

# Optimised Selection of Structure of Pictures for Video Coding

Vigneswaran Poobalasingam and Ebroul Izquierdo  
School of Electronic Engineering and Computer Science  
Queen Mary University of London (UK)

Saverio G. Blasi and Marta Mrak  
BBC Research & Development  
London (UK)

**Abstract**— Encoders based on the High Efficiency Video Coding (HEVC) standard consider an input sequence as a succession of slices grouped in Structures of Pictures (SOP). The SOP used while encoding specifies many parameters, such as the coding order of frames, or the reference frames used during interprediction. Reference encoders typically make use of a fixed SOP structure of a given size, which is periodically repeated throughout the whole sequence. In this paper, the usage of unconventional SOP structures is first analysed, showing that most sequences benefit from usage of larger SOPs, and that the selection of the optimal SOP is highly content dependent. As a result, an algorithm is proposed to automatically select the optimal SOP size based on a low-complexity texture analysis of neighbouring frames. The algorithm is capable of adaptively changing SOP size during the encoding. Extensive evaluation shows that consistent bitrate reductions are reported at the same objective quality as an effect of using the proposed algorithm.

**Keywords**—HEVC, video coding

## I. INTRODUCTION

H.265/High Efficiency Video Coding (HEVC) standard [1] was developed by the Joint Collaborative Team on Video Coding (JCT-VC) as a successor to H.264/Advanced Video Coding (AVC) [2]. Although HEVC is based on a similar architecture to AVC, the standard includes many new coding tools and enhancements than its predecessor. Thanks to these improvements, HEVC can reportedly achieve more than 50% bitrate reductions for the same subjective quality compared to AVC [3].

Similar to AVC, a video sequence is encoded in HEVC as a sequence of entities called slices. In most cases it is common to code each frame in the sequence as exactly one slice, as will be assumed in the rest of this paper. Three slice types are supported, namely I (in which only intra-prediction may be used), P (in which uni-prediction may also be used) or B (in which biprediction may also be used). In order to exploit as much as possible the temporal redundancy within consecutive frames, the coding order, namely the order in which slices are compressed in the bitstream, may be different to the output order, namely the order in which they should be displayed.

Slices are grouped in sets of consecutive frames referred to as Structures of Pictures (SOP). The SOP used while encoding defines a multitude of parameters for each slice in the SOP: the coding order, the reference frames used while performing interprediction, the Quantisation Parameter (QP) offset used for the quantisation step, etc. Typical HEVC implementations, such as the one in the HEVC test Model (HM) reference software developed by the JCT-VC, consider three main SOP

structures [4], which are periodically repeated during encoding depending on the access configuration. In the All-Intra configuration, all frames are coded as I slices (SOP size is 1). This is only useful for particular applications such as editing, where no inter-frame dependencies are allowed. Coding efficiency of HEVC in the Intra configuration is limited, due to temporal redundancy not being exploited. In the Low-Delay (LD) configuration, the frames are coded in display order. Frames can be encoded as P or B slices, where only frames whose temporal index is lower than the temporal index of the current slice can be used as reference frames. The SOP size is set to 4. The LD configuration is useful in applications which require the lowest delay during decoding: due to the fact that slices are compressed in the same order as they should be displayed, frames can be displayed immediately after decoding. Conversely, the Random-Access (RA) configuration makes use of a more complex coding order where frames are encoded as P or B slices and are compressed in a different order than they are displayed. SOP size is set to 8. Due to temporal redundancy being exploited efficiently, RA configuration can in general achieve higher compression efficiency than LD. The SOP structures typically used in the LD and RA configurations are shown in Figure 1.

