SIP (2015), vol. 4, e6, page 1 of 24 © The Authors, 2015.

doi:10.1017/ATSIP.2015.4

ORIGINAL PAPER

Seam carving modeling for semantic video coding in security applications

MARC DÉCOMBAS[1,2], YOUNOUS FELLAH[1], FRÉDERIC DUFAUX[1], BEATRICE PESQUET-POPESCU[1], FRANCOIS CAPMAN[2] AND ERWANN RENAN[2]

In some security applications, it is important to transmit just enough information to take the right decisions. Traditional video codecs try to maximize the global quality, irrespective of the video content pertinence for certain tasks. To better maintain the semantics of the scene, some approaches allocate more bitrate to the salient information. In this paper, a semantic video compression scheme based on seam carving is proposed. The idea is to suppress non-salient parts of the video by seam carving. The reduced sequence is encoded with H.264/AVC while the seams are encoded with our approach. The main contributions of this paper are (1) an algorithm that segments the sequence into group of pictures, depending on the content, (2) a spatio-temporal *seam clustering method, (3) an isolated seam discarding technique, improving the seam encoding, (4) a new seam modeling, avoiding geometric distortion and resulting in a better control of the seam shapes, and (5) a new encoder which reduces the overall bit-rate. A full reference object-oriented quality metric is used to assess the performance of the approach. Our approach outperforms traditional H.264/AVC intra encoding with a Bjontegaard's rate improvement between 7.02 and 21.77% while maintaining the quality of the salient objects.*

Keywords: Seam carving, Video compression, Seam modeling, Security application

Received 2 April 2014; Revised 10 May 2015

I. INTRODUCTION

The objective of traditional video coding approaches like H.264/AVC [1] and high efficiency video coding (HEVC) [2] is to minimize mean squared error (MSE) for a given bitrate, but they do not explicitly consider visually salient regions for the rate allocation. From a psycho-visual point of view, or for a well-defined task involving certain objects, these approaches may not be optimal. For defense and security applications, with limited infrastructure and bandwidth, video transmission is often constrained to low data rates. In this context, the transmitted information has to be the most pertinent for human understanding, at the lowest possible rate. The *overall* image quality, as generally estimated by video codecs, is not a well-suited criterion. The objective is rather that the users can correctly interpret the content and take decisions in critical conditions by maintaining the semantic meaning of the sequences. Based on these considerations, in this paper, we consider that the background can be partly suppressed, still preserving enough information to understand the context, while objects should be preserved and well-positioned.

[1]Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France
[2]Laboratory MultiMedia – Thales Communications & Security, Gennevilliers, France

Corresponding author:
M. Décombas
Email: marc.decombas@gmail.com

For this purpose, a semantic content-aware video coding scheme based on seam carving is proposed in this paper. Seam carving is a content-aware resizing method, initially introduced in [3]. The advantage of using seam carving for video compression is that salient information is concentrated in a reduced resolution sequence and the non-informative background is suppressed, leading to a significant bitrate reduction and a better preservation of the salient regions. Since the seam carving process is not reversible, the correct position of the objects may be lost during the seam reduction. It is therefore necessary to transmit some additional information about the seams in order to properly recover the original dimension of the video sequence and the position of the salient regions. In most of the papers using seam carving for image or video compression, the cost of the seams texture is too important and synthesis algorithms are usually applied to reconstruct the background. As our objective is to concentrate on transmitting the salient information, no background texture synthesis will be applied in the proposed approach. Nevertheless, any kind of inpainting algorithms [4] or other synthesis algorithms could be used for this purpose.

Approaches in [5–7] use seam carving to perform image compression, trying to find at each iteration an optimal seam that minimizes a cost function combining coding cost and visual distortion. Moreover, in order to obtain a reduced resolution video without annoying temporal artifacts, the

seams have to be temporally linked. The temporal constraint on the seams during their computation further reduces the flexibility. The ratio of spatial resizing is also often limited due to the lack of flexibility when fast moving objects are crossing the scene. Alternatively, in [8, 9], the temporal constraint is added after having computed all the suppressible seams, showing better efficiency. Significant bitrate savings have been reached compared with conventional H.264/AVC, while salient objects are preserved and the scene geometry is well reconstructed. But some limitations remain. The seam carving is applied on group of pictures (GOPs) with a predefined length. As the same number of seams has to be suppressed in each GOP, one frame with several salient objects will hinder the seam carving process.

This paper builds upon our earlier work in [8, 9] and proposes new contributions, as follows: (1) a content-aware adaptive GOP segmentation algorithm that automatically defines the GOP depending on the content, (2) a spatio-temporal seam clustering based on spatial and temporal distances, (3) an isolated seam discarding algorithm, improving the seam encoding, (4) a new seam modeling approach, avoiding geometric distortions and resulting in a better control of the seam shapes at the decoder without the need of a saliency map, and (5) a new encoder that reduces the number of bits to be transmitted. Our experiments show that, using full intra coding and compared with a traditional H.264/AVC encoding, we achieve better performances for numerous test sequences, leading to bitrate savings between 7.02 and 21.77% using the Bjontegaard's metric in conjunction with a full reference object-oriented quality metric.

The remaining of this paper is structured as follows. We will first give an overview of the related work in Section II. In Section III, the proposed approach with its new contributions will be presented. Section IV will introduce the methodology of evaluation and the influence of the different parameters on the final results. For different sequences, the encoding cost of the seams and the rate-distortion performances will be assessed. Our proposed scheme will be compared with a traditional H.264/AVC encoding and the bitrate savings will be detailed. Section V will conclude this paper and present some perspectives for future work.

II. RELATED WORK

Seam carving has been proposed by Avidan and Shamir and developed to resize images or video sequences while preserving the semantic content [3]. This iterative algorithm suppresses/adds at each iteration a seam in the image, passing through its less significant parts. A seam is an optimal 8 connected path of pixels in an image, crossing from left to right or from top to bottom. Let I be an $N \times M$ image. A *vertical seam* is characterized by the set of points:

$$s^X = \left\{s_i^x\right\}_{i=1}^{N} = \{x(i), i\}_{i=1}^{N}, s.t. \quad \forall i, |x(i) - x(i-1)| \leq 1, \tag{1}$$

where x is the horizontal coordinate of the point. Similar to the removal of a row or column from an image, removing the pixels of a seam from an image has only a local effect: all the pixels of the image are shifted left (or up) to compensate the missing path. The cost of a seam is defined as a cumulative energy function. The energy function highlights the important parts in the image and the cumulative energy function defines the optimal path.

A) Energy function

The energy function e defines the salient parts of an image. In [3], Avidan and Shamir proposed an energy function based on a gradient on the luminance, which has the advantage to efficiently highlight the borders of the objects. However, textured areas are not necessarily linked to salient parts of the image and if a salient object is smooth and the background is textured, the seam carving will first suppress the salient object. To solve this kind of problems, Anh *et al.* use a combination of a saliency map and the magnitude of gradient in the image [5]. Domingues *et al.* use another approach in [10] to generate saliency maps by merging several features like gradient magnitude, faces, edge, and straight line detection. In [11], Achanta and Susstrunk apply their saliency map [12] that uniformly assigns saliency values to the entire salient regions, rather than just edges or texture regions. This is achieved by relying on the global contrast rather than local contrast, measured in terms of both color and intensity features.

B) Cumulative energy function

After having defined the salient areas in the image, it is necessary to find the optimal seam.

The cumulative energy function E is used to determine the optimal path by using dynamic programming. Let I be an $N \times M$ image, s a seam, and i the pixel position index. The pixels of the path of seams s will be noted:

$$I_s = \{I(s_i)\}_{i=1}^{N} = \{I(x(i), i)\}_{i=1}^{N}. \tag{2}$$

Given an energy function e, the cost of a seam s is defined as:

$$E(s) = E(I_s) = \sum_{i=1}^{N} e(I(s_i)). \tag{3}$$

We look for the optimal seam $s*$ that minimizes the following seam cost:

$$s* = \underset{s}{\operatorname{argmin}}\, E(s) = \operatorname{argmin} \sum_{i=1}^{N} e(I(s_i)). \tag{4}$$

In the case of a vertical seam, the first step is to scan the image from top to bottom and compute the cumulative minimum energy $CM = E(S^*)$ for all possible connected seams for each entry (i, j).

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

Avidan and Shamir [3] propose to find the optimal seam path using backward energy, defined as:

$$CM(i,j) = e(i,j) + \begin{cases} CM(i-1,j-1) \\ CM(i-1,j) \\ CM(i-1,j+1) \end{cases}, \quad (5)$$

where e represents the energy function computed on the image. However, this approach suppresses the seam having the smallest energy in the image without taking into account the consequence of this suppression. After the suppression, a new border may be created and some artifacts may appear. Therefore, Rubinstein *et al.* propose in [13] to take into account the new neighbors created after the suppression of a seam. This new cumulative function, referred to as forward energy, is defined as:

$$CM(i,j) = e(i,j) + \begin{cases} CM(i-1,j-1) + C_L(i,j) \\ CM(i-1,j) + C_U(i,j) \\ CM(i-1,j+1) + C_R(i,j) \end{cases}, \quad (6)$$

with

$$C_L(i,j) = |I(i,j+1) - I(i,j-1)| + |I(i-1,j) - I(i,j-1)|, \quad (7)$$

$$C_U(i,j) = |I(i,j+1) - I(i,j-1)|, \quad (8)$$

$$C_R(i,j) = |I(i,j+1) - I(i,j-1)| + |I(i-1,j) - I(i,j+1)|, \quad (9)$$

where C_L, C_U, and C_R represent the cost of the new edges created after removing a seam and $e(i,j)$ is an additional pixel based on energy function as above. This function has proven its efficiency in numerous cases and is used in most of the seam carving literature.

C) Temporal aspects

As our objective is to use seam carving for video applications, we will shortly review different improvements that have been proposed to pass from still images to video.

Rubinstein *et al.* are the first ones to define seam carving for video retargeting [13]. Dynamic programming is replaced by graph cuts that are suitable for handling three-dimensional (3D) volumes. Temporal coherence of the energy maps is obtained through a linear combination of the temporal and spatial gradients of the luminance. As the human visual system (HVS) is more sensitive to motion than to texture, larger weighting is given to the temporal gradient. The seam can only move from one pixel to the left or to the right following the temporal axis, which can be a severe limitation in the case of a moving object that crosses the scene. Chao *et al.* [14] propose a solution to this problem of seams flexibility following the temporal axis. A seam is computed in a frame and block-based motion estimation and Gaussian masks are used to predict the coarse location of the seam in the next frame. This allows both a reduction of the search range of dynamic programming and having seams that can move with more than one pixel from one frame to another.

In [8, 9], the idea is to add temporal coherence in the saliency map to manage the temporal aspect, instead of constraining the seam. The optical flow proposed by Chambolle and Pock [15] is directly used in the saliency map, based on Rahtu and Heikkila [16], by taking into account the intensity of the movement. In addition, the optical flow is also used for temporal tracking of the saliency map. In this way, the current saliency map can be combined with the previous one to improve the temporal coherence.

D) Content-aware compression

Having reviewed existing seam carving approaches, we will now focus on compression applications. Content-aware video compression is a broad subject. An overview of different perceptual video coding approaches is first presented hereafter; thorough reviews can be found in [17–20]. Then, we will see in detail the existing seam carving techniques used in compression applications.

In [17], Wu and Rao present a review of the basics of compression and HVS modeling. They describe subjective quality evaluation methods and objective quality metrics. Some practical applications such as video codecs based on the HVS, restoration or error correction are finally presented.

In [18], Chen *et al.* address the incorporation of the human perception in video coding systems to enhance the perceptual quality. This topic is challenging, given the limited understanding and high complexity of computational models of the HVS. First, the visual attention and sensitivity modeling is treated, considering bottom-up and top-down attention modeling, contrast sensitivity functions, and masking effects. Then, perceptual quality optimization for constrained video coding is described. Finally, an overview of the impact of the human perception on new applications like high dynamic range video or 3D video is presented.

In [19], Ndjiki-Nya *et al.* survey perception-oriented video coding based on image analysis and completion. The relevance, limitations, and challenges of these coders for future codec designs are brought forward. It is also concluded that additional work on the evaluation has to be carried out to obtain a new rate-quality metric.

In [20], Mancas *et al.* present a review of human attention modeling and its application for data reduction. After a presentation of attention modeling and saliency maps, they address perceptual coding, but also the application of attention modeling for perceptual spatial resizing.

Perceptual video coding is a coding approach based on the HVS and trying to give more detailed, bitrate, to the important parts. The using of a coder and attention modeling is needed. Three main approaches can be considered: the interactive one, the indirect, and the direct one.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

The "interactive approach" requires eyes tracking device and are consequently not common at all. The idea is to follow where the user is watching and allocate more bitrate to this region. Other problems are that it works only if there is one viewer, it is dependent of the viewing distance, and it changes also with the eye tracking system. The idea to automate this system without eye tracking device is very challenging. The use of saliency maps is necessary, and some problems may appear when, for example, no salient objects are in the scene, people will watch in all the video. Two approaches are possible. The first one is the indirect approach and will modify the video before being encoded. In this approach, the coder is not modified. The second approach uses direct approach and modifies directly the coders.

In the "indirect approaches", the idea is to modify the video in input in order not to modify the coder.

Itti propose in [21] to use their model [22] and apply a smooth filter in all the non-salient regions. This allows to have a higher spatial correlation, a better prediction, and consequently to reduce the bitrate of the video. This allows to reduce by 50% the number of bit needed with MPEG-1 and MPEG-4 encoders. Another approach proposed by Tsapatsoulis *et al.* [23] combine bottom-up and top-down information in a wavelet decomposition to obtain a multiscale analysis. A bitrate saving of 10.4–28.3% with a MPEG-4 coder is obtained. Mancas apply in [24] their saliency model in image compression. An anisotropic filtering is applied on non-salient regions and allows to decrease twice the number of bit compared with the Joint Photographic Experts Group (JPEG) standards. Some approach use resizing before the encoding to reduce the bitrate and obtain more flexibility on the spatial dimension of the image or video.

In the direct approach, the coder is directly modified to reduce the quantity of information to encode. These approaches can also use image synthesis.

