context, the overall image quality is not a crucial criterion. Instead, the objective is to maintain the semantic meaning 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. In this way, the background is suppressed and the bitrate is greatly reduced. Some extra information is transmitted to rebuild seams at the decoder. Therefore, the quality and position of the salient objects in the video are preserved.

Our previous work [2] is based on the content-aware image resizing seam carving algorithm in [3] which efficiently preserves semantic content. The coding scheme in [2] is shown to achieve significant bitrate savings compared to conventional H.264/AVC, while the salient objects are preserved and the geometry scene is well reconstructed.

In this paper, we improve upon [2] on several points. Firstly, we introduce a new energy map with superior temporal robustness, including a better combination of the saliency and gradient maps, and an improved stopping criterion. A new way to define groups of seams based on k-median clustering is then proposed. Finally, an improved background synthesis is used based on Shift-map [4], which gives better visual results.

The proposed semantic video coding scheme based on seam carving has been evaluated with the object-based quality metric in [5] on three objects from the sequences *coastguard* and *container*. In the case of *coastguard*, our approach gives notably better results at all bitrates for the two objects. In the case of *container*, smaller gains are obtained.

## II. Seam carving review

Seam carving is an approach to resize images or video sequences while preserving the semantic content [3]. 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

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

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 allows defining the salient parts of an image. The first one has been proposed by Avidan and Shamir [3]and is based on a gradient on the luminance. The gradient is efficient to highlight the textured areas, although these areas are not necessarily salient. To address this problem, saliency map information has been integrated in [6] and [7].

### B. Cumulative energy function

After defining the salient part in the energy function, it is necessary to define the optimal seam to remove. This is done by dynamic programming in the cumulative energy function. Avidan and Shamir first proposed backward energy [3] to suppress paths with minimum energy. However, this function

fails to measure the consequence of seams suppression and leads to some visual artifacts. Therefore, Rubinstein *et al.* proposed in [8] to use forward energy, which takes into account seam suppression. It is 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, and I the image.

### C. Content synthesis

When seam carving is used to enlarge an image, content synthesis is needed. Linear interpolation, one of the simplest methods to synthesize missing areas, is often used in seam carving [3] and [8]. It performs well as long as seams do not cross textured areas or get too close from each other. In [9], Võ proposes a multiframe interpolation using neighbor frames. However, these interpolation techniques tend to fail when the missing area to be recovered is too large.

Domingues et *al.* propose in [7] to use the inpainting approach from Bertalmio et al.[10]. This inpainting is quite efficient for rebuilding structures, but often fails to reconstruct textured areas.

### D. Compression application

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

In [6], Anh et al. introduce a content-aware multi-size image compression based on seam carving, because existing spatial scalable coders only support dyadic resolutions and are not content-aware. They reduce the spatial dimension until touching the Region Of Interest (ROI) and encode the reduced image and all the information of the seams (position and content).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.

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

To address the problem of overhead information, a seam can be simplified by a straight line. Data pruning has been used by Võ in [12]to spatially reduce the frames and encode them with H.264/AVC. Based on the same principle, Wang et al. propose to also reduce the temporal dimension [13]**.** The problem of these approaches 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]. The temporal aspect is considered in [15].

All these techniques perform well when 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.

Based on the above considerations, the scheme in [2] approximates the seams by regrouping them and only encoding some key points. Given that the seams avoid salient objects, seams are regrouped with a concentration criterion and each group is 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 and the synthesis needs enhancement.

The new scheme proposed in this paper addresses the shortcomings of [2] by introducing several improved modules leading to better performances.

## III. Quality Metrics

Assessing the performance of a semantic content-aware video coding scheme remains a challenge.

Metrics such as Peak-Signal-to-Noise-Ratio (PSNR) or Structural SIMilarity (SSIM) [16] compare corresponding pixels or blocks in the reference and processed images. However, these metrics fail when geometric deformations occur.