While most encoder implementations make use of a fixed SOP structure throughout the encoding of a sequence, the HEVC specification does not enforce this restriction. SOP size and structure can be changed throughout the encoding while still producing a bitstream that is standard compliant. The benefits of using different SOP structures to those in Figure 1 have already been investigated. In fact, the usage of longer SOP sizes of 16 or 32 slices is being considered in the investigative work carried out towards the definition of the successor to HEVC [5]. In this paper, an analysis of the effects of different SOP structures and sizes is presented, with the objective of evaluating the impact of different structures on the coding efficiency while encoding different content with a wide variety of temporal and spatial characteristics. An algorithm is then proposed to flexibly change the SOP structure while encoding, depending on specific features of the sequence. The algorithm is shown to be beneficial to HEVC coding efficiency, selecting the optimal SOP size in the majority of the cases and hence resulting in lower bitrates for the same objective visual quality of the decoded signal.

## II. STATE OF THE ART

Very little work has been presented in the past to show the benefits of using different SOP structures in HEVC. Recently, some experiments were presented in the context of the Joint

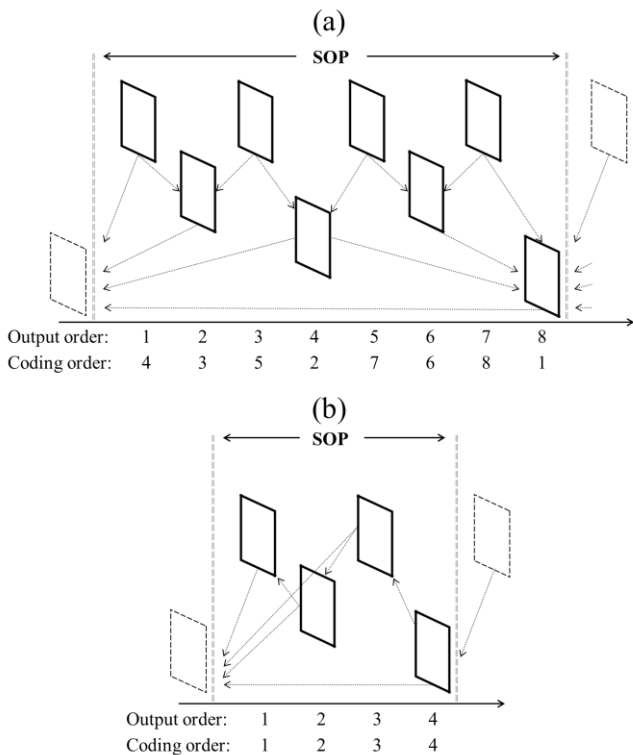


Figure 1 - SOP structures used in the (a) RA and (b) LD configurations

Video Exploration Team (JVET) responsible for exploring promising technology towards the definition of new video coding standards. New SOP structures and sizes of 16 and 32 were tested [5], resulting on average in bitrate reductions of 7% and 10%, respectively, for the same objective quality. Still, these experiments focused only on using a fixed SOP size throughout the encoding, and did not consider methods to detect the optimal SOP size based on the content.

Some work was introduced in the context of Distributed Video Coding (DVC). A method was presented [6], based on the calculation of different “motion activity metrics” to compare visual characteristics of neighbouring frames. Global (at frame level) and local (at block level) changes are detected, and used to derive the optimal SOP size for the current portion of the sequence being encoded. A similar approach was proposed [7] where the rate necessary to encode a given frame is estimated with no knowledge on the motion information. A few approaches were presented in the context of the AVC standard. In [8] methods from perceptual video coding are used to estimate the optimal SOP size. The idea is that of detecting large perceptual differences in a video sequence, defined using two dimensional entropy and pixel dissimilarity. Finally, a method was proposed [9] in which the motion vectors from motion estimation are used, to form the sum of the absolute motion vector (SAMV) and motion residuals, which is then used to perform a decision on the SOP size. Both these methods require knowledge on the motion activity of the current portion of the sequence being encoded.

The method proposed in this paper is different in that it is designed in the context of an HEVC encoder scheme without

requiring any prior knowledge on the content of the portion of the sequence being encoded. As such, the approach has a negligible impact on computational complexity while achieving high accuracy in detecting the optimal SOP size for each portion of the encoded sequence.