Li *et al.* [25] use a saliency map to generate a guidance map that will modify the quantization parameter of the coder. By this way, more bitrate will be allocated for the salient regions. It is underlined that some studies should be done to measure the influence of the artifacts in the non-salient areas that can become salient if the artifacts are too disturbing. Gupta and Chaudhury [26] improve the model of Li *et al.* [25] and propose a learning-based feature integration algorithm incorporating visual saliency propagation that decreases the complexity of the method. Hou and Zhang [27] and Guo and Zhang [28] propose approaches based on the spectrum of the images: the Spectral Residual for Hou and Zhang and the Phase spectrum of Quaternion Fourier Transform for Guo and Zhang. In the Guo approach, the object in the spectrum domain is identified and some frequencies in the background are suppressed. A bitrate saving between 32.6 and 38% compared with the traditional H.264/AVC is reported. These methods have the advantage to be less computationally intensive but are also less linked to the HVS. This approach does not work when the salient object is too important because only the boundaries will be detected and the background is more textured than the salient objects. In [29], Chen *et al.* notice that temporal prediction does not work well for video sequences with nonlinear motion and global illumination change between the frames. They propose a new algorithm for dynamic texture extrapolation using H.264/AVC encoding and decoding system. They use as virtual reference frames some synthetized frames that are built with a dynamic texture synthesis. Their evaluation was for a range of QP = {22–37} and IPPP coding. The idea here is to use dynamic texture synthesis to improve some parts of the video sequences. The perceptual results are improved for the entire video sequences. In [30], Bosch integrates several spatial texture tools into a texture based video coding scheme by testing different texture techniques and segmentation strategies to detect texture regions in video sequences. These textures are analyzed using temporal motion techniques and are labeled as skipped areas that are not encoded. After the decoding process, frame reconstruction is performed by inserting the skipped texture areas into the decoded frames. Some side information, such as texture, masks motion parameters which ensure that temporal consistency of the decoder is sent with the modified video sequence. In terms of data rate savings, it is shown that a combination of the gray level co-occurrence matrix to describe the textures and a K-means algorithm to classify them performed the best. On all the sequences tested, the average shows that when the quantization parameter is larger than 36, the side information becomes an overcost. In [31], the coding efficiency of the texture based approaches relative to fast motion objects has been improved. A texture analyzer and a motion analyzer have also been tested. These methods were incorporated into a conventional video coder, e.g. H.264/AVC, where the regions modeled by both the texture analyzer and the motion analyzer were not coded in the usual manner as texture and motion model parameters were sent to the decoder as side information. Both schemes are strongly influenced by the HVS, and described as a set of subjective experiments to determine the acceptability of these methods in terms of visual quality. During the experiment, two sequences are compared, one with a traditional coding and one with their approach. The subjects had no limit on making their decisions and had three options: the first sequence is better in terms of perceptual quality than the second one, the second is better than the first one, and there is no difference between the two sequences. At a quantization parameter equal to 44, a bitrate savings around −20% for the texture based approach and around 3% for the motion based approach are reached. The second approach gives better results because more skip blocks are identified but, in terms of quality their subjective evaluation, it shows that 49% prefer H.264/AVC, only 14% prefer motion based method, and 37% see no difference. Their approach based on skip block is working only on B frame. The I and P frames, which open and close the GOP, are not modified.

Seam carving has been applied to image/video compression using different approaches.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

As existing spatial scalable codecs [32] only support fixed down-sampled resolutions and are not content-aware, Anh *et al.* propose in [5] a content-aware multi-size image compression based on seam carving. Seam carving is used for content-aware reduction until reaching the region of interest (ROI). However, the reduced image and all the information of the seams (position and texture) are encoded, leading to an important bitrate overhead. Moreover, severe block artifacts occur on the boundaries of the ROI and non-ROI regions. In [33], Deng *et al.* solve this problem by combining the advantages of seam carving and wavelet-based coding. A novel content-based spatially scalable compression scheme is thus obtained.

To address the problem of overhead information, a seam can be simplified by a straight line. This approach named selective "data pruning" has been used by Võ *et al.* in [34] to spatially reduce the frames and encode them with H.264/AVC. In this approach, the seam is strongly constrained and cannot easily avoid salient objects, which may lead to visual distortions or to a low spatial reduction.

Tanaka *et al.* [6] introduce a compromise between selective data pruning and seam carving. The seam positions are encoded using piecewise vertical or horizontal straight lines, referred to as "pillars". To define the pillar length, a top-down approach is applied on a modified cumulative energy function. This function is a combination of the forward energy and the seams bitrate using a Lagrangian multiplier. The process stops when the seam defined by this function has a cumulative energy superior to a threshold.

The method proposed in [7] is improved compared with [6] and limit artifacts created during interpolation. Artifacts appear during interpolation when seams are in texture areas, around/near objects or too close to each other. Instead of changing the interpolation method, they propose to change the seams path. A bottom-up approach is used instead of a top-down one to define the length of an optimal pillar and to update the Lagrangian multiplier. Therefore, seams can pass through salient regions of the image. In [35], a piecewise linear approximation is proposed to find the optimal seam. The novelty is that the pieces of seams can have different directions and lengths. In [36], the work of Tanaka *et al.* is extended by approximating each seam with piecewise functions after a rate-dependent optimization. The rest of the image is encoded with a set partitioning in hierarchical trees (SPIHT)-based wavelet coding scheme.

In [37], the authors apply the method used in [6] for video reduction based on the graph-cut approach from [13]. The same seam is deleted for the current GOP and an 8-connexity is allowed from the previous seam to the next GOP in order to avoid artifacts at the transition between GOPs. To compute the seam for the current GOP, all its frames are used but the intra-frame is given more importance.

The temporal aspect is further considered in [38]. The authors propose a trade-off between seam carving and selective data pruning called generalized selective data pruning (GenSDP). GenSDP considers both the retargeted image quality and the bitrate for side information, and a suitable compromise must be considered between these two extreme cases. GenSDP significantly reduces the required bitrates for seam path information compared with the approach based on the original seam carving.

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. To solve this problem, only little information should be transmitted and an approach that encodes seams independently is thus not optimal. In [8, 9], the idea is to transmit just enough information to reposition the salient objects. We observe that seams are mostly concentrated between the salient objects. By defining groups of seams between the salient objects and only transmitting their positions, seam shapes are well approximated and correctly positioned between the salient objects. This approach was validated in [8]. In [9], a new energy map with a better temporal component has been proposed, along with a better combination of the saliency and gradient maps. This approach allows a better preservation of the salient objects. Groups of seams are defined with a k-median algorithm in a manner that each cluster is modeled independently and with more flexibility. All these contributions have improved the coding performance and led to better visual results.

In most of the previous approaches, the texture of seams cannot be transmitted in order to avoid an overhead. Thus, seam synthesis is needed at the decoder. A quick overview of seam synthesis is presented here. Linear interpolation, one of the simplest methods to synthesize missing areas, is used in [3, 13]. It performs well as long as seams do not cross textured areas nor get too close to each other. In [34], Võ *et al.* propose a multi-frame interpolation using neighboring frames. However, these interpolation techniques tend to fail when the areas to be recovered are too large. Domingues *et al.* propose in [10] to use the inpainting approach from Bertalmio *et al.* [4]. This inpainting is quite efficient for rebuilding structures, but often fails to reconstruct textured areas. Seam synthesis remains a challenging issue; however, it is not the focus in our seam carving video coding system.

III. PROPOSED SEAM CARVING APPROACH FOR SEMANTIC CODING

In this section, after reviewing the general approach of semantic coding by seam carving, we present the main contributions of this paper: (1) an algorithm that automatically and adaptively separates the sequence into GOPs depending on the content, (2) a spatio-temporal seam clustering method based on spatial and temporal distances, (3) an isolated seam discarding technique, improving the seam encoding, (4) a new seam modeling, avoiding geometric distortion and resulting in a better control of the seam shapes at the decoder without the need of a saliency map, and (5) a new encoder that reduces the global bitrate.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

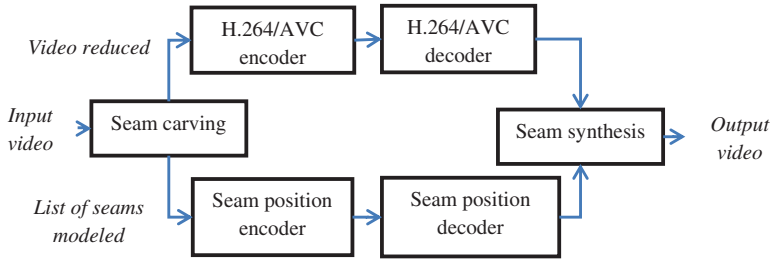


Fig. 1. Architecture of the proposed semantic video coding using seam carving.

A) General approach

The proposed seam carving approach for semantic coding builds upon [8, 9]. Seam carving is used to reduce the spatial dimensions of the video sequence as much as possible, while still preserving the salient objects. Then, the reduced video is encoded with a traditional encoder such as H.264/AVC [1]. In parallel, the seams are modeled and encoded with our proposed scheme. After transmission, the video sequence is reconstructed at the decoder side, in order to recover the original dimensions and to preserve the scene geometry. Finally, a technique such as image inpainting can optionally be used to synthesize the missing texture. Figure 1 shows the global approach.

Figure 2 details the seam carving module in Fig. 1. During the seam carving process, lists of seams are computed for each frame of the sequence and used to identify the variation of the number of seams in time. The sequence is then adaptively subdivided into GOPs. This is executed during the content-aware GOP segmentation. Then, for each GOP, a spatio-temporal seam clustering is performed in order to model the groups of seams and to discard isolated seams. In this way, a list of modeled seams is obtained for each frame of the sequence; which are all subsequently suppressed from the original video sequence to obtain a reduced video.

The seam computation module depicted at the top of Fig. 2 is further detailed in Fig. 3. An energy function is defined for each frame from the saliency model in [39]. This model identifies the salient objects by finding the rarity on different maps. The most pertinent maps are combined together to obtain a unique saliency map. The model uses static (L,a,b) and dynamic (motion amplitude and direction) components in order to identify salient areas for static and moving scenes. The saliency map is then combined with a gradient map that highlights the outlines of the objects. On this energy map, a spatial median filter is then applied to remove the noise, followed by a dilation filter to preserve salient objects and their neighborhood. The forward cumulative energy map is then computed to define the seams to suppress.

In parallel, the energy map is binarized to obtain a control map. The binarization threshold is defined as $T = 2 \cdot mean\ (Saliency)$ as proposed in [40]. The control map is used to decide when the process of seam suppression is stopped. More precisely, seam carving is iterated until reaching an object in the control map. Thus, a list of seams is obtained for each frame. The process is successively applied vertically and horizontally.

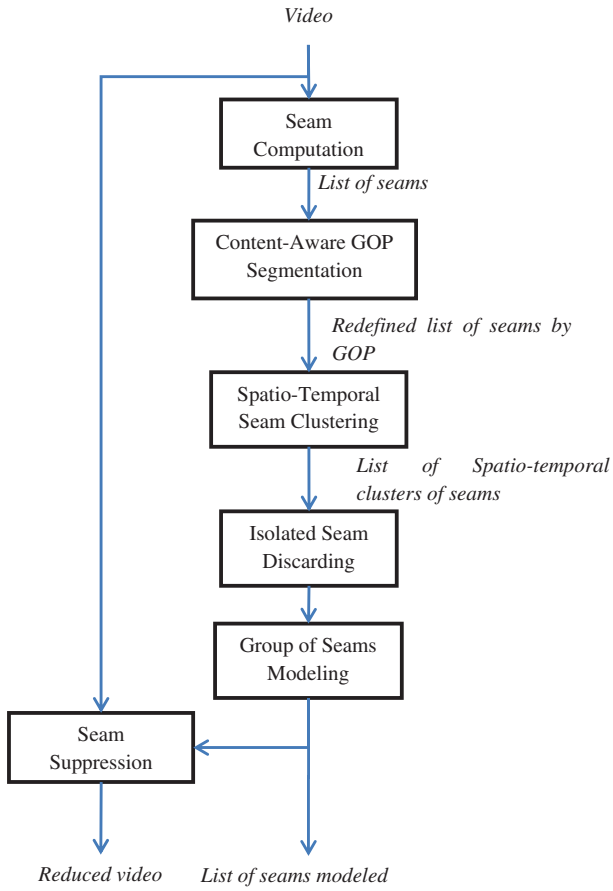


Fig. 2. Overall scheme of the proposed seam carving process (corresponding to the module "seam carving" in Fig. 1).

B) Content-aware adaptive GOP segmentation

To use seam carving in a video compression application, the sequence is divided into GOPs. In a GOP, in order to avoid padding, it is preferable that all the frames have the same spatial dimension, defined as a multiple of 16 pixels (corresponding to the size of a macro block in the subsequent video coding scheme). In our previous work [8, 9], the length of the GOP was predefined and the number of seams suppressed for each GOP was linked with the frame having the largest salient objects. That led however to a suboptimal reduction of the sequence dimensions.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

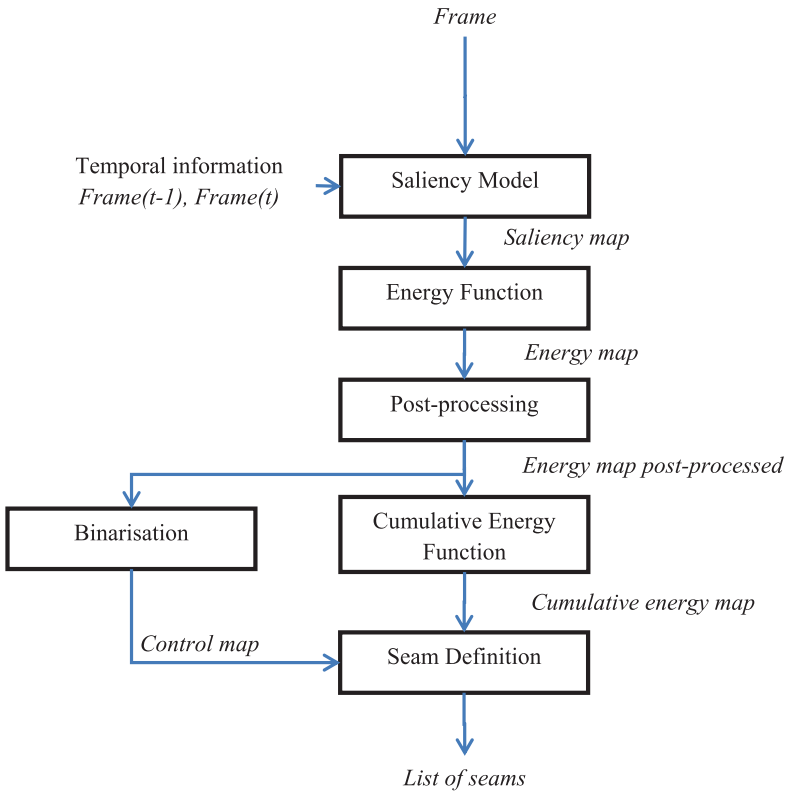


Fig. 3. Seam computation process to obtain from a frame a list of seams (corresponding to the module "seam computation" in Fig. 2).