Some metrics have been defined for image retargeting. Azuma et al. have defined in [18] a full reference metric based on Scale-Invariant Feature Transform (SIFT) [17] and SSIM to evaluate images resized by different retargeting algorithm. Liu et al. presented an objective metric simulating the Human Vision System (HVS) based on global geometric structures and local pixel correspondence based on SIFT in [19]. The objective of [18]**,**[19] is to evaluate resized (smaller) images. These metrics are not designed to evaluate compression artifacts and do not take into account the geometric deformations that can occur in some content-based video coding schemes.

Therefore, in [5] we have presented a metric especially designed for this problem. This metric measures two kinds of artifacts, geometric artifacts and compression artifacts. Compression artifacts due to encoders like H.264/AVC are measured by SSIM windows around SIFT points. These windows are totally included in the object, so the result is not disturbed by the synthesis of the content of the seams. To measure the geometric artifacts, the standard deviation between the matching of the SIFT points is computed. Subjective tests in [5] show that the metric achieves high correlation with Mean Opinion Scores (MOS).

## IV. Proposed seam carving for semantic coding

### A. General approach

The proposed seam carving scheme for semantic video coding is based on our earlier work in [2]. The main idea is to use seam carving 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 High Profile and the seams are encoded with our proposed scheme.

After transmission, the video sequence is reconstructed at the decoder side. Missing areas are synthesized by seam synthesis in order to recover the original dimension and to preserve scene geometry. Fig. 1 shows the global approach.

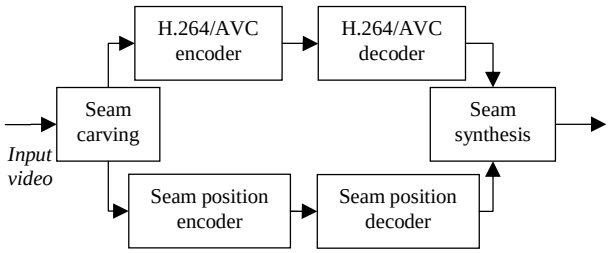


Fig. 1. Architecture of the proposed semantic video coding using seam carving

In this paper, we introduce several improvements compared to [2]. First, we propose a new energy map, achieving superior temporal robustness and providing a better stopping criterion. The stopping criterion is automatically defined to preserve the salient content. Second, we develop a new way to define and encode the seams, resulting in an improved representation. Thirdly, we propose to use the shift-map method [4] to synthesize the content at the decoder. These contributions are discussed in more details hereafter. Finally, our proposed approach is evaluated with the object-based quality metric in [5] to assess rate-distortion performance.

### *B. Proposed energy map*

The input video sequence is first divided into GOPs with a predefined size. Seam carving is next applied to each frame of the GOP as illustrated in Fig. 2. For each frame, an energy function is defined based on gradient, saliency and temporal information. Post processing is applied on this energy map to better preserve salient objects. The forward cumulative energy map is then computed to define the seam to be suppressed. In parallel, the energy map is binarized to obtain a control map. This control map is used to decide when to stop seams suppression. More precisely, seam carving is iterated until touching an object in the control map.

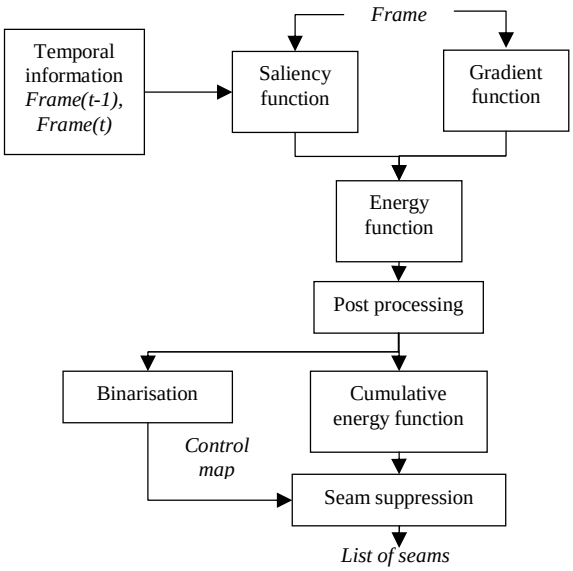


Fig. 2. Overall seam carving scheme

*1) Saliency map*: In the proposed scheme, we use the saliency map from [20]. It has the advantage to be multiscale, needs no training and is computed in the perceptual color space CIE Lab.