### III. EFFECTS OF SOP SIZE ON HEVC CODING

In order to analyse the impact of using different SOP sizes while encoding using HEVC, five SOP structures were selected. In particular, the two SOP structures in Figure 1 (a) and (b) were considered, referred to in this paper as RA8 and LD4, respectively. Then, RA8 was used as a template to design three additional SOP structures of size 4, 16 and 32, referred to as RA4, RA16 and RA32, respectively. These five structures were then used to encode a test set of video sequences with spatial resolution of  $1920 \times 1080$  luma samples, and varying temporal resolutions ranging from 24 to 60 frames per seconds (fps). Encoding was performed using HM version 16.7 under the Common Test Conditions (CTC) [4] specified by the JCT-VC, using four Quantisation Parameter (QP) values of 22, 27, 32 and 37. While the HEVC standard does not impose any correlation between the SOP size and the Intra-Period (IP, namely the interval between two consecutive I slices), in most common HEVC encoders (including the HM reference software), the IP is assumed to be a multiple of the SOP size, so that an integer number of SOPs fits in each IP. For this reason, in order to use the RA16 and RA32 SOP structures, all subsequent tests were performed using IPs of either 32 or 64 (whichever is closest to the fps of the corresponding sequence).

Results of these tests are presented in Table 1. In particular, the table presents a comparison of results obtained with LD4, RA4, RA16 and RA32, compared with RA8 used as an anchor. Compression efficiency is presented in terms of BjøntegaardDelta rate (BD-rate) [10] where only luma values are shown, and negative values correspond to savings with respect to the anchor. The results show that in most cases, usage of larger SOP sizes provide gains. Only in one case, LD4 provided the best results (the Riverbed sequence), and in one case RA8 was the best (the DucksTakeOff sequence).

### IV. SOP CHANGE DETECTION

The results in Section III show that different SOP sizes from the conventionally used RA8 or LD4 can lead to potentially very high gains in compression efficiency, but also highlight that

Table 1-IMPACT OF USING DIFFERENT SOP STRUCTURES COMPARED TO RA8. BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Sequence	fps	BD-rates for SOP structure [%]			
		L4	RA4	RA16	RA32
BQTerrace	60	5.91	8.59	-2.20	<b>-9.74</b>
DucksTakeOff	50	15.07	0.36	5.03	3.20
Riverbed	25	<b>-4.74</b>	-0.30	0.37	-0.25
Station	25	14.31	5.81	-3.15	<b>-8.13</b>
Tractor	25				
FourPeople	60	8.90	13.05	-8.66	<b>-13.71</b>
Average	-	12.35	7.55	-3.54	-9.03

results greatly vary depending on content. Ideally, the encoder should be able to select the optimal SOP structure for each portion of the sequence being encoded. In theory, the encoder could analyse a set of look-ahead frames and select the SOP structure based on Rate-Distortion (RD) decisions. For instance, the encoder could perform a multi-pass coding of a portion of the sequence using the available SOP structures, and therefore select the best configuration. Unfortunately this solution is impractical in most situations due to the high complexity overload: the encoder would need to perform the time consuming inter-prediction, transform and entropy coding steps multiple times, hence considerably increasing the time needed for compression.