In this paper, we propose to adaptively split the sequence as a function of the number of seams that can be suppressed. This way, the GOP can be further spatially reduced without damaging the salient objects and the quantity of information to transmit is lowered. Three main cases can be identified: (1) an object of interest is appearing into the video and the number of seams that can be suppressed is decreasing; (2) an object is disappearing from the video and the number of seams that can be suppressed is increasing; and (3) a constant number of still or moving salient objects are present in the scene and the number of seams that can be suppressed remains constant.

To implement these cases, rupture detection is applied on the number of vertical and horizontal seams to identify important changes. More specifically, let us define $Nb\_VertSeam(t)$, the number of vertical seams for the frame at time t. The first step is to apply a median filter on $Nb\_VertSeam(t)$ to reduce the local variations due to noise or salient object detection errors. Then, rupture detection is applied to define the segments. Formally, for a new segment starting at the frame t_g, the median value at the current frame t_{cur} is defined as:

$$Val\_Vert_{med}(t_{cur}) = Median\left(\{Nb\_VertSeam(t)\}_{t=t_g}^{t_{cur}}\right). \quad (10)$$

If the condition

$$|Val\_Vert_{med}(t_{cur}) - Nb\_VertSeam(t_{cur})| < th_{Gop}, \quad (11)$$

holds, the current frame is included into the segment, otherwise a new segment is created. For this purpose, a *GOP_Threshold* th_{GOP} is used to define when a new segment is created.

Likewise, the same process is applied to the horizontal seams $Nb\_HorizSeam(t)$.

The number of removable seams for a segment is defined as the median of the number of seams over this segment. This way, the reduction process is improved while preserving the salient objects, especially for the monotonic segments. Finally, the number of removable seams in each segment is rounded to the nearest multiple of 16.

The combination of these two analyses gives an adaptive cut of the video with different dimensions. More precisely, if a rupture appears horizontally or vertically, a new GOP is created.

Figure 4 summarizes this approach. For the horizontal and the vertical seams, the rupture detection is applied, as illustrated in Figs 4(a) and 4(c). Then, for each segment, a number of seams is calculated and rounded to a multiple of 16, as shown in Figs 4(b) and 4(d). By combining the vertical and the horizontal analyses, the GOPs are defined with their dimensions as in Fig. 4(e).

C) Spatio-temporal seam clustering

After having defined the GOP and the corresponding number of seams that can be suppressed, spatio-temporal clustering is applied to identify groups of seams. These groups of seams will be used to model and encode the most important

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

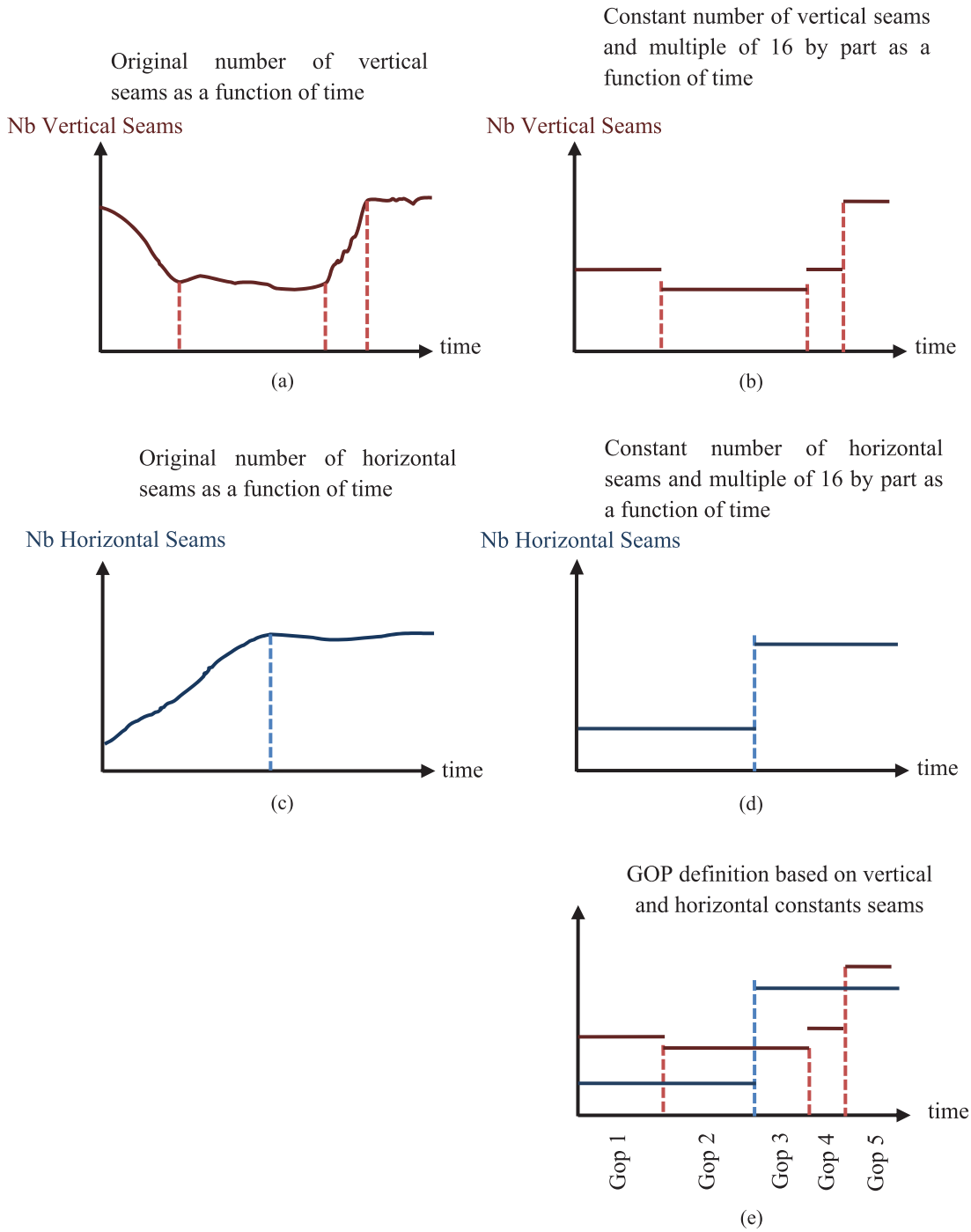


Fig. 4. Group of picture (GOP) definition based on the content. Horizontal continuous lines represent the number of seams. Vertical dotted lines represent the GOP segmentation. Blue, resp. red, indicates horizontal, resp. vertical, seams.

seams and to identify isolated ones. Figure 5 highlights the different step to obtain from a redefined list of seams by GOP, a list of spatio-temporal clusters of seams.

As the seam carving is an iterative process, the coordinates of the seams position at iteration k are expressed in function of the reduced image at iteration k-1. In addition, subsequent seams can cross one another. Therefore, in order to unequivocally define them, the seams positions are changed to be stated in the coordinates of the original image and then rearranged by ordering the horizontal, respectively, vertical, coordinates in an increasing order as in [9]. Figure 6 illustrates the seam reordering process with three seams that are represented in blue, orange, and green. In the first schema, before reordering, seams can cross each other. But in the second schema, it is no longer the case after reordering. At the same time, it can be noticed that the new seams go through the same coordinates.

To perform spatio-temporal grouping, the seams are first grouped together spatially and then temporally. The process is explained for vertical seams and can be easily transposed

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

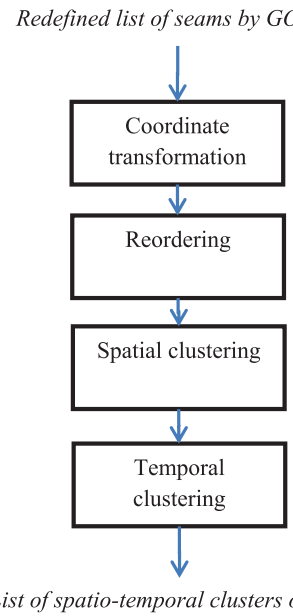


Fig. 5. Spatio-temporal clustering process to obtain a list of spatio-temporal clusters of seams from a list of seams by GOP (corresponding to the module "spatio-temporal seam clustering" in Fig. 2).

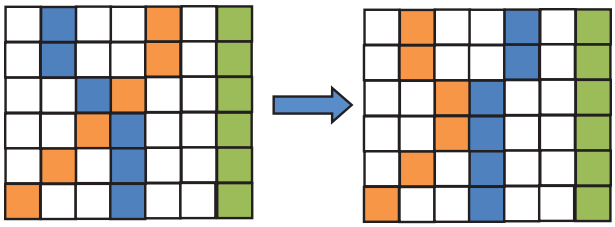


Fig. 6. Seams reordering: on the left, seams before reordering; on the right, seams after reordering.

for horizontal ones. For this purpose, the *Spatial Distance* (*SD*) is defined as:

$$\forall j \in [1, J-1], SD_{(j,t)}(Seam_{(j,t)}, Seam_{(j+1,t)}) = \max_{i=1\ldots N}(|Seam_{(j+1,t)}(i) - Seam_{(j,t)}(i)|), \quad (12)$$

where N is the length of a vertical seam (in the case of a $N \times M$ image), i the pixel position index inside the seam, and J the number of seams suppressed for the frame at t. For a vertical seam, $Seam_{(j,t)}$ is the horizontal coordinate of the jth seam in the frame t and $Seam_{(j+1,t)}$ is the horizontal coordinate of the seam $(j+1)$th in the frame t. This maximum distance has been chosen as it successfully identifies salient objects between seams, contrary to the mean distance or the median distance.

The groups of seams are spatially defined for each frame independently: $\forall j \in [1, J-1], \forall t \left[1, Length\_GOP\right]$, $b = 1$ and

$$\text{if } SD(Seam_{(j,t)}, Seam_{(j+1,t)}) < Sp_{Th} \Rightarrow Seam_{(j+1,t)} \in G_{Seam}(b,t), \quad (13)$$

$$\text{if } SD(Seam_{(j,t)}, Seam_{(j+1,t)}) \geq Sp_{Th} \Rightarrow b = b+1, Seam_{(j+1,t)} \in G_{Seam}(b,t), \quad (14)$$

$G_{Seam}(b,t)$ is the bth group of seams in the frame t and b is initialized at 1 for each new frame. Sp_{Th} is the *Spatial Threshold* and represents the maximal distance between two consecutive seams found in the same group. It has been experimentally set to 12 pixels. Consequently, the maximal number of groups of seams $B(t)$ can be different in each frame.

Then, the groups are temporally linked together using the symmetric difference between the groups of seams at time t and $t+1$. Let $Border\_Seam\_Left_{(b,t)}$ (resp. $Border\_Seam\_Right_{(b,t)}$) the most leftward (*resp. rightward*) seam of the bth group. Define $G_{Seam}(b,t)$ by all the pixels between $Border\_Seam\_Left_{(b,t)}$ and $Border\_Seam\_Right_{(b,t)}$:

$$G_{Seam}(b,t) = \left\{ \begin{array}{c} (x,y) \in N^2; 1 \leq y \leq N \text{ and} \\ Border_{Seam_{Left}(b,t)}(y) \leq x \leq \\ Border_{Seam_{Right}(b,t)}(y) \end{array} \right\}. \quad (15)$$

The *Symmetric Difference SymDif* between the group $G_{Seam}(b,t)$ and $G_{Seam}(k,t+1)$ is defined as the area of the union of $G_{Seam}(b,t)$ with $G_{Seam}(k,t+1)$ minus the area of the intersection of $G_{Seam}(b,t)$ with $G_{Seam}(k,t+1)$:

$$\begin{aligned} &\forall b \in [1, B(t)], \forall k \in [1, B(t+1)], \\ &\quad Area(SymDif(G_{Seam}(b,t), G_{Seam}(k,t+1))) \\ &= Area(G_{Seam}(b,t) \Delta G_{Seam}(k,t+1)) \\ &= Area \frac{\{(G_{Seam}(b,t) \cup G_{Seam}(k,t+1))}{(G_{Seam}(b,t) \cap G_{Seam}(k,t+1))\}}, \end{aligned} \quad (16)$$

where $B(t)$ is the number of groups of seams for the frame t.

Then, temporal regrouping is applied using *SymDif*. More precisely, the groups of seams in frame t, respectively, in frame $t+1$, with the smallest distance, are linked together.

l is defined as the label of the group of seams in the complete GOP and is initialized at 1 at the beginning of the GOP. If *SymDif* is inferior to the *Temporal_Threshold* T_{Th}, these two groups of seams will share the same label l. Otherwise, a new class is created and $l = l+1$.

Figure 7 illustrates the spatio-temporal clustering of two consecutive frames. The seams are represented with different colors. The spatial clustering is first applied on the first frame and leads to two groups of seams. The first group $G_{Seam}(1,1)$ contains three seams (red, dark blue, and green) and takes the label $l = 1$. The second group $G_{Seam}(2,1)$ contains two seams (purple, blue) and takes the label $l = 2$.

The $Area\_G_{Seam}(b,t)$ is illustrated by the orange outlines. For the second frame, three groups of seams are identified. The first one contains the red and dark blue seams, the second one the green seam, and the third one the purple and blue seams.

Then, the temporal clustering is carried out using $Area\_G_{Seam}(b,t)$ and *SymDif*, and each group of seams in frame 2 is compared with the group of seams in frame 1. $G_{Seam}(1,2)$ takes the same label than $G_{Seam}(1,1)$ and $G_{Seam}(3,2)$ takes the same label than $G_{Seam}(2,1)$. $G_{Seam}(2,2)$ is considered as a new group of seams and take the label $l = 3$.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

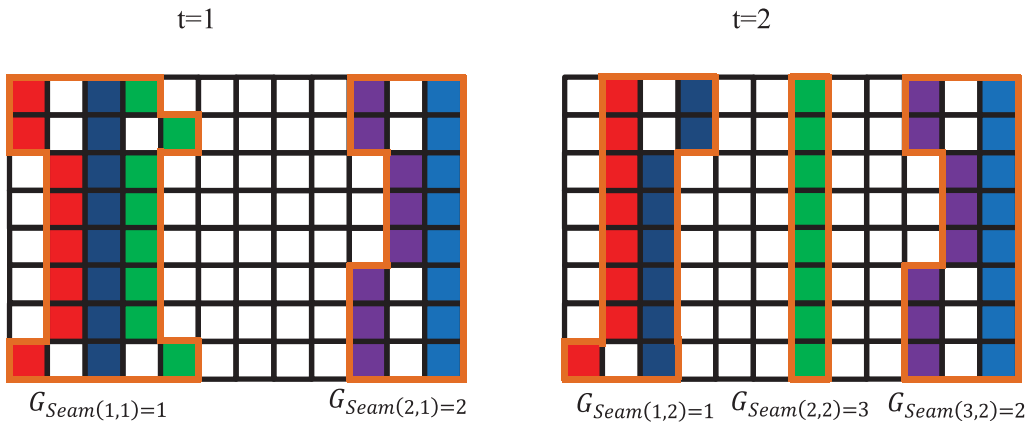


Fig. 7. Illustration of the spatio-temporal seam clustering.