In order to improve temporal robustness, the motion intensity between the current frame and the previous or the next frame is integrated in the saliency map. The saliency map is defined based on 4 features: 3 from the CIE Lab color space and the motion intensity. In the case of a moving camera, the background has a global motion, whereas other objects follow local motion. In order to emphasize salient moving objects, global motion is computed and substracted from the motion intensity obtained by optical flow. Finally, a temporal weighting of the saliency map is applied. The temporal processing to compute the saliency map is illustrated in the Fig. 3. β has been empirically fixed to 0.3.

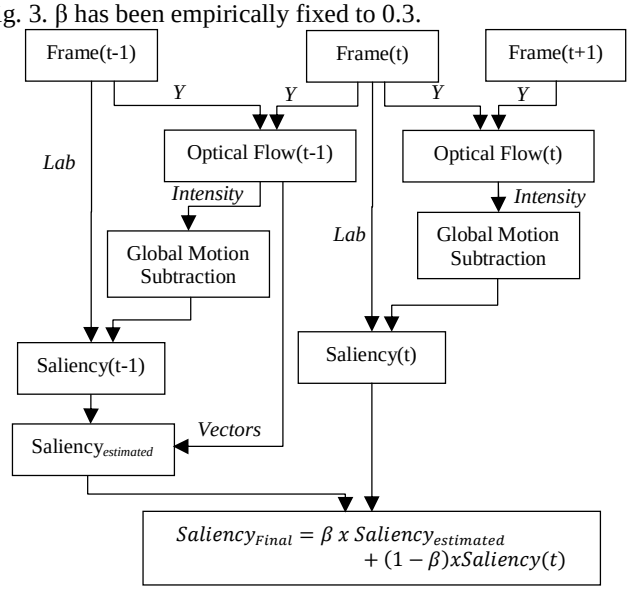


Fig. 3. Temporal processing for saliency map

*2) Energy function*: The energy function is then obtained by combining the saliency map and the gradient information. The gradient has a high value on the objects borders only. Therefore, a simple multiplication of the saliency map and gradient, as proposed in [2], may lead to the elimination of salient regions within the objects where the gradient is small.

To alleviate this weakness, a new combination is proposed here, based on a linear combination of the saliency map and the gradient

$$E_{out} = \alpha \frac{E_{Grad} - \min(E_{Grad})}{\max(E_{Grad}) - \min(E_{Grad})} + \cdots$$
$$(1-\alpha) \frac{E_{Saliency} - \min(E_{Saliency})}{\max(E_{Saliency}) - \min(E_{Saliency})}$$

where $E_{Grad}$ is the energy obtained from the gradient, $E_{Saliency}$ the energy obtained by the process defined in Fig. 3 and α is a weighting parameter empirically fixed to 0.3. This combination allows a better preservation of the information obtained from the saliency map (in particular inside the objects) and enhances the importance of the objects borders.

3) *Post-processing*: In [2], median filter and dilatation are applied on the control map in order to improve salient objects detection. This approach has the advantage to well preserve salient objects but stops the seam suppression too early. To alleviate this shortcoming, the same post-processing is now applied to the energy map.

In this way, the reduction process can go further with a better preservation of the salient objects, because the seams will avoid the area highlighted in the control map, as the control map is only a binarisation of the energy map.

4) *Isolated seams discarding*: It is desirable to keep the spatial dimension for the GOP constant and in multiple of 16. In this way, no pixel padding is needed and the effectiveness of the subsequent H.264/AVC video coding is improved.

For this purpose and as the number of seams to suppress is not constant from frame to frame, we propose in the present scheme to discard isolated seams. More specifically, For each seam, the distance between the current seam and the seam at his left, respectively at his right, is computed. Next, the lowest value among the two distances is kept and associated to the current seam. Finally, the seam having the largest distance is suppressed. The process is iterated till the frame size reaches a multiple of 16 to avoid pixel padding.