For this reason an alternative solution is proposed in this paper based on a pre-analysis of the content of a portion of the sequence. The pre-analyser should be capable of detecting the optimal SOP structure, without performing motion searches or other computationally expensive encoder tasks. For this purpose, in this paper the MPEG-7 Homogeneous Texture Descriptor (HTD) [11] was selected as basis for the proposed pre-analysis step. The HTD was introduced in the MPEG-7 standard as a tool to quickly retrieve images based on texture similarity. The idea is that of comparing the energy deviations of selected frequency bands of the image, where the partitioning in bands is performed according to characteristics of the human visual systems. In doing so, texture characteristics are efficiently exploited to obtain very high matching rates [12]. The HTD as defined in the MPEG-7 standard is computed on image areas of  $128 \times 128$  samples. Only the luminance (luma) component is considered. First, the mean ( $f_{mn}$ ) and standard deviation ( $f_{sd}$ ) of all luma samples in the image area are computed. Subsequently, the image is transformed using Fourier transform and expressed in the polar coordinates. After that, a Gabor filter is applied to strengthen the image directional information. The outcome of the Gabor filter is denoted  $H_i(\omega, \theta) = G_{sr}(\omega, \theta) \times F(\omega, \theta)$ , where  $i$  is the number of frequency sub-bands,  $i \in \{1, 2, \dots, 30\}$ . Finally, 30 energy coefficients [ $e1, e2, \dots, e30$ ] and corresponding energy deviations [ $d1, d2, \dots, d30$ ] are computed. The HTD of a given block in a frame is then defined as a 62-component feature vector:

$$HTD = [f_{av}, f_{sd}, e_1, \dots, e_{30}, d_1, \dots, d_{30}] \quad (1)$$

In the rest of this paper we will refer to the  $j$ -th element in each HTD vector as  $HTD[j]$ .

In the context of this paper, the HTD can be used to identify abrupt variations in the texture content of neighbouring frames. Some tests were performed to validate this assumption. In particular, the first 50 frames of the same sequences as in Table 1 were considered. Denote as  $H$  and  $W$  the height and width (in luma samples) of the sequence, respectively. Each frame was partitioned in  $K$  image areas of  $128 \times 128$  luma samples. In case  $H$  is not a multiple of 128, a number of  $\lfloor H/128 \rfloor$  image areas covering the top portion of each frame is considered, while the remaining few bottom-most rows of samples are discarded. Similarly, in case  $W$  is not an integer multiple of 128, few rightmost columns of luma samples are discarded. For each image area  $k=0, \dots, K-1$  in frame  $n=0, \dots, 49$ , the HTD was then computed, denoted as  $HTD_{k,n}$ . As a measure of similarity of two HTDs, the Sum of Absolute Differences (SAD) was calculated for each pair of HTD vectors extracted from collocated image areas in neighbouring frames in the sequences, as:

$$SAD_{k,n} = \sum_{j=0}^{61} |HTD_{k,n-1}[j] - HTD_{k,n}[j]|, n = 1, \dots, 49$$

Finally, SAD values computed from image areas extracted from the same frames were cumulated and normalised to the total number of image areas per frame, to obtain a single value for each frame pair, or:

$$SAD_n = \sum_{k=0}^{K-1} \frac{SAD_{k,n}}{K}, n = 1, \dots, 49 \quad (2)$$

The results of these tests are shown in Figure 2, where  $SAD_n$  for each frame pair  $n$  is plotted for some selected sequences in the test set. As can be noticed, there is a high correlation between the obtained SAD values, and the results in Table 1. Sequences which benefit from using short SOP sizes, such as *Riverbed*, tend to return higher SAD values, which corroborates the idea that sequences which present high temporal activity among neighbouring frames are better encoded using shorter SOP sizes. Conversely, sequences which benefit from large SOP sizes tend to assume lower SAD values. Unfortunately, the SAD is not always a good classifier, as can be observed for instance from the results obtained with the *DucksTakeOff* sequence. This sequence is best encoded using the RA8 structure, even though it returns lower SAD values than sequences, such as *Tractor*, that are best encoded with RA32. This behaviour can be illustrated by analysing the texture activity in the sequence in the spatial domain. The average SAD between intensities of collocated luma samples in pairs of neighbouring frames was computed for the first 50 frames in the sequence, compared with the same value obtained with the *Tractor* sequence. *DucksTakeOff* resulted in average SAD of 62.10 per sample, while *Tractor* returned higher SAD of 65.07. This reflects the results obtained in the frequency domain in Figure 2. Conversely, the percentage of different