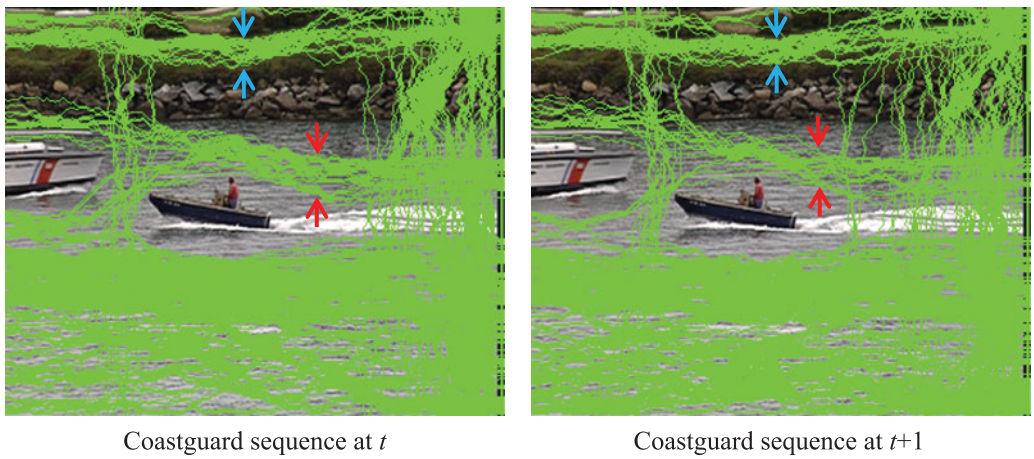


Fig. 8. Illustration of seams in two consecutive frames of the Coastguard sequence. Arrows show examples of group of seams.

In Fig. 8, seams are illustrated in green in two consecutive frames of the coastguard sequence. Blue and red arrows show examples of group of seams that are clustered together.

D) Isolated seam discarding

During the previous step, groups of seams are defined for the whole GOP. Moreover, for each group, we know the number of seams associated with each group and the number of frames where it is present. As each group will be modeled and encoded, it is important to encode only meaningful groups in order to avoid a high overhead.

For this purpose, all the small groups, with a percentage of the total number of seams inferior to a threshold *Outliers_Number_Threshold*, are deleted. In addition, groups of seams have to be present in a sufficient number of consecutive frames to be temporally consistent. Therefore, groups that are only present in a few frames are also discarded, using a threshold *Threshold_Outliers_Length*. In Fig. 9, examples of isolated seam are indicated by red arrows.

Finally, since the number of seams in each frame should be constant throughout the GOP, discarded seams have to be reallocated to other groups of seams in the same frame. This will also stabilize the temporal variations of the groups of seams. For this purpose, in each frame, the group having the strongest temporal variation of its number of seams will receive the same number of seams as those subtracted by the isolated seam discarding.

E) Group of seams modeling

After being defined, the groups of seams have to be modeled before being encoded. As we assume that salient regions do not contain seams, the modeling should not modify the outside of the groups of seams while creating the maximum of diversity within the groups of seams.

In the proposed approach, border seams are approximated in a different way than seams within the group. Figure 10 illustrates two groups of seams, with the border seams in red and the inner seams in blue. The first group has three inner seams between the two border seams and the second group has one inner seam between the two border seams.

Figure 11 summarizes the process of modeling both the border seams (red seams in Fig. 10) and the inner

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

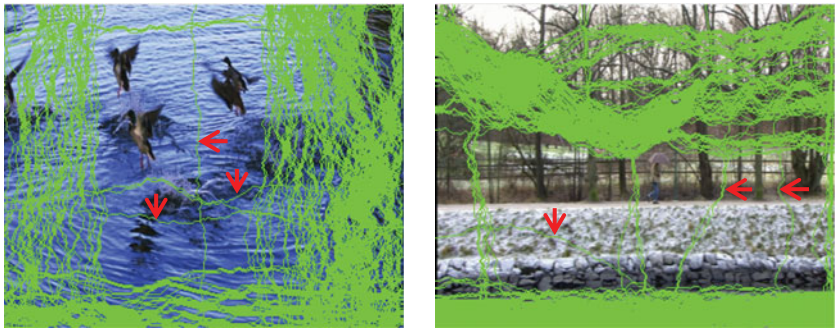


Fig. 9. Example of isolated seams for a frame of the Ducks sequence (left) and the Parkrun sequence (right). Isolated seams are shown by the red arrows.

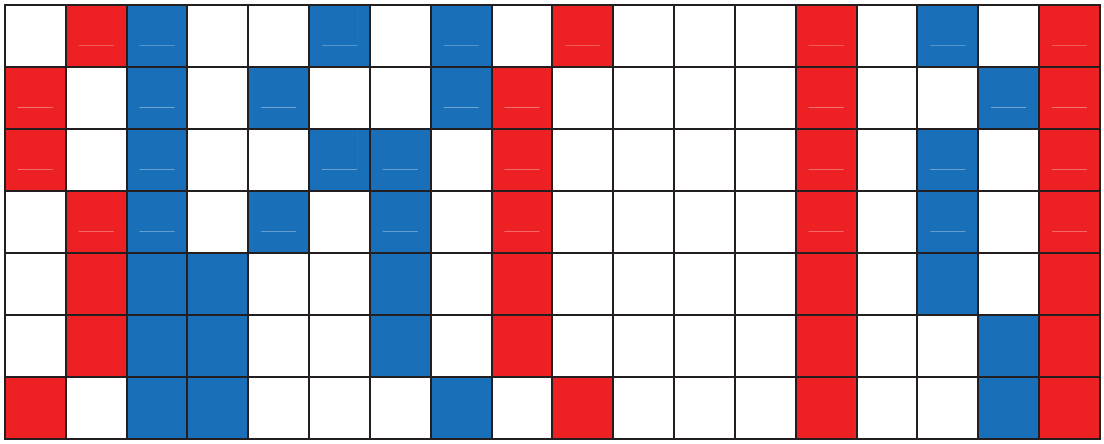


Fig. 10. Illustration of two groups of seams with in red the border seams and in blue the inner seams within the group of seams.

seams (blue seams in Fig. 10). Each border seam will be approximated by a polynomial that needs less information to be encoded. As the outside of the group of seams may include salient regions, the polynomial seams have to be totally included inside the group. In other words, the whole approximated left polynomial seam should be at the right of the left border seam. Conversely, the whole approximated right polynomial seam should be at the left of the right border seam. The first step of the group of seam modeling is the combination between the polynomial seam and the border seam to create a combined seam, with the following operations:

$$\begin{aligned} \forall i \in [1, N] & Combined\_seam\_Left(i) \\ & = \max(Polynomial\_seam\_Left(i), \\ & \quad Border\_seam\_Left(i)), \end{aligned} \tag{17}$$

$$\begin{aligned} \forall i \in [1, N] & Combined\_seam\_Right(i) \\ & = \min(Polynomial\_seam\_Right(i), \\ & \quad Border\_seam\_Right(i)), \end{aligned} \tag{18}$$

where N is the length of the seam, *Polynomial_seam_Left* is the left polynomial seam, and *Border_seam_Left* is the seam border at the left of the group of seam. *Polynomial_seam_Right* is the right polynomial seam and *Border_seam_Right* is the seam border at the right of the group of seams. For the first iteration, the combined seam is initialized with the border seam.

The combined seam obtained is then approximated by a polynomial seam during the border seams approximation step. Next, we check if the polynomial seam is totally included within the group of seams:

$$\begin{aligned} \text{if} \sum_{i=1}^{N} & ((Polynomial\_seam\_Left(i) - Border\_seam\_Left(i)) > 0) \\ & = N \Rightarrow included, \end{aligned} \tag{19}$$

$$\begin{aligned} \text{if} \sum_{i=1}^{N} & ((Border\_seam\_Right(i) - Polynomial\_seam\_Right(i)) > 0) \\ & = N \Rightarrow included. \end{aligned} \tag{20}$$

If the polynomial seam is totally included, its shape is encoded, otherwise the process of combination and approximation is reiterated.

As for the inner seams in the group of seams (blue seams in Fig. 10), they are modeled by a uniform distribution between the two approximated polynomial border seams. If the distance between two modeled border seams is locally narrower than the number of seams in the group, it is obviously not possible for all inner seams to pass between the two borders. In this case, some inner seams are allowed to locally go over the borders.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

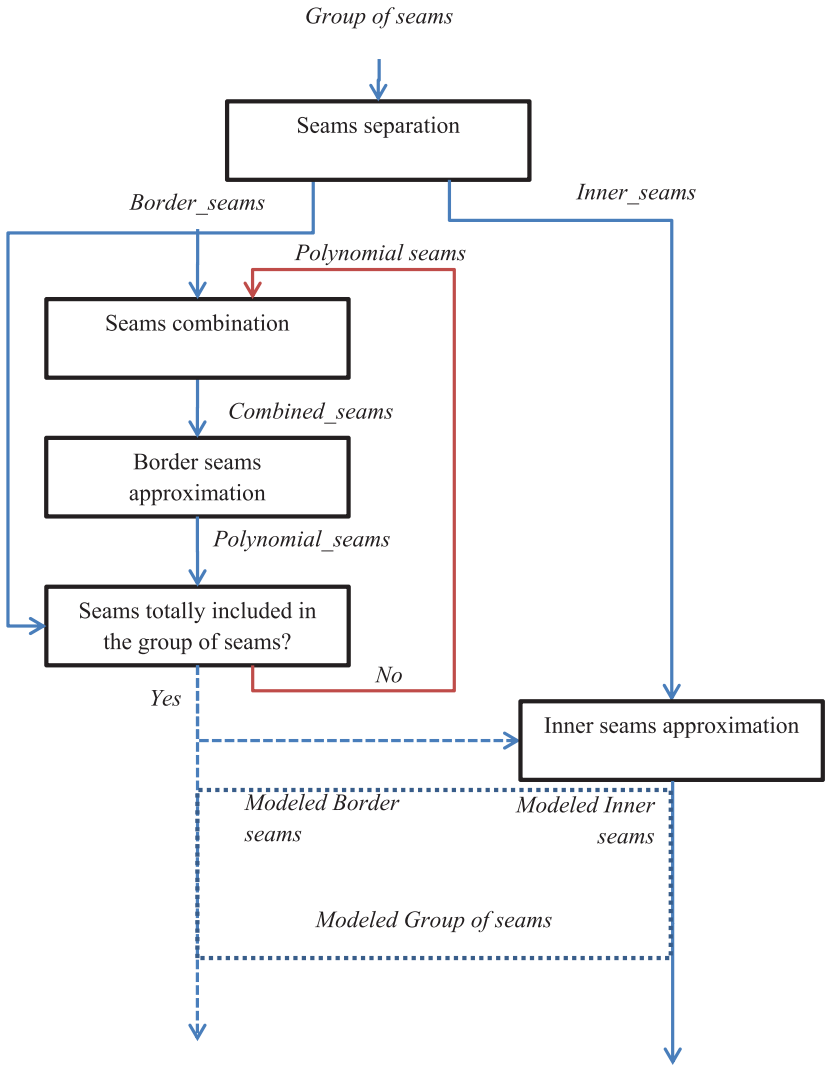


Fig. 11. Group of seams modeling (corresponding to the module "group of seams modeling" in Fig. 2).

Fig. 12. Illustration of the original seams (left) and the modeled seams (right) for the Parkjoy sequence.

Figure 12 illustrates the result after having applied the seam clustering, the isolated seam discarding, and the group of seams modeling. In this example, six vertical groups of seams are identified and an isolated vertical seam is discarded. The shape borders of groups of seams are approximated by polynomial seams. In this way, the quantity of

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

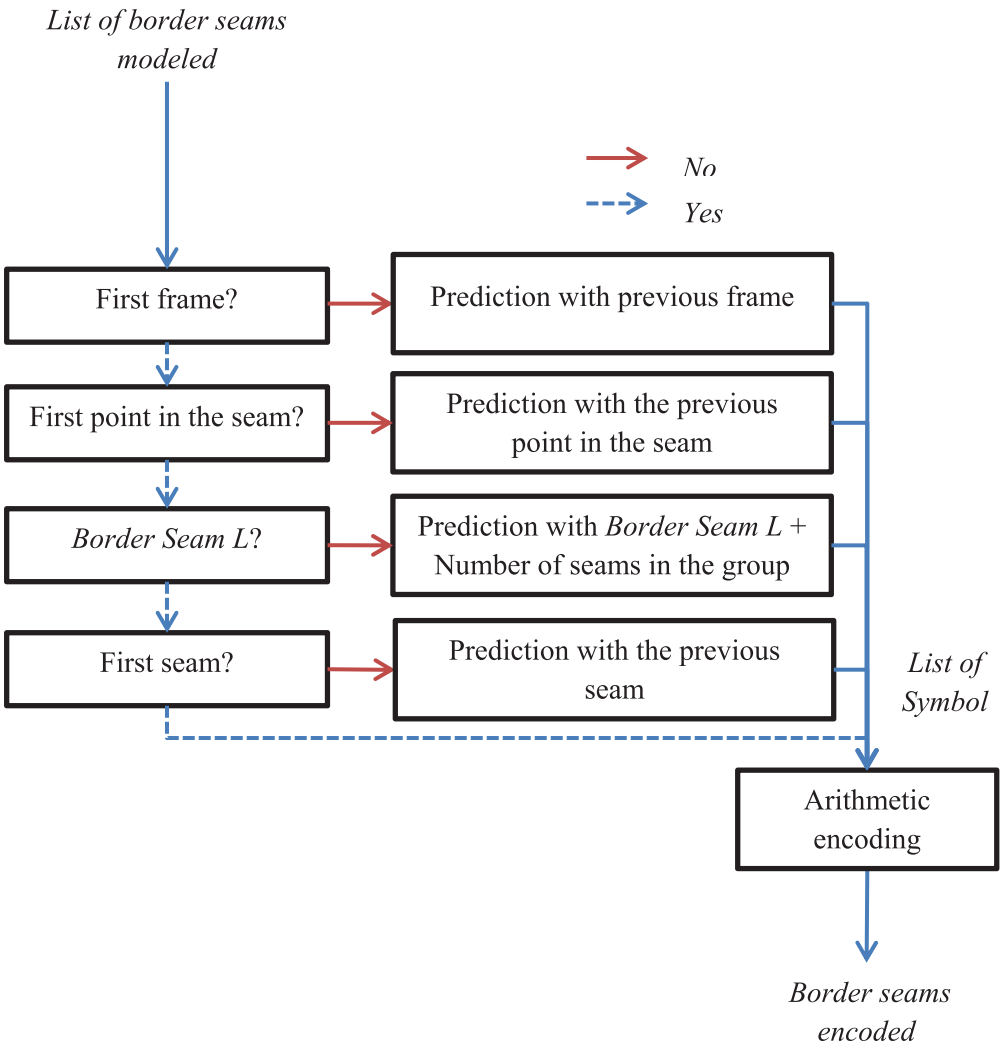


Fig. 13. Predictive model for the group of border seams and encoding (corresponding to the module "group of seams modeling" in Fig. 1).

information to transmit is reduced and the texture outside the groups of seams is preserved. Horizontally, only one group of seams is identified.