### *C. Proposed seam carving clustering*

Seam carving is first applied in the vertical direction and then in the horizontal direction. In each direction, seam carving suppresses seams one by one in an iterative process.

The coordinate of the seams are defined relative to the original image dimension. Next, all the seams are rearranged by ordering the horizontal, respectively vertical, coordinates in an increasing order. In this way, the seams are defined from left to right, respectively top to bottom, and they do not intersect. This is illustrated with 3 seams in Fig. 4.

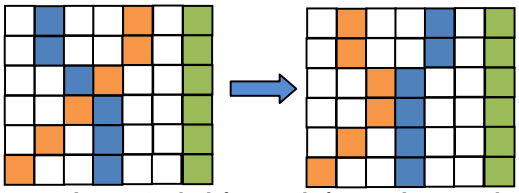

Fig. 4. Seams reordering : On the left, seams before reordering, on the right, seams reordered

In order to constraint the amount of overhead information needed to represent seams, they are clustered into groups of seams whose coordinates will be transmitted. For this purpose, a k-median algorithm is used to find median seams. The number of median seams is defined by the user. Hereafter, the number of median seams has been empirically chosen in a range between 3 and 6. Each seam is then associated to a cluster by minimizing its Euclidian distance.

The median seams are used to define the coordinates of each cluster. Given that the seam carving reinsertion process is performed from left to right and therefore pushes the right part of the image, it is necessary to translate the median seams coordinates accordingly. By this way, after the reinsertion, the groups of seams are properly centered.

We now identify concentration areas in the perpendicular direction in order to define group of seams. We have noted that the seams are concentrated between salient objects. The coordinates of these concentration areas have to be encoded to correctly reposition the salient objects at the decoder. To find these areas, for each median seam, the cumulative Euclidean distance between the median seam and its associated group of seams is calculated. This is done for each point of the median seam. The concentration areas are then defined as the points with shortest distances.

To avoid keeping many points from the same concentration area, the seam is beforehand divided into subparts and the minimum cumulative Euclidean distance is searched for each subpart. The number of subparts is defined by the user. Clearly, defining more subparts results in more precise seams at the synthesis. An example is shown in Fig. 5 where 6 seams are represented. In this example, 2 median seams are defined but not represented here and the seams are divided into 2 subparts. The red rectangle represents the group of seams.

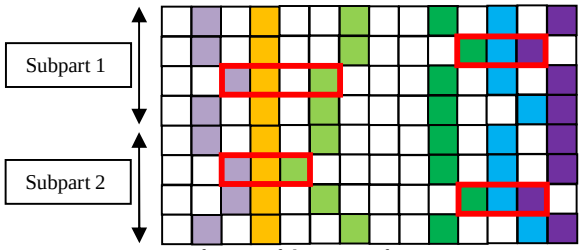


Fig. 5. Definition of the group of seams

The position of the group of seams, and the number of seams for each group is encoded.

All the process is resumed in the Fig. 6.

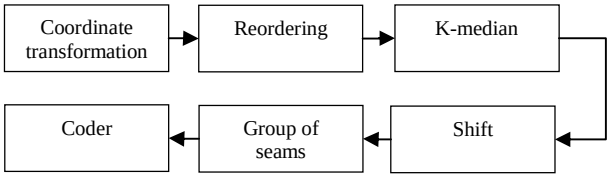


Fig. 6. Seams encoding by regrouping approach

### D. Proposed synthesis

In seam carving, synthesis commonly uses linear interpolation. This interpolation performs reasonably well when the seams do not cross textured areas or are not close to each other. The use of saliency map to create the energy map allows the seams to cross non salient textured areas and preserve the salient parts. Moreover, our cluster-based seams positions encoding approach ensures that the seams are grouped together. Therefore, linear interpolation is not suited for our seam carving approach.

To solve this problem, the inpainting algorithm from Svarm and Strandmark based on Shift-map has been chosen [4].This algorithm is based on energy minimization over a large space of labels and is optimized using Gaussian pyramid.