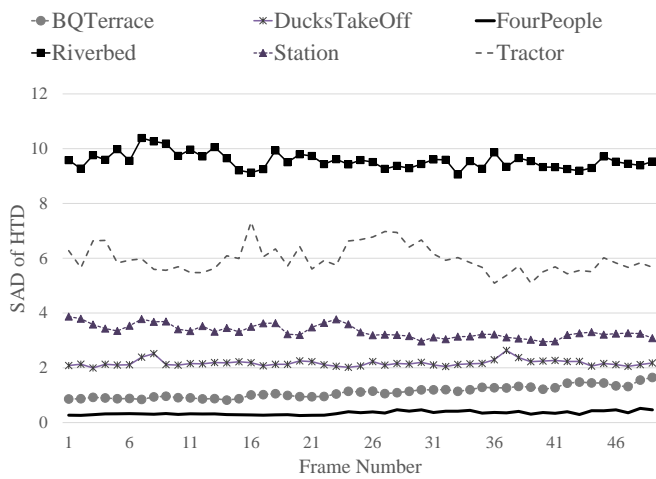


Figure 2 - SAD of HTD computed on neighbouring frames

luma samples was then computed among pairs of neighbouring frames for the first 50 frames in both sequences. In other words, while most samples among pairs of neighbouring frames in the *DucksTakeOff* sequence are different, the average distance is low, which results in low SAD values.

For these reasons a different distance metric is proposed here between two HTD vectors. Formally, given two vectors  $\mathbf{HTD}_{k,n-1}$  and  $\mathbf{HTD}_{k,n}$  of length 62 extracted from collocated image areas of  $128 \times 128$  luma samples in a pair of neighbouring frames at time indexes  $(n-1)$  and  $n$ , first the vectors are normalised.

In particular, denote the maximum element in  $\mathbf{HTD}_{k,n}$  as  $C_{k,n}^{\max} = \mathbf{HTD}_{k,n}[\mathbf{m}]$ , such that  $\mathbf{HTD}_{k,n}[\mathbf{m}] \geq \mathbf{HTD}_{k,n}[\mathbf{j}], \forall \mathbf{j}$ , and similarly the minimum element as  $C_{k,n}^{\min}$ . Denote the normalised HTD vector as  $\overline{\mathbf{HTD}}_{k,n}$  where:

$$\overline{\mathbf{HTD}}_{k,n}[\mathbf{j}] = \frac{\mathbf{HTD}_{k,n}[\mathbf{j}] - C_{k,n}^{\min}}{C_{k,n}^{\max} - C_{k,n}^{\min}}$$

Equivalently the normalised HTD vector  $\overline{\mathbf{HTD}}_{k,n-1}$  can be computed. Finally, the Inverse Hamming Distance (IHD) between  $\mathbf{HTD}_{k,n-1}$  and  $\mathbf{HTD}_{k,n}$  is defined as:

$$\mathbf{IHD}_{k,n} = 62 - \sum_{\mathbf{j}=0}^{61} d\{\overline{\mathbf{HTD}}_{k,n-1}[\mathbf{j}], \overline{\mathbf{HTD}}_{k,n}[\mathbf{j}]\}, \quad (3)$$

$\mathbf{n} = 1, \dots, 49$

Where  $\mathbf{n} = 1, \dots, 49$ , and  $d\{x, y\} = 0$  if  $x = y$ , or  $d\{x, y\} = 1$  otherwise.

The HTD in (3) is computed for all  $\mathbf{K}$  collocated image areas in each frame and then averaged to obtain  $\mathbf{IHD}_n$  as a measure of the texture activity between frames at time indexes  $\mathbf{n} - 1$  and  $\mathbf{n}$ .