F) Seam encoding

To rebuild the seams at the decoder side, some information has to be transmitted. More specifically, for each group of seams in each frame, the numbers of seams and border seams polynomial models are sent. As previously detailed, the border seams are polynomials of degree m, with $m = 3$. A vertical (resp. horizontal) border seam of length N (resp. M), is totally defined by its horizontal position (resp. vertical) for each vertical position y, $y \in [1 \ldots N]$. The Matlab function polyfit and polyval are used for this purpose. Due to the degree 3 of the polynomial, each border seam can be represented with four horizontal positions at $[1, N/m, 2N/m, \text{and } 3N/m]$.

A list of symbol containing the horizontal (resp. vertical) position in four points for all the vertical (resp. horizontal) border seams in the GOP is obtained and used in the predictive scheme. Figure 13 illustrates with more details the predictive encoding models of the border seams models. For the whole process, the prediction is a subtraction and the residuals are encoded in a lossless way.

The first horizontal (resp. vertical) position of the first vertical (resp. horizontal) border seam in the first frame of the GOP is simply represented by its horizontal (resp. vertical) position. Then, the next horizontal (resp. vertical) position of this seam is predicted from the previous one. For the next seams, if it is a *Border_seam_Right*, the first horizontal (resp. vertical) position is predicted from the first horizontal (resp. vertical) position of the previous seam plus the number of seams inside the group of seams. Otherwise, it is a *Border_seam_Left* and the prediction of the first horizontal (resp. vertical) position is just the difference with the first horizontal (resp. vertical) position of the previous seam, *Border_seam_Right*. After having performed the prediction of the first frame of the GOP, the prediction is done temporally for the other frames.

Finally, the list of predicted residuals is encoded by using an arithmetic coder [41].

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

IV. PERFORMANCE ASSESSMENT RESULTS

A) Evaluation methodology

To perform the evaluation, four test sequences in 352 × 288 pixels (CIF) format have been chosen: Coastguard, Ducks, Parkrun, and Parkjoy. These sequences show different spatial and temporal characteristics representative of a broad range of application scenarios.

Traditional image quality metrics, such as peak signal to noise ratio (PSNR) or structural SIMilarity (SSIM) [42], compare corresponding pixels or blocks in the reference and processed images. However, they provide a global quality measure rather than an object-oriented measure. In addition, they commonly fail in the presence of geometric deformations, object displacements, and background synthesis. In our context, a quality metric has to compare the salient regions, and to match the pixel positions in the original and processed images. The full reference image retargeting metric proposed by Azuma *et al.* in [43] solves the problem of matching by using scale-invariant feature transform (SIFT) [44] and SSIM. With the same idea, Liu *et al.* presented in [45] an objective metric simulating the HVS based on global geometric structures and local pixel correspondence based on SIFT. However, these two metrics are not designed to evaluate compression artifacts, do not take into account geometric deformations and are influenced by a synthesized background. In [46], a metric, referred to as SSIM_SIFT, especially designed for the problem at hand has been presented. More precisely, it uses windows around SIFT points to measure by SSIM the compression artifacts due to encoders like H.264/AVC and return also the quantity of geometric deformation introduced by the approximation of the seam carving encoding. Since these windows are totally included in the salient parts, the result is not degraded by the synthesized parts in the background. This metric has been validated with subjective tests [46] and obtained a Spearman and a Pearson correlation of 0.86. As our approach creates no geometric artifacts, evaluation has been done with another simple metric, referred to as SSIM_MASK. It consists in merely computing SSIM on a predefined region of the image only.

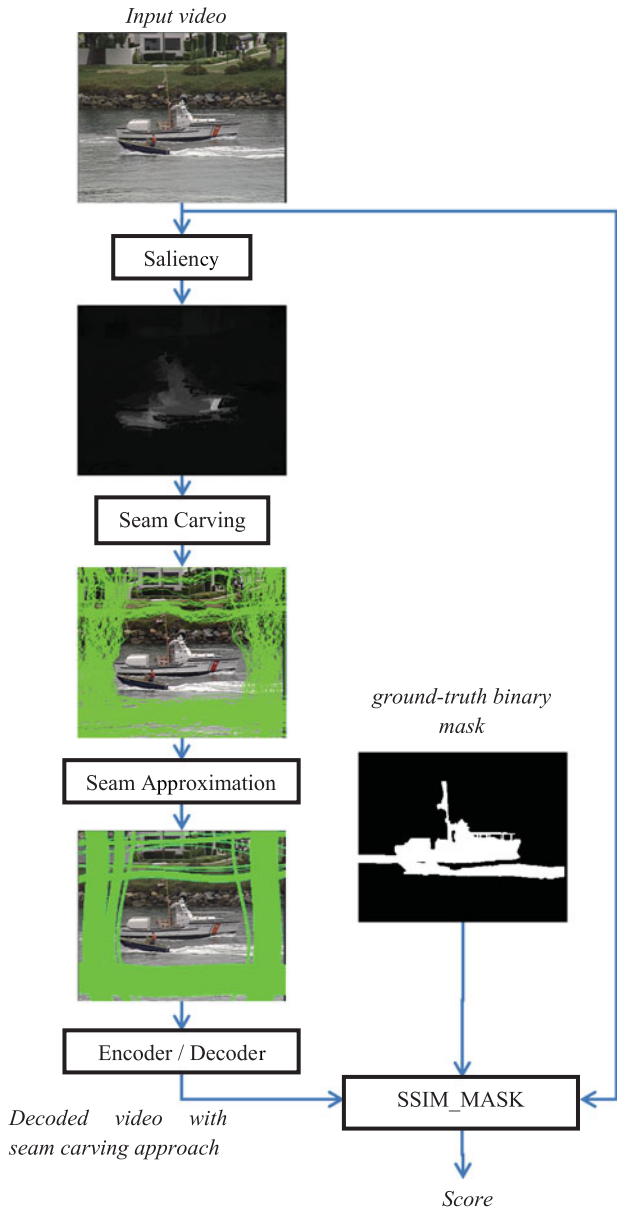


Fig. 14. Evaluation protocol. Comparison of the decoded video with seam carving approach with the original video. SSIM_MASK is computed on the ground-truth binary mask.

B) Evaluation protocol

When using the proposed seam carving scheme, perceived visual quality can be assessed in two steps. The first one is the evaluation of the salient object (object deformation, displacement, and quality) and the second one is the interpretability of the background. To the best of our knowledge, no techniques can reliably evaluate the second step. For this reason, the evaluation is focused hereafter on the salient object only.

Figure 14 illustrates the proposed evaluation protocol for the salient object using SSIM_MASK. After computing the saliency map as proposed in [39], seam carving is applied as described in Section II.B. Then, the proposed seam modeling (Section III) is applied to approximate the seams. The white parts of the saliency maps illustrate the important regions and the black parts the less important ones. The seams are illustrated in green. The reduced-size sequence is encoded, along with the modeled seams. To evaluate the results, the reduced sequence is expanded at the decoder side and the final results are compared with the reference using the SSIM_MASK.

The SSIM_MASK metric use a ground-truth binary segmentation mask to identify the salient object. In our experiments, we have used the manual binary masks from [39]. It should be underlined that these binary masks are only used in the quality metrics and are not involved in the encoding process. With this metric, a SSIM score is computed only on the salient object.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

Table 1. Value and influence of various parameters.

Parameter	Value	Influence
Saliency map binarization threshold (Section III.A)	$T = 2 \cdot mean(saliency)$	Defines what is salient in the video and when the seam carving should stop
Parameter for the content-aware group of picture (GOP) segmentation (Section III.B)	$GOP\_Threshold$, $th_{GOP} = 15$	Defines when a new GOP is created
Parameters for seam modeling (Sections III.C–III.E)	$Spatial\_Threshold$, $Sp_{Th} = 12$	Define the groups of seams
	$Temporal\_Threshold$, $T_{Th} = 100$	
	$Outliers\_Number\_Threshold = 1$	Define the outliers
	$Threshold\_Outliers\_Length = Length\_GOP/2$	

C) Parameters

In our approach, several parameters have been defined. Their theoretical optimization is quite complex due to the fact that the proposed approach depends on the video content. Moreover, the evaluation is difficult due to the lack of appropriate metrics for some improvements like the seam modeling. The influence of each parameter is described in Table 1. Extensive experimental tests have been done on different sequences with different ranges of values in order to heuristically define the optimal value of the parameters.

D) Ratio of spatial resizing

As our approach reduces the spatial dimensions of the sequences depending on the content, the ratio of spatial resizing varies accordingly. The performance is directly linked to two parameters: the binarization coefficient applied on the saliency maps to obtain the control maps and the content-aware GOP segmentation. The ratio of spatial resizing, or in other words the percentage of suppressed pixels, is defined as:

$$R_{Resizing} = 100 * \left(1 - \frac{\text{Spatial dimension of the reduced sequence}}{\text{Spatial dimension of the original sequence}}\right). \quad (21)$$

Figure 15 illustrates the evolution of the seams as a function of time for the Parkjoy sequence. The number of seams that can be suppressed without any constraint is shown in blue, and the proposed approximation in green. A compromise is obtained between a high number of GOPs, giving a better approximation, but more subsequences to encode.

In Table 2, a comparison of the ratio of spatial resizing between an approach without any constraint on the number of suppressible seams, an approach with fixed length GOP = 5 [8, 9] and our approach is presented. We can see that the proposed approach achieves a very good ratio of spatial resizing, nearly equal to an approach without any constraint. In contrast, the fixed GOP approach has a much smaller ratio of spatial resizing.

E) Seams approximation

Our approach tries to find a trade-off between the flexibility of seam representation and their coding cost. Figure 16 illustrates some visual results for the different test sequences. In general, in all the sequences, the salient objects obtained using [39] are preserved after modeling. Isolated seams are deleted and reallocated to other groups of seams. Some isolated seams are however kept, when they are consistent in time. Figure 17 illustrates the temporal aspect of the modeling for the Coastguard sequence.

F) Evaluation of the seams modeling

In this experiment, we evaluate the efficiency of our seam modeling.

Different methods have been compared:

- The first one is based on the reduced video after the seam carving *without* modeling. No seam is encoded and this approach illustrates the achievable bitrate saving due to the spatial resizing. In this case, the scene geometry cannot be correctly reconstructed at the decoder.
- The second approach is based on the same reduced video and all seams positions are fully encoded *without* seam modeling. The first horizontal position (in the case of a vertical seam) of each seam is encoded on 10 bits. Then the next coordinate is predicted from the previous one. As the seam has only three possible position (right, straight, and left), it is encoded on 2 bits.
- The third approach is based on the new reduced video after the seam carving *with* modeling. The seams are not encoded and this approach illustrates the achievable bitrate saving due to the spatial resizing when the seams suppressed are grouped and modeled. In this case, the scene geometry cannot be correctly reconstructed at the decoder.
- The fourth approach is based on the new reduced video after the seam carving *with* modeling and the seams are modeled and encoded with our approach. This is the result of our proposed approach.

H.264/AVC [1] is used in full intra coding that is consistent with video surveillance applications. The bitrate of the entire sequence is taken into account during the

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

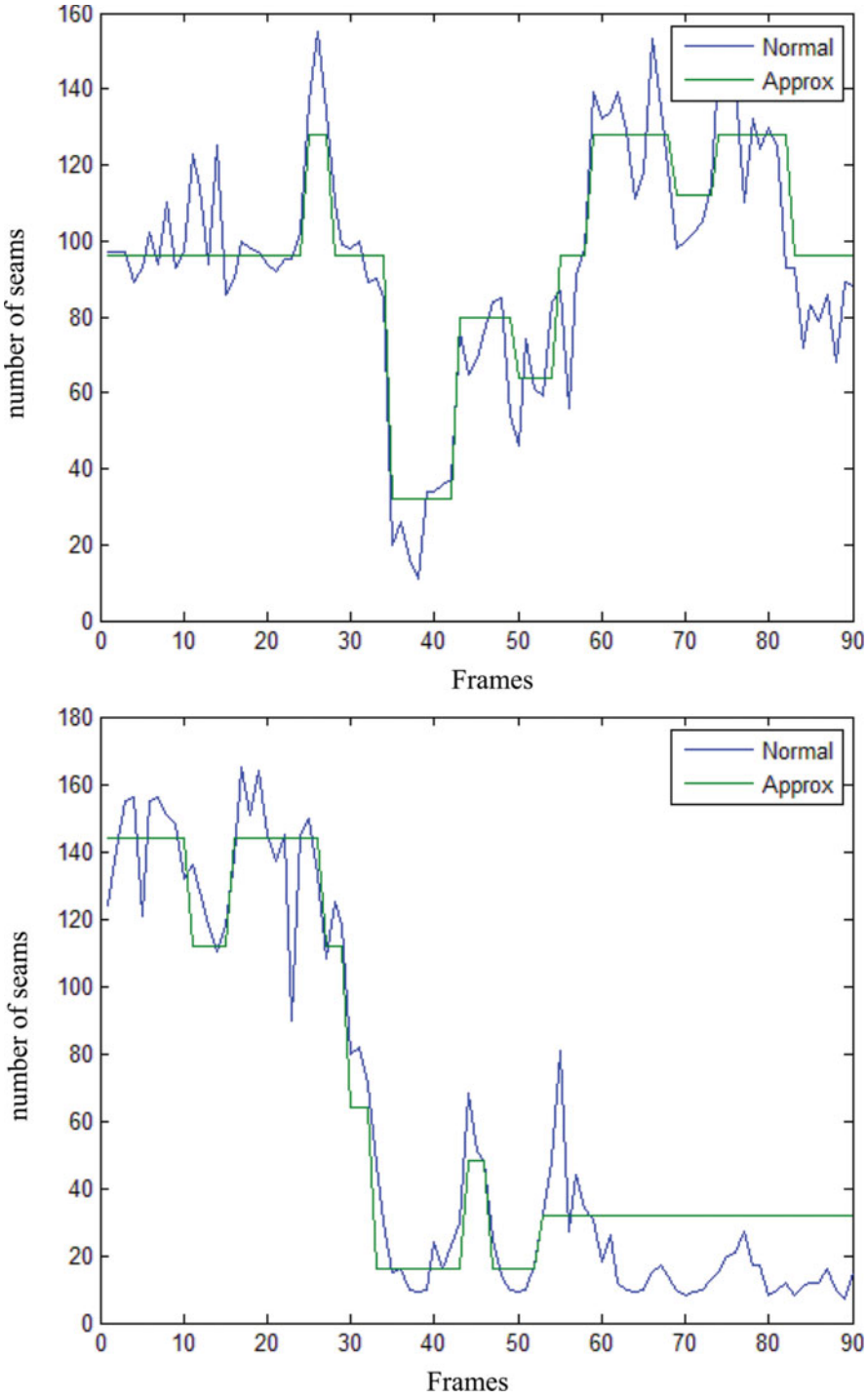


Fig. 15. Approximation of the number of suppressed seams for Parkjoy: evolution of the number of vertical seams (top) and horizontal seams (bottom).