More specifically, the area to be synthesized is filled after the reinsertion of one group of seams. The synthesis is performed over a search area. On the one hand, this search area should not be too large to avoid using weakly correlated data. On the other hand, it should be sufficiently large to preserve enough information to rebuild the texture. Therefore, the search area is defined as the smallest rectangular box including the group of seams, as illustrated in blue in Fig. 7. The search area is then extended to twice the number of seams in the group, in order to have enough texture information, as represented in red in the Fig. 7.

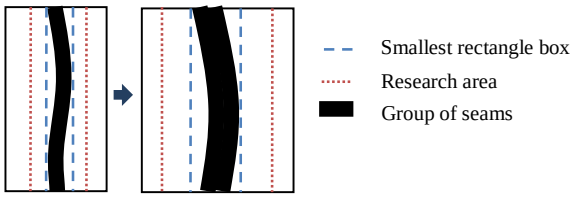


Fig. 7. Example of seam synthesis

## V. Results

To assess the performance of the proposed coding scheme, we use 2 CIF sequences with available binary segmentation masks [21]: c*oastguard* and *stefan*. The sequences are temporally subsampled to 5 fps and we make our tests on 5 GOPs of 3 frames. 6 median seams with 3 key points are defined horizontally and vertically. The threshold of binarisation has been fixed at 0.75 for *coastguard* and 0.9 for *stefan*. These parameters have been empirically defined. The encoding is in full-Intra mode.

With the proposed seam carving scheme, the process is iterated until touching an object in the control map. A size reduction of 41.6 % is reached for *coastguard* and 18.9% for *stefan*. In comparison, a size reduction of 17.8% was reached for *coastguard* and 10 % for *stefan* with the seam carving scheme in [2]. The percentage of bitrate saved compared to H.264/AVC High Profile for different methods is illustrated in Table I.

TABLE I
PERCENTAGE OF BITRATE SAVED COMPARED TO H.264/AVC AS A FUNCTION OF QPI: (1) NO SEAM CODING, (2) OUR PROPOSED APPROACH, (3) TOTAL SEAM CODING, (4) APPROACH FROM [2]

| | coastguard | | | | stefan | | | |
|---|---|---|---|---|---|---|---|---|
| R | 41.6% | | | 17.8% | 18.9% | | | 10% |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| 24 | 34.3 | 34.0 | -19.6 | 8.43 | 11.5 | 11.3 | -8.05 | 1.17 |
| 27 | 33.3 | 32.9 | -38.7 | 8.03 | 11.1 | 10.9 | -13.4 | 0.89 |
| 30 | 31.8 | 31.1 | -70.8 | 7.06 | 11.0 | 10.7 | -20.5 | 0.84 |
| 33 | 30.5 | 29.6 | -116.8 | 6.47 | 11.0 | 10.6 | -30.5 | 0.60 |
| 36 | 28.6 | 27.2 | -196.6 | 5.68 | 11.3 | 10.7 | -45.1 | 0.60 |
| 39 | 26.3 | 24.1 | -317.0 | 3.86 | 11.5 | 10.8 | -64.8 | 0.41 |
| 42 | 22.8 | 19.3 | -538.7 | 1.26 | 11.7 | 10.6 | -99.2 | 0.29 |

Compared to an approach without seam encoding, where the objects are deformed, our approach represents a small overhead. At the opposite, an approach where all the seams are encoded cannot be envisaged for low bitrate applications. Compared to the approach in [2], we reach a higher reduction due to a better seams selection. By this way, for QP=42, the bitrate saved reaches 19.3% and 10.6% compared to H.264/AVC for *coastguard* and *stefan* respectively. The gain is due to the proposed seam definition and representation.

Fig. 8, Fig. 9 and Fig. 10 illustrate the rate distortion curves for the two boats of *coastguard* and *stefan* considered as the salient objects. To measure the distortion, we use the object-based quality metric developed in [5]. It is clear that for *coastguard*, our approach gives notably better results at all bitrates. For *stefan*, seams have suppressed the crowd leading to good performance gains.