Using the expression in Equation (3), an automatic strategy for optimal SOP structure selection was derived. This strategy is based on a simple statistical analysis of the behaviour of the mean and standard deviation of the vectors  $\mathbf{IHD}_n$ , for all sequences  $l = 1, \dots, L$  in the test set. The mean  $\mu_l$  and standard deviation  $\sigma_l$  are then computed. First, it was observed that sequences which seem to benefit from usage of short SOPs, such as LD4 or RA8, tend to cluster within the first quartile of all estimated means in the observation data. Formally, consider the first quartile of the sample distribution of  $\mu_l$ , referred to as  $\lambda$ . Then, if  $\mu_l < \lambda$ , the optimal SOP size for sequence  $l$  is likely to be short (namely LD4, RA4 or RA8). Conversely, in case  $\mu_l \geq \lambda$ , the optimal SOP size for  $l$  is with high probability RA16 or RA32. Assuming that the vectors  $\mathbf{IHD}_n$  are normalised to the unit and the data obeys a uniform density function, then we can set  $\lambda = 0.25$ . Note that such assumption is reasonable, since for a given video it is impossible to predict its corresponding texture activity without any prior analysis or knowledge.

Another important aspect is that the distributions of the means  $\mu_l$  clustered with respect to the optimal SOP structure present significant overlaps. On the other hand, the

corresponding distributions of the standard deviations  $\sigma_l$  significantly differ in each case, and therefore can be used to better discriminate among the different SOP structures being considered.

Finally, the aforementioned observations were aggregated to form the following algorithm:

Algorithm 1:

- If  $(\mu_l < \lambda/3)$  then use LD4
- else if  $(\mu_l < 2\lambda/3) \wedge (\sigma_l^2 > \varepsilon)$ , then use RA4
- else if  $(\mu_l < \lambda) \wedge (3\sigma_l^2 \in [\lambda, 2\lambda])$ , then use RA8
- else if  $(\mu_l < 4\lambda/3)$ , then use RA16
- else if  $(\mu_l \geq 4\lambda/3)$ , then use RA32,
- else use RA8

Here, the value of the parameter  $\varepsilon$  serves to separate the selection of RA4 and RA8. From empirical observations this parameter is set to  $\varepsilon=0.005$ , in the rest of this paper.

Algorithm 1 can be used in the context of an HEVC encoder to select the optimal SOP size during the encoding. Assume encoding is performed with an  $\mathbf{IP}$  of  $N$  frames. The first  $\mathbf{IP}$  in the sequence is considered. Each frame is split into  $\mathbf{K}$  image areas of  $128 \times 128$  luma samples. The  $\mathbf{HTD}_{k,n}$  vector is then computed for each image area  $k$ , frame  $n$ . Correspondingly,  $\mathbf{IHD}_n$  is computed. The vector  $\mathbf{IHD}_n, n = 1, \dots, N$  is then used to compute mean  $\mu$  and standard deviation  $\sigma$ . Finally Algorithm 1 is applied to determine the optimal SOP structure, which is used throughout the encoding of the frames in the current  $\mathbf{IP}$ . The process is repeated for each  $\mathbf{IP}$  until completion of the whole sequence.

## V. RESULTS

The proposed algorithm was implemented using the HM reference software (version 16.7), on a test set of 14 sequences with spatial resolution of  $1920 \times 1080$  luma samples, extracted from a variety of well-known test sets. Notice that for completeness, the six sequences presented in Table 1 are included in this test set, but eight additional sequences are also presented to show that the algorithm works well across sequences with different texture characteristics and that the derived parameters are representative of a wide variety of conditions. Encoding was performed according to the CTC using QP values of 22, 27, 32 and 37. The method is compared with an anchor HEVC encoder based on HM making use of a fixed SOP structure set to RA8. Results are presented in terms of luma BD-rates, where again negative BD-rate values represent efficiency gain with respect to the anchors.