Table 2. Comparison of the ratio of spatial resizing between different approaches.

Sequence	No constraint	Fixed GOP length = 5 ([8, 9])	Proposed approach
Coastguard	53.38	41.38	54.19
Ducks	27.74	13.32	27.09
Parkjoy	43.84	34.60	45.20
Parkrun	33.57	20.65	32.56

comparison. The quantization parameter intra (QPI) varies from 27 to 39 with a step of 3. Table 3 illustrates the percentage of bitrate saved compared with conventional H.264/AVC as a function of QPI.

We can see that the rate of spatial reduction is not directly linked to the bitrate saved following (1) and (3). This is due to the fact that seams can pass in areas that did not need many bits to be encoded and that some high frequencies can be created after the seam carving. The

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

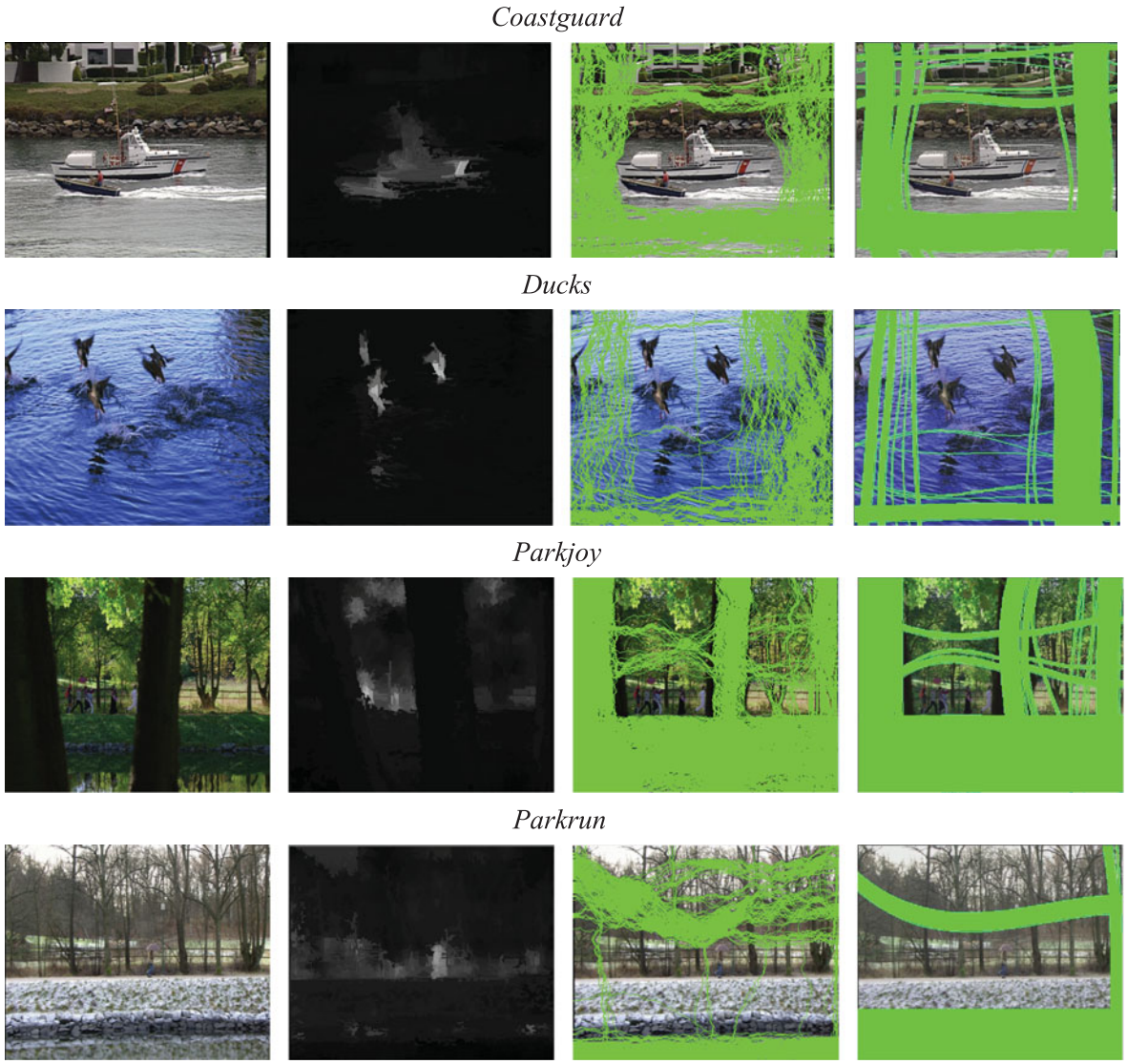


Fig. 16. Visual seam modeling for Coastguard/Ducks/ParkJoy/ParkRun sequences. First column: sample original frame; second column: salient objects [39]; third column: initial seams; and fourth column: seams after the proposed modeling.

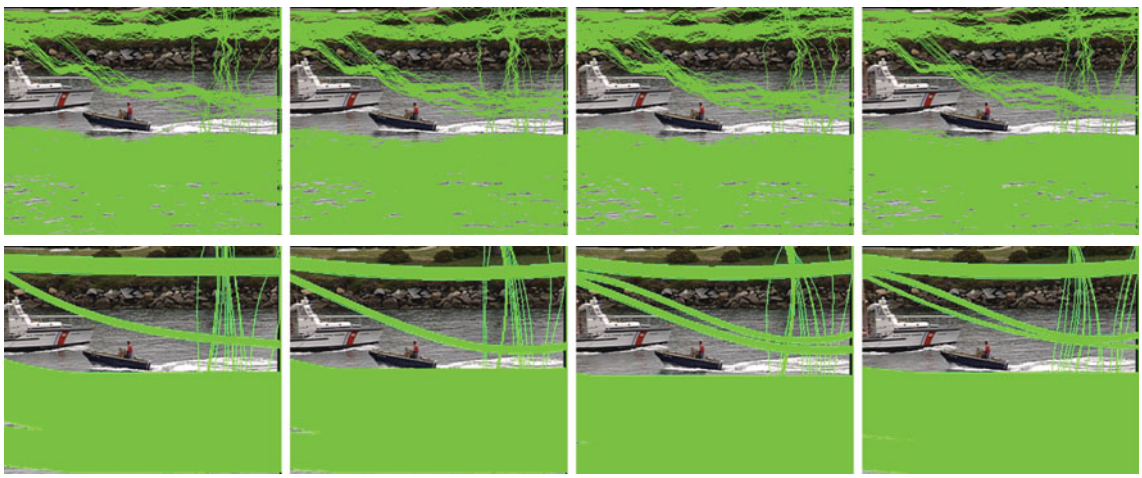


Fig. 17. Visual seam modeling for Coastguard at different times $t = 27, 28, 29, 30$. On first line, initial seams, on second line, the seams after the proposed modeling.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

Table 3. Percentage of bitrate saved compared with H.264/AVC as a function of quantization parameter intra (QPI) (positive value means bitrate is decreased, negative value means bitrate is increased): (1) reduced video after the seam carving without seam coding, (2) reduced video after the seam carving with seam position encoded without modeling, (3) reduced video after the seam carving modeled without seam coding, and (4) our proposed approach with reduced video after the seam carving modeled and the seams modeled and encoded.

Sequence	Approach	Spatial rate of reduction	QPI = 27	QPI = 30	QPI = 33	QPI = 36	QPI = 39
Coastguard	(1)	54.19	38.41	36.15	34.25	32.02	30.01
	(2)		−54.93	−94.53	−150.32	−243.45	−377.32
	(3)		39.40	37.10	35.21	33.06	31.21
	(4)		38.83	36.29	34.07	31.36	28.69
Ducks	(1)	27.09	24.20	23.77	22.76	21.96	21.40
	(2)		−6.53	−18.03	−36.41	−66.19	−110.77
	(3)		23.49	22.84	21.87	21.29	20.98
	(4)		22.98	22.16	20.89	19.84	18.81
Parkjoy	(1)	45.20	25.38	23.25	21.40	19.85	19.06
	(2)		−30.72	−53.07	−88.05	−148.37	−242.25
	(3)		26.19	24.06	22.17	20.70	19.88
	(4)		25.78	23.52	21.39	19.49	18.01
Parkrun	(1)	32.56	25.78	24.96	24.22	23.45	22.80
	(2)		−8.81	−21.17	−39.59	−71.15	−119.78
	(3)		25.72	24.81	23.95	23.19	22.63
	(4)		25.31	24.26	23.20	22.07	20.94

Table 4. Percentage of bitrate saving compared with H.264/AVC as a function of QP (positive value means bitrate is decreased, negative value means bitrate is increased): (1) reduced video after the seam carving without seam coding, (2) reduced video after the seam carving with seam position encoded without modeling, (3) reduced video after the seam carving modeled without seam coding, and (4) our proposed approach with reduced video after the seam carving modeled and the seams modeled and encoded.

Sequence	Approach	Spatial rate of reduction	QPI = 27	QPI = 30	QPI = 33	QPI = 36	QPI = 39
Coastguard	(1)	54.19	23.61	20.29	17.96	15.87	15.62
	(2)		−178.21	−290.25	−477.20	−773.32	−1220.96
	(3)		24.16	20.00	16.87	14.39	13.91
	(4)		22.91	18.08	13.81	9.52	6,29
Ducks	(1)	27.09	3.36	1.16	−0.55	−0.77	1.39
	(2)		−62.75	−97.96	−159.53	−259.19	−433.29
	(3)		−1.67	−4.88	−7.31	−7.22	−3.81
	(4)		−2.76	−6.51	−9.93	−11.47	−10.96
Parkjoy	(1)	45.20	4.36	0.20	−1.33	−0.40	3.44
	(2)		−121.73	−194.55	−319.43	−521.71	−857.87
	(3)		6.03	1.60	−0.49	−0.37	3.22
	(4)		5.12	0.20	−2.77	−4.11	−2.94
Parkrun	(1)	32.56	−12.23	−15.96	−16.80	−15.73	−10.60
	(2)		−132.06	−196.38	−297.95	−455.78	−707.27
	(3)		−12.06	−16.88	−19.02	−22.17	−23.81
	(4)		−13.48	−19.02	−22.17	−23.81	−21.85

approach (3) shows that the reduced video after grouping and modeling the seams leads to better bitrate saving compared with (1) for Coastguard, Parkjoy, and Parkrun. This is due to the fact that the modeling limits the creation of high frequencies in the reduced video. But in the approach (1) and (3), it is not possible to reconstruct the geometry scene because no seam information is transmitted and as the seam carving is not a reversible process, artifacts may appear at the decoder in this case. In the approach (2), all the seam positions are encoded with the reduced video after the seam carving. The overhead cost is too important and a classical video encoding performs better. This justifies the fact that the seams positions have to be approximated. It is possible to see that with our proposed approach in (4), the overhead due to the seam encoding is low and the bitrate saving is very close to the (3).

Table 4 illustrates the percentage of bitrate saved compared with conventional H.264/AVC as a function of quantization parameter in inter coding. As in Table 3, QPI varies from 27 to 39 with a step of 3 and the quantization parameter P-frame (QPP) is defined as: QPP = QPI + 1.

It is interesting to see that even if a good rate of spatial reduction is reached and no seam are encoded, like in (1) and (3), the bitrate saving is very small for Ducks and Parkjoy at QPI = 27. For Parkrun, the reduced sequence

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

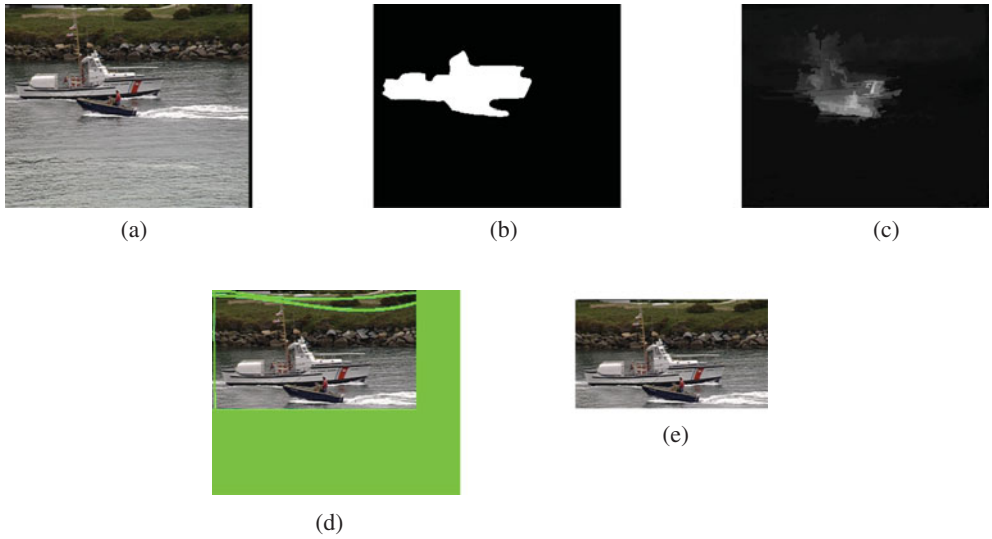


Fig. 18. Illustration of the different processes on Coastguard sequence: (a) original image, (b) binary mask of reference, (c) saliency model, (d) image with proposed seam carving and seams in green, and (e) image reduced with proposed seam carving.