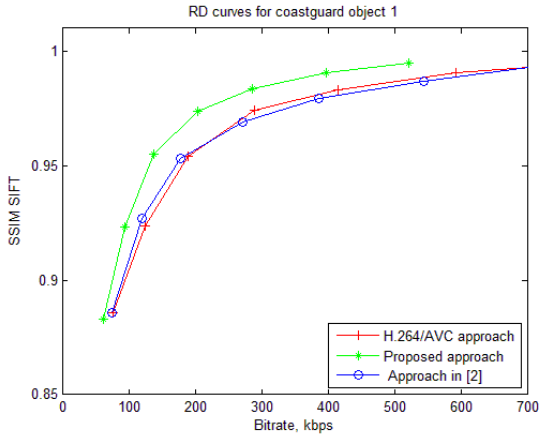


Fig. 8. RD curves for coastguard – Object 1

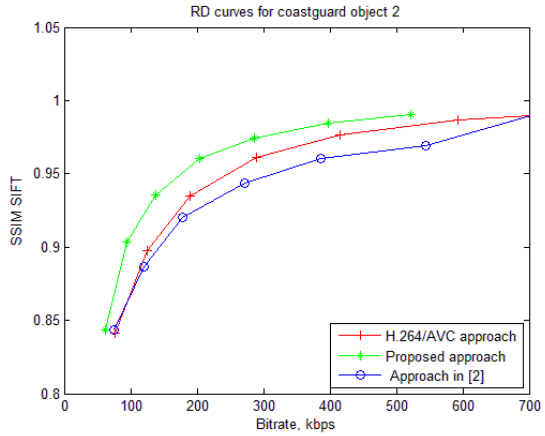


Fig. 9. RD curves for coastguard – Object 2

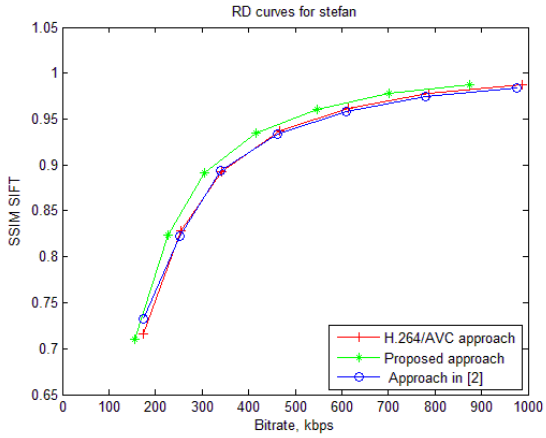


Fig. 10. RD curves for stefan

The Fig. 11 shows the visual quality of the synthesized frame obtained using the traditional linear interpolation and the proposed approach based on shift-map. The water is better synthesized. However, some artifacts remain visible due to the lack of information in this large removed zone.

Fig. 11. Synthesis results for coastguard. Left: Shift-map, Right: linear interpolation

## VI. Conclusion

This paper describes a new semantic content-aware video coding scheme based on seam carving. Compared to our earlier work [2], we have introduced several new contributions leading to improved coding performance. More specifically, we have introduced a new energy map with a better temporal aspect, as well as a better combination of the saliency and gradient maps, which allows a better preservation of the salient objects. Groups of seams are now defined with a k-median algorithm. The utilization of the Shift-map [4] to synthesize the seams content at the decoder side gives better visual results on the background. Performance assessment shows notable coding gains on the *coastguard* and *stefan* test sequence, using an object-based quality metric [5]. At a quality of SSIM=0.9, we reach a bitrate saving of 24% for the left object of *coastguard* and 10.7% for *Stefan*.

In future work, we will consider improving the Shift-map synthesis by exploiting the temporal aspects. Moreover, we will also investigate seam carving along the time axis.

## 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 & B. Pesquet-Popescu, "Seam carving for semantic video coding ," *SPIE Proc. Application of Digital Image Processing*, 2011.

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

[4] L. Svarm & P. Strandmark, "Shift-map Image Registration," *International Conference on Pattern Recognition, 2010.*

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

[6] N. Anh, W. Yang & J. Cai, "Seam carving extension: a compression perspective," *ACM 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 key points," *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