The main results are presented in Table 2. As can be seen, the proposed algorithm provides an average BD-rate gain of -5.11% with respect to the anchor. As a consequence of the high sensitivity to different content characteristics, results vary a lot from sequence to sequence. In fact, the proposed approach provides losses with respect to the anchor in two cases (the *DucksTakeOff* and *Tractor* sequences, resulting in losses of

5.16% and 2.44%, respectively), while it achieves very high gains up to -14.81% (in the case of the *Library* sequence).

Table 2 - RESULTS OF THE PROPOSED SOP SELECTION METHOD.

Sequence	fps	BD-rate [%]	Min. SOP	Max. SOP
BQTerrace	60	-6.51	RA8	RA32
DucksTakeOff	50	5.16	RA8	RA16
Riverbed	25	-4.74	LD4	LD4
Station	25	-6.65	RA16	RA32
Tractor	25	2.44	RA4	RA4
FourPeople	60	-13.71	RA32	RA32
Netball	50	-1.39	RA4	RA32
RushHour	30	-5.75	RA32	RA32
Vault	50	-1.09	RA4	RA32
Wood	30	-10.87	RA16	RA32
CampfireParty	30	-5.3	RA16	RA16
Hurdles	50	-9.78	RA16	RA32
Library	30	-14.81	RA16	RA32
LongJump	50	-1.86	RA8	RA32
<i>Average</i>	-	<b>-5.11</b>	-	-

Table 2 also presents the minimum and maximum size of SOP selected throughout encoding of each sequence (where the SOP structure LD4 is considered “smaller” than RA4). It can be seen that the majority of sequences make use of a maximum SOP of RA32, while the minimum SOP size is highly content dependent. By comparing the results in Table 2 with the figures in Table 1, it can be observed that the proposed algorithm is considerably accurate in selecting an optimised SOP structure for each sequence. For instance, L4 was correctly selected for all IPs in the Riverbed sequence, which according to the results in Table 1 is the optimal choice in an RD sense. Similarly, RA32 was selected for all IPs in the *FourPeople* sequence. A visualisation of the SOP selected by the proposed algorithm in the first 162 frames of the *LongJump* and *Vault* sequences is shown in Figure 3. The algorithm adapts to variations in the texture activity, and adaptively selects optimised SOP sizes throughout the encoding.

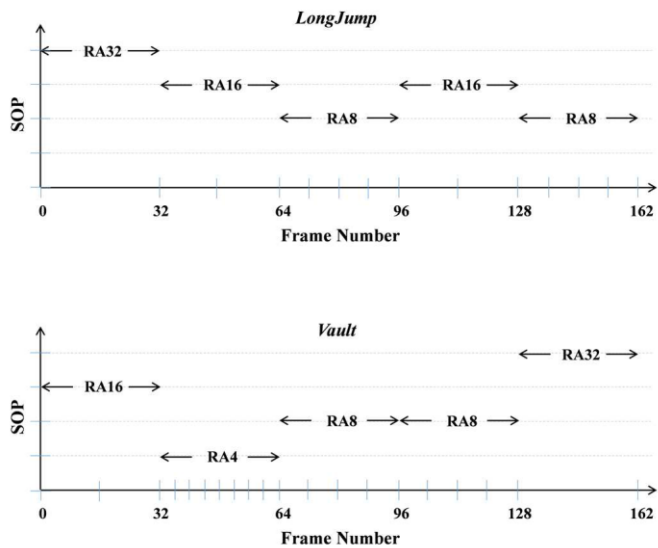


Figure 3 - SOP selected in first 162 frames of the LongJump and Vault sequences

Finally, it should be noticed that the algorithm has a negligible impact on the encoding complexity. In fact, due to the fact that the algorithm predominantly selects larger SOP sizes, and that encoding using such SOP structures is typically faster than encoding using RA8, the algorithm required in average 3.7% less encoding time with respect to the anchor, averaged with respect to all sequences and all QPs in the test set. Decoding complexity is virtually unaffected by the proposed method.