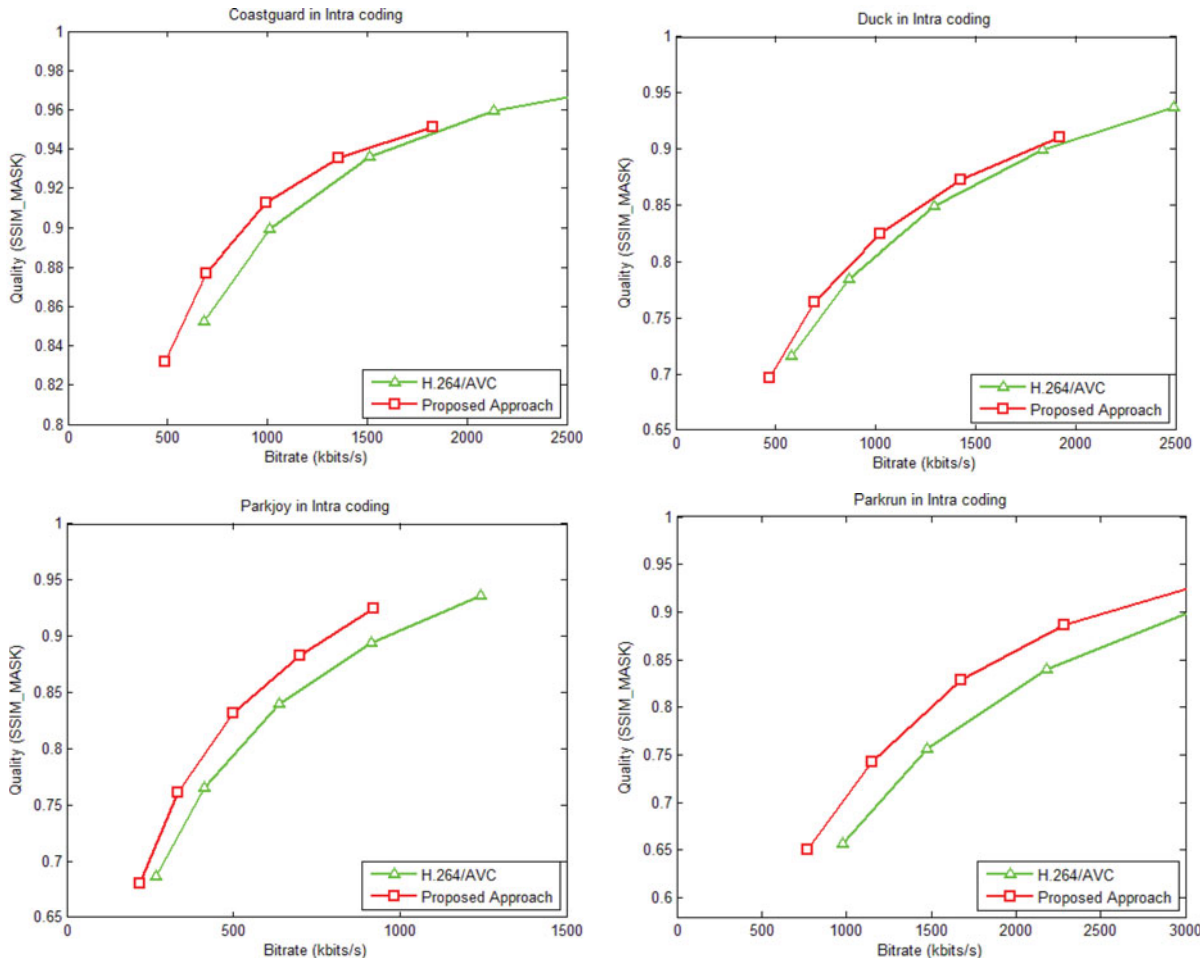


Fig. 19. Rate-distortion performance in full intra coding for Coastguard, Ducks, Parkrun, and Parkjoy, using the SSIM_MASK. In green, H.264/AVC and in red the proposed approach.

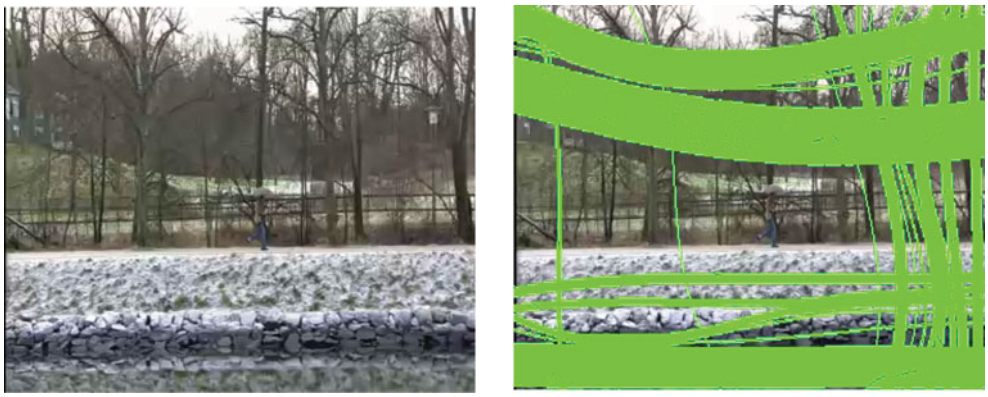
has a more important bitrate than the original one. This highlights the importance to have a really good temporal saliency model and consistent seams in the time. For Coastguard, the saliency model and the seam modeling performs well and lead to a bitrate saving in (1) and (3) even at QPI = 39.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

Table 5. Bjontegaard's scores (percentage and Delta SSIM_MASK) in full intra coding for Coastguard, Ducks, Parkrun, and Parkjoy, using the SSIM_MASK.

Coastguard	Percentage	Delta SSIM_MASK	Ducks	Percentage	Delta SSIM_MASK
Proposed approach	9.01	−0.0078	Proposed approach	7.02	−0.0105
Parkjoy	Percentage	Delta SSIM_MASK	Parkrun	Percentage	Delta SSIM_MASK
Proposed approach	21.77	−0.0305	Proposed approach	21.24	−0.0356

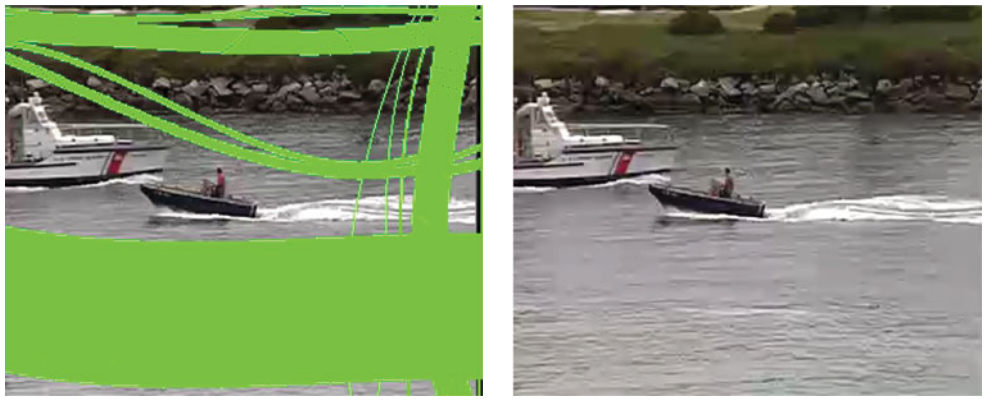
The scores are computed between H.264/AVC and our content-aware video compression approach.



Compressed video with seam carving (2183Kbits/s)

H.264 /AVC (1676Kbits/s)

Fig. 20. Rate-distortion performance. Visual results for Parkrun in intra coding with a SSIM_MASK = 0.83 and a bitrate saving of 23%.



Compressed video with seam carving (SSIM_MASK= 0.88)

H.264 /AVC (SSIM_MASK = 0.85)

Fig. 21. Rate-distortion performance. Visual results for Coastguard in intra coding with a bitrate of 488 Kbits/s.

G) Rate-distortion performance assessment

Finally, we assess the rate-distortion performance of the proposed semantic video coding scheme based on seam carving. For this purpose, comparison is made in full intra coding, which is very common and pertinent in security applications. Experiment in inter coding are also done to highlight the next challenges. The sizes of the GOP are defined by the content-aware adaptive GOP segmentation. The traditional H264/AVC [1] encoder is used as reference, with the same GOP structure than our proposed approach. Figure 18 illustrates different processes on the Coastguard sequence. Figures 18(a) and 18(d) are used to plot the rate-distortion performance.

Experimental results on the four test sequences with the metrics SSIM_MASK are presented in Fig. 19 and Table 5.

With SSIM_MASK, the proposed approach consistently outperforms H.264/AVC for all test sequences. A bitrate saving between 7.02 and 21.77% can be reached using the Bjontegaard's metric. The Delta SSIM_MASK is between −0.0078 and −0.0356. We remind that the Bjontegaard's metric with the SSIM_MASK allows to compute the average gain in SSIM only computed on the binary mask or the average per cent saving in bitrate between two rate-distortion curves. We use the Matlab implementation from Giuseppe Valenzise done in 2010.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

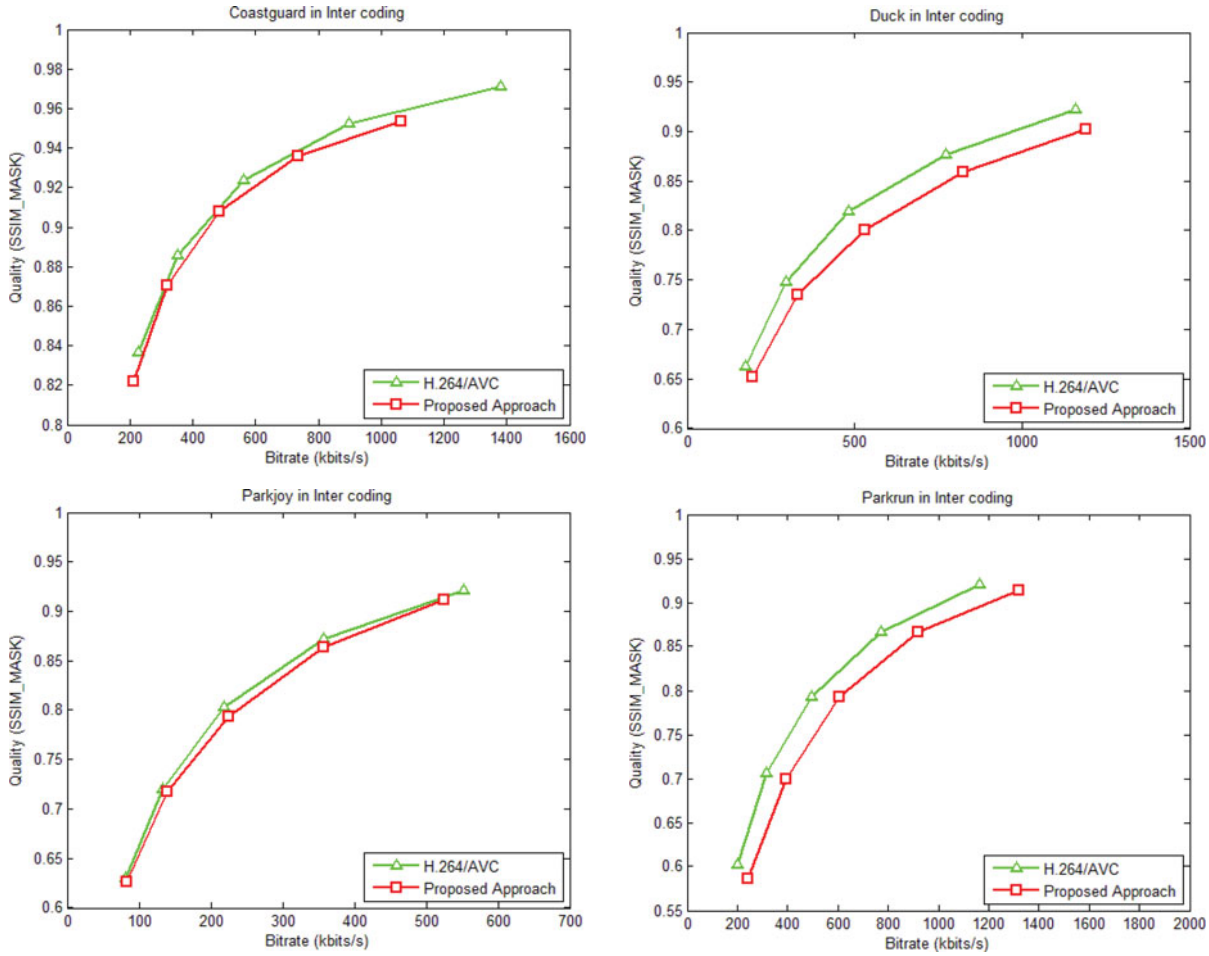


Fig. 22. Rate-distortion performance in inter coding for Coastguard, Ducks, Parkrun, and Parkjoy, using the SSIM_MASK. In green, H.264/AVC and in red the proposed approach.

Table 6. Bjontegaard's scores in inter coding for Coastguard, Ducks, Parkrun, and Parkjoy, using the SSIM_MASK.

Coastguard	Percentage	Delta SSIM_MASK	Ducks	Percentage	Delta SSIM_MASK
Proposed approach	−8.81	0.0063	Proposed approach	−18.73	0.0289
Proposed approach	−5.49	0.0071	Proposed approach	−22.34	0.0440

The scores are computed between H.264/AVC and our content-aware video compression approach.

Figure 20 illustrates a bitrate saving of 23% in intra coding on the Parkrun sequence with the same SSIM_MASK quality on the object of interest. Figure 21 illustrates the visual result for Coastguard at an equivalent bitrate of 488 Kbits/s. The compression by seam carving has been done with a QP = 36, while the traditional compression has been done with a QP = 39. The video obtained without our approach has a SSIM_MASK = 0.88 and the video with the H.264/AVC has a SSIM_MASK = 0.85. Perceptually, the quality of the object is a little bit better even if we are in the presence of strong artifacts. For example, both the body and the head of the driver in the boat are better.

To generalize our performance and highlight the next challenge, experiments with inter coding were also carried out. The GOP coding structure is IPPP, and the QPP is defined in function of the quantization parameter of the I frame, QPI, as QPP = QPI + 1.

Experimental results are presented in Fig. 22 and Table 6.

We can observe that with inter coding, the performance of the proposed approach cannot do better than H.264/AVC. This lower performance in inter coding, as opposed to intra coding, can be explained by the lack of temporal stability for the saliency model. In addition, motion compensated prediction is not as efficient when encoding the reduced video. More precisely, some static parts in the original video are shifted in the reduced sequence and from frame to frame. Also, seam carving is not always suppressing the same paths. For these reasons, the bitrates of the P frames sometimes increase and lead to these results.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

Table 6 shows the performance with the Bjontegaard's metric based on the SSIM_MASK. The performance is an increase of bitrate between 5.49 and 22.34% and a Delta SSIM_MASK between 0.0063 and 0.044.