## VI. CONCLUSIONS

While the HEVC standard allows frames in a sequence to be flexibly encoded using variable SOP sizes, most reference encoders assume a fixed SOP structure which is periodically used to encode all frames throughout the sequence. This paper investigates the usage of unconventional SOP structures of different sizes than those typically used in the RA or LD profiles defined by the JCT-VC. It is shown that most sequences benefit from usage of larger SOP sizes, but also that the optimal SOP structure is highly content-dependent. As a result, an algorithm was presented in the paper, which makes use of the HTD descriptor to obtain a low-complexity estimation of the texture activity in the current portion of the sequence being encoded. This is then used to select which SOP structure to use dynamically during the encoding of the sequence. The algorithm is shown to provide consistent improvements in encoding efficiency with respect to reference HEVC implementations in the majority of the cases, achieving an average of -5.11% bitrate reductions at the same objective quality. Future work will focus on extending the algorithm to additional SOP structures as well as trying to further improve the accuracy of the selection process.

## ACKNOWLEDGEMENT

This work was supported by an EPSRC I-CASE studentship, in collaboration with BBC Research & Development.

## REFERENCES

- [1] Sullivan, G.J.; Ohm, J.; Woo-Jin Han; Wiegand, T., "Overview of the High Efficiency Video Coding (HEVC) Standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649,1668, Dec. 2012
- [2] Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.13, no.7, pp.560,576, July 2003
- [3] Tan, T. K.; Weerakkody, R.; Mrak, M.; Ramzan, N.; Baroncini, V.; Ohm, J.-R.; Sullivan, G. J. , "Video quality evaluation methodology and verification testing of HEVC compression performance," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 26, no. 1, pp. 4–19, Jan 2016
- [4] Bossen, F., "Common HM test conditions and software reference configurations", Tech. Rep. JCTVC-L1100, Jan. 2013
- [5] Andersson, K.; Wennersten, P.; Sjöberg, R.; Samuelsson, J.; Ström, J.; Hermansson, P.; Pettersson, M., "Non-normative JEM encoder improvements", Tech. Rep. JVET-B0039, Feb. 2016
- [6] Ascenso, J.; Brites, C.; Pereira, F., "Content adaptive Wyner-Ziv video coding driven by motion activity", *Image Processing, 2006 IEEE International Conference on*, Atlanta, GA, 2006
- [7] Yaacoub, C. , Farah, C., Pesquet-Popescu, B., "Content adaptive GOP size control with feedback channel suppression in distributed video coding," *IEEE International Conference on Image Processing (ICIP)*, Cairo, 2009

- [8] Zatt, B. , Porto, M. , Scharcanski J. , Bampi, S. , “Gop structure adaptive to the video content for efficient H.264/AVC encoding,” 2010 IEEE International Conference on Image Processing, Hong Kong, 2010
- [9] Ding, J. R.; Yang, J. F.; Lin, J. K., “Motion-based Adaptive GOP Algorithms for Efficient H. 264/AVC Compression”, *Journal of Communication and Information Systems*, 2006
- [10] Bjøntegaard, G., “Calculation of average PSNR differences between RD-curves”, VCEG-M33, 13<sup>th</sup> meeting, Austin, TX, April 2001
- [11] Yong Man Ro, Munchurl Kim, Ho Kyung Kang, B.S.Manjunath, Jinwoong Kim, “MPEG-7 Homogeneous Texture Descriptor,” *Electronics and Telecommunications Research Institute Journal*, vol. 2, pp. 41–51, July 2001
- [12] Shao, H. , Ji, J. , Kang, Y. ,Zhao, H., “Application Research of Homogeneous Texture Descriptor in Content-Based Image Re-trieval,” *Information Engineering and Computer Science International Conference*, Wuhan, China, Dec 2009