V. CONCLUSION AND FUTURE WORK

In conclusion, we have proposed in this paper a new approach based on seam modeling for semantic video compression in video surveillance applications. More precisely, (1) a method was first presented that automatically segments the GOP as a function of the content. For each GOP, (2) a spatio-temporal seam clustering based on spatial and temporal distances and (3) an isolated seam discarding technique have been applied to improve the encoding of the reduced-size sequence and to help seam modeling. (4) A new seam modeling, avoiding geometric distortion and resulting in a better control of the seam shapes at the decoder without the need of a saliency map and (5) a new encoder that reduces the global bitrate have also been proposed.

The seam modeling was first validated, both visually and in terms of bitrate overhead to transmit the seam information. Next, the proposed semantic video coding scheme was compared with H.264/AVC using the SSIM_MASK quality metric. We used full intra coding, which is consistent with security applications but also inter coding to evaluate the temporal consistency of our approach and define future works. For the Coastguard, Ducks, Parkrun, Parkjoy test sequences, our approach outperforms H.264/AVC in intra coding with a bit-rate reduction ranging from 7.02 to 21.77% using Bjontegaard's metric [47].

In future works, we will pursue the improvement of this scheme. As we have seen, several parameters have been experimentally defined. One upgrading is to automatically define them depending on the content. In our results, the seam texture is not synthesized at the decoder side. For this purpose, some inpainting algorithms can be used. Studying the influence of the synthesis quality would also be a valuable continuation of this study. Others improvements could be done on temporal stability of saliency models and seam carving clustering to obtain better performance in inter coding.

REFERENCES

[1] Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol., **13** (7) (2003), 560–576.

[2] Sullivan, G.J.; Ohm, J.-R.; Han, W.-J.; Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol., **22** (12) (2012), 1649–1668.

[3] Avidan, S.; Shamir, A.: Seam carving for content-aware image resizing. ACM Trans. Graph., **26** (10) (2007), 10.

[4] Bertalmío, M.; Bertozzi, A.; Sapiro, G.: Navier-Stokes, fluid-dynamics and image and video inpainting, in *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, Kauai, HI, vol. 1, 2001, 355–362.

[5] Anh, N.; Yang, W.; Cai, J.: Seam carving extension: a compression perspective, in *IEEE Proc. Int. Conf. Multimedia*, Beijing, China, 2009, 825–828.

[6] Tanaka, Y.; Hasegawa, M.; Kato, S.: Image coding using concentration and dilution based on seam carving with hierarchical search, in *IEEE Proc. Int. Conf. Acoustics Speech and Signal Processing*, Dallas, TX, 2010, 1322–1325.

[7] Tanaka, Y.; Hasegawa, M.; Kato, S.: Improved image concentration for artifact-free image dilution and its application to image coding, in *IEEE Proc. Int. Conf. Image Processing*, Hong Kong, China, 2010, 1225–1228.

[8] Décombas, M.; Capman, F.; Renan, E.; Dufaux, F.; Pesquet-Popescu, B.: Seam carving for semantic video coding, in *SPIE Proc. Applications of Digital Image Processing*, San Diego, CA, 2011, 81350F.

[9] Décombas, M.; Dufaux, F.; Renan, E.; Pesquet-Popescu, B.; Capman, F.: Improved seam carving for semantic video coding, in *IEEE. Proc. Int. Workshop MultiMedia Signal Processing*, Banff, Canada, 2012, 53–58.

[10] Domingues, D.; Alahi, A.; Vandergheynst, P.: Stream carving: an adaptive seam carving algorithm, in *IEEE Proc. Int. Conf. Image Processing*, Hong Kong, China, 2010, 901–904.

[11] Achanta, R.; Susstrunk, S.: Saliency detection for content-aware image resizing, in *IEEE Proc. Int. Conf. Image Processing*, Cairo, Egypt, 2009, 1005–1008.

[12] Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S.: Frequency-tuned salient region detection, in *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, Miami, FL, 2009, 1597–1604.

[13] Rubinstein, M.; Shamir, A.; Avidan, S.: Improved seam carving for video retargeting. ACM Trans. Graph., **27** (3) (2008), 16.

[14] Chao, W.-L.; Su, H.-H.; Chien, S.-Y.; Hsu, W.; Ding, J.-J.: Coarse-to-fine temporal optimization for video retargeting based on seam carving, in *IEEE Proc. Int. Conf. Multimedia and Expo*, Barcelona, Spain, 2011, 1–6.

[15] Chambolle, A.; Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Springer J. Math. Imaging Vis., **40** (1) (2011), 120–145.

[16] Rahtu, E.; Heikkila, J.A.: Simple and efficient saliency detector for background subtraction, in *IEEE Proc. Int. Conf. Computer Vision Workshops*, Kyoto, Japan, 2009, 1137–1144.

[17] Wu, H.R.; Rao, K.R. (Eds): Digital Video Image Quality and Perceptual Coding. *CRC Press*, London, 2005.

[18] Chen, Z.; Lin, W.; Ngan, K.N.: Perceptual video coding: challenges and approaches, in *IEEE Proc. Int. Conf. Image Processing*, 2010, 784–789.

[19] Ndjiki-Nya, P.; Doshkov, D.; Kaprykowsky, H.; Zhang, F.; Bull, D.; Wiegand, T.: Perception-oriented video coding based on image analysis and completion: a review. Signal Process., Image Commun. **27** (6) (2012), 579–594.

[20] Mancas, M.; De Beul, D.; Riche, N.; Siebert, X.: Human Attention Modelization and Data Reduction. *INTECH Open Access Publisher*, Croatia, 2012.

[21] Itti, L.: Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans. Image Process., **13** (10) (2004), 1304–1318.

[22] Itti, L.; Koch, C.; Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell., **20** (11) (1998), 1254–1259.

[23] Tsapatsoulis, N.; Rapantzikos, K.; Pattichis, C.: An embedded saliency map estimator scheme: application to video encoding. Int. J. Neural Syst., **17** (4) (2007), 1–16.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

[24] Mancas, M.; Gosselin, B.; Macq, B.: Perceptual image representation. J. Image Video Process., **2007** (2) (2007), 1–9.

[25] Li, Z.; Qin, S.; Itti, L.: Visual attention guided bit allocation in video compression. Image Vis. Comput., **29** (1) (2011), 1–14.

[26] Gupta, R.; Chaudhury, S.: A scheme for attentional video compression, In Springer Pattern Recognition and Machine Intelligence, 2011, 458–468.

[27] Hou, X.; Zhang, L.: Saliency detection: a spectral residual approach, in *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, 2007, 1–8.

[28] Guo, C.; Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans. Image Process., **19** (1) (2010), 185–198.

[29] Chen, H.; Hu, R.; Mao, D.; Thong, R.; Wang, Z.: Video coding using dynamic texture synthesis, in *IEEE Proc. Int. Conf. Image Processing*, 2010, 203–208.

[30] Bosch, M.; Zhu, F.; Delp, E.: Spatial texture models for video compression, in *IEEE Proc. Int. Conf. Image Processing*, vol. 1, 2007, 93–96.

[31] Bosch, M.; Zhu, F.; Delp, E.: Segmentation based video compression using texture and motion models. IEEE Proc. Sel. Top. Signal Process., **5** (7) (2011), 1366–1377.

[32] Schwarz, H.; Marre, D.; Wiegand, T.: Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans. Circuits Syst. Video Technol., **17** (9) (2007), 1103–1120.

[33] Deng, C.; Lin, W.; Cai, J.: Content-based image compression for arbitrary-resolution display devices, in *IEEE Proc. Int. Conf. Communications*, Kyoto, Japan, 2011, 1127–1139.

[34] Vō, D.; Sole, J.; Yin, P.; Gomila, C.; Nguyen, T.: Selective data pruning-based compression using high-order edge-directed interpolation. IEEE Trans. Image Process., **19** (2) (2010), 399–409.

[35] Tanaka, Y.; Hasegawa, M.; Kato, S.: Seam carving with rate-dependent seam path information, in *IEEE Proc. Int. Conf. Acoustics Speech and Signal Processing*, Prague, Czech Republic, 2011, 1449–1452.

[36] Tanaka, Y.; Yoshida, T.; Hasegawa, M.; Kato, S.; Ikehara, M.: Rate-dependent seam carving and its application to content-aware image coding. APSIPA Trans. Signal Inf. Process., **2** (1) (2013), e1.

[37] Tanaka, Y.; Hasegawa, M.; Kato, S.: Image coding using concentration and dilution based on seam carving with hierarchical search, in *IEEE Proc. Int. Conf. Acoustics Speech and Signal Processing*, Dallas, TX, 1322–1325.

[38] Tanaka, Y.; Hasegawa, M.; Kato, S.: Generalized selective data pruning for video sequence, in *IEEE Proc. Int. Conf. Image Processing*, Brussels, Belgium, 2011, 2113–2116.

[39] Riche, N.; Décombas, M.; Mancas, M.; Dufaux, F.; Pesquet-Popescu, B.; Gosselin, B.; Dutoit, T.; Fellah, Y.: STRAP: a spatio-temporal rarity-based algorithm with priors for human fixations prediction and objects detection in videos. Submitted to Image and Vision Computing, 2014.

[40] Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S.: Frequency-tuned salient region detection, in *IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL 2009, 1597–1604.

[41] Sayood, K.: Introduction to Data Compression *Newnes*, Oxford, 2012.

[42] Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process., **13** (4) (2004), 600–612.

[43] Azuma, D.; Tanaka, Y.; Hasegawa, M.; Kato, S.: SSIM based image quality assessment applicable to resized images. IEICE Tech. Rep., **110** (368) (2011), 19–24.

[44] Lowe, D.: Distinctive image features from scale invariant key points. Int. J. Comput. Vis., **60** (2) (2004), 91–110.

[45] Liu, J.; Luo, X.; Xuan, Y.M.; Chen, W.F.; Fu, X.L.: Image retargeting quality assessment. Eurographics, **30** (2) (2011), 583–592.

[46] Décombas, M.; Dufaux, F.; Renan, E.; Pesquet-Popescu, B. Capman, F.: A new object based quality metric based on SIFT and SSIM, *in IEEE Proc. Int. Conf. Image Processing*, Orlando, FL 2012, 1493–1496.

[47] Bjontegarrd, G.: Calculation of average PSNR differences between RD curves, in *VCEG Meeting*, Austin, TX, April 2001, 2–4.

Marc Décombas graduated in 2010 from the French engineering school Telecom Sud-Paris has obtained his PhD about "Content aware video compression for very low bitrates application" with Telecom ParisTech and Thales Communications & Security, Paris, France in 2013. His research areas are content aware video compression, saliency models, video summary, subjective evaluation and people reidentification. He is the author or co-author of more than 10 research publications and holds 1 international patents issued or pending.

Frédéric Dufaux is a CNRS Research Director at Telecom ParisTech. He is also Editor-in-Chief of Signal Processing: Image Communication. He received his M.Sc. in physics and Ph.D. in electrical engineering from EPFL in 1990 and 1994 respectively. Frédéric has over 20 years of experience in research, previously holding positions at EPFL, Emitall Surveillance, Genimedia, Compaq, Digital Equipment, MIT, and Bell Labs. He has been involved in the standardization of digital video and imaging technologies, participating both in the MPEG and JPEG committees. He is currently co-chairman of JPEG 2000 over wireless (JPWL) and co-chairman of JPSearch. He is the recipient of two ISO awards for these contributions. Frédéric is an elected member of the IEEE Image, Video, and Multidimensional Signal Processing (IVMSP) and Multimedia Signal Processing (MMSP) Technical Committees. His research interests include image and video coding, distributed video coding, 3D video, high dynamic range imaging, visual quality assessment, video surveillance, privacy protection, image and video analysis, multimedia content search and retrieval, and video transmission over wireless network. He is the author or co-author of more than 100 research publications and holds 17 patents issued or pending.

Beatrice Pesquet-Popescu (IEEE Fellow) is a Full Professor at Télécom ParisTech since 2007, Head of the Multimedia Group. She was also the Scientific Director of the UBIMEDIA common research laboratory between Alcatel-Lucent Bell Labs and Institut Mines Télécom. She is a member of the FET Advisory Board of the European Commission for the H2020 program, as well as member of the Advisory Group on Part IV of Horizon 2020, "Spreading Excellence and Widening Participation" (2014–2016). She was an EURASIP Board of Governors member between 2003–2010, and an IEEE Signal Processing Society IVMSP TC (2008–2013, Vice-Chair 2015) and IDSP SC (2010–2015, Vice-Chair 2012) member and an MMSP TC member (2006–2009). In 2013–2014 she served as a Chair for the Industrial DSP Standing Committee. She is also a member of the IEEE Comsoc Technical Committee on Multimedia Communications (2009–2014, chair of the Awards Board, 2015). In

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press

2008–2009 she was a Member at Large and Secretary of the Executive Subcommittee of the IEEE Signal Processing Society (SPS) Conference Board, and she was (2012–2014) a member of the IEEE SPS Awards Board and of the IEEE SPS Conference Board (2015–2016). Beatrice Pesquet-Popescu served as an Editorial Team member for IEEE Signal Processing Magazine (2012–2014), as an Associate Editor for IEEE Trans. on Multimedia (2008–2014), IEEE Trans. on Circuits and Systems for Video Technology, IEEE Transactions on Image Processing, Area Editor for Elsevier Image Communication journal, Associate Editor for APSIPA Transactions on Signal and Information Processing, and Associate Editor for the Hindawi Int. J. Digital Multimedia Broadcasting journal and for Elsevier Signal Processing (2007–2010). She was a Technical Co-Chair for the PCS 2004 conference, and General Co-Chair for IEEE SPS MMSP2010, EUSIPCO 2012, and IEEE SPS ICIP 2014 conferences. She is a recipient of the "Best Student Paper Award" in the IEEE Signal Processing Workshop on Higher-Order Statistics in 1997, of the Bronze Inventor Medal from Philips Research and in 1998 she received a "Young Investigator Award" granted by the French Physical Society. In 2006, she was co-recipient of the IEEE Trans. on Circuits and Systems for Video Technology "Best Paper Award". In April 2012, Usine Nouvelle cited her among the "100 who matter in the digital world" in France. She holds 29 patents in video coding and has authored more than 320 book chapters, journal and conference papers in the field. She is a co-editor of the book "Emerging Technologies for 3D Video: Creation, Coding, Transmission, and Rendering", Wiley Eds., 2013. She has (co-)directed 27 PhD students and was in 30+ PhD committee defense juries. Her current research interests are in source coding, scalable,robust and distributed video compression, multi-view video, 3DTV, holography, sparse representations and convex optimisation.

https://doi.org/10.1017/ATSIP.2015.4 Published online by Cambridge University Press