EFFICIENT VIDEO RESOLUTION ADAPTATION USING SCALABLE H.265/HEVC

Hoda Roodaki *† *Kemal Ugur* * *Miska M. Hannuksela* * *Moncef Gabbouj* †

*Nokia Research Center, Tampere, Finland
† Tampere University of Technology, Tampere, Finland

ABSTRACT

Dynamically changing the spatial resolution in a video conferencing session is useful for seamlessly adapting the bitrate to changing network conditions and for improving the user experience. Similar to earlier standards, the emerging High Efficiency Video Coding (H.265/HEVC) standard does not allow prediction across different resolutions, so an Instantaneous Decoding Refresh (IDR) picture must be sent to reinitialize the stream when a resolution change happens. IDR pictures take significantly more bits compared to predictively coded pictures. Thus, using them for resolution switching significantly reduces coding efficiency and increases the delay. In this paper we propose a method to support efficient adaptive resolution change using the emerging scalable H.265/HEVC standard. The proposed approach utilizes the inter-layer predicted random access pictures at the enhancement layer for resolution switching, instead of IDR pictures. The experimental results show that when the proposed method was used, the bitrate was reduced at the switching point by 34% on average for the tested video sequences. In addition, visual examples are shown demonstrating the improved visual quality with the proposed method.

Index Terms— *H.265/HEVC*, video coding, scalable, adaptive resolution change, SHVC.

1. INTRODUCTION

In video conferencing applications, it is often desired to dynamically change the resolution of a participant within the conferencing session for better adaptation to changing network conditions. The benefits for changing the resolution dynamically can be listed as follows [1][2]:

- Better network adaptation and error resilience: For better adaptation to changing network requirements for different content, it is desired to be able to change both the temporal/spatial resolution in addition to quality.
- Fast start: The start-up time of the session could be increased by first sending a low resolution picture and then increasing the resolution over time.
- Conference compose: When the person is silent, a low resolution video is sent and when he/she starts speaking, the corresponding video resolution is increased.

Changing the resolution within the sequence can be done by transmitting an Instantaneous Decoder Refresh (IDR) picture as predicting between different resolutions is not allowed in video coding standards, such as the High Efficiency Video Coding (H.265/HEVC) [3] and the Advanced Video Coding (H.264/AVC) [4]. IDR pictures are coded only by using intra-prediction techniques, which means that they take significantly more bits compared to predictively coded pictures (P and B pictures). For this reason, utilizing IDR pictures to change the resolution reduces the coding efficiency and increases the end-to-end delay of video conferencing session as it causes a sudden increase in the bitrate. In order to reduce the delay, the quality of the IDR picture could be reduced in a manner that the bitrate of the IDR picture is not higher than that of the corresponding predictively coded picture. However, reducing the quality of the IDR picture causes a sudden and clearly visible blip in the video quality and reduces the user experience. In addition, an IDR picture coded at a reasonable quality uses significantly more bits compared to an inter coded picture, so the decoding process of an IDR picture can be considered to be more complex.

In this paper we propose a method to support efficient adaptive resolution change using the emerging scalable H.265/HEVC standard [5][6]. Instead of using IDR pictures, the proposed approach utilizes the inter-layer-predicted pictures from the low-resolution base layer to the high-resolution enhancement layer for resolution switching. Since the scalable video coding inherently includes mechanisms for upsampling and inter-layer prediction, it can efficiently support the adaptive resolution change use cases listed above. It should be noted that this functionality was proposed to the Joint Collaborative Team on Video Coding (JCT-VC) and it is supported by the draft scalable extension of the H.265/HEVC standard [7]. The provided experimental results show that the proposed method for adaptive resolution change reduces the bitrate at the switching point by 34% on average for the tested video sequences. The rest of the paper is organized as follows. First, an overview of H.265/HEVC and scalable H.265/HEVC are provided in Section 2. The proposed method of using scalable H.265/HEVC for adaptive resolution change is described in Section 3. Section 4 provides the experimental results and finally the paper ends in Section 5 with the concluding remarks.

2. H.265/ HEVC AND ITS SCALABLE EXTENSIONS

The High Efficiency Video Coding (H.265/HEVC) standard was developed by the Joint Collaborative Team on Video Coding (JCT-VC), which is comprised of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group and it was technically finalized in January 2013. Similar to earlier video coding standards, H.265/HEVC employs a hybrid video coding approach, where a prediction signal is first formed by inter and intra prediction means and the residual signal is then coded with transform coding. H.265/HEVC achieves similar visual quality as H.264/AVC but roughly at half the bitrate [8]. The gains in coding efficiency are achieved by improving and re-designing many parts of the design, such as flexible quad-tree partitioning of coding blocks, improved in-loop and interpolation filtering, more flexible intra prediction, etc. A more detailed explanation of the new coding tools of H.265/HEVC can be found in [9].

At the October 2012 meeting, JCT-VC initiated new work to extend H.265/HEVC to include scalability. The planned extension is called Scalable H.265/HEVC (SHVC) and it is aimed to support spatial and quality scalability use cases. The current SHVC draft standard achieves high coding efficiency by exploiting the correlation between different layers. For example, for the case of spatial scalability, the base layer is first upsampled and made available for prediction of higher resolution enhancement layer samples. An example of the inter-layer prediction mechanism is illustrated in Figure 1, where the enhancement layer improves both the frame rate and the spatial resolution of the base layer. The current SHVC draft standard [10] includes a way to achieve scalability using high-level syntax only changes (i.e. without including any new tools below slice header level). This design decision is different than in the earlier scalable extension of H.264/AVC, which included changes in block-level coding tools, and it is expected that this codec design would increase the adoption of SHVC as existing H.265/HEVC decoders could easily be extended to decode SHVC bitstreams with minimal changes. For spatial scalability, the SHVC draft standard achieves coding efficiency improvements over simulcast ranging from 19.4% to 28.0% for different test conditions [11].

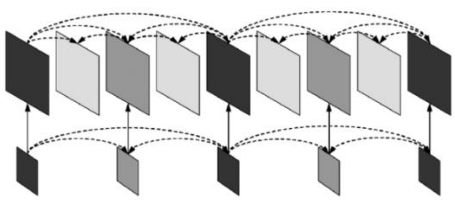


Figure 1. Inter-layer prediction for enabling spatial scalable coding [12]

3. ADAPTIVE RESOLUTION CHANGE USING SHVC

As described earlier, changing the resolution of the video dynamically within a video conferencing session is useful for various use cases. However, conventionally this can only be done by sending an IDR picture when the resolution changes, as single-layer video coding standards do not include mechanisms to predict from reference pictures having different resolution than that of the current picture being coded or decoded. This causes the following problems:

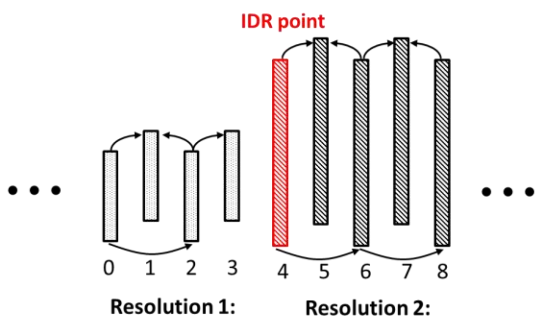
- As IDR pictures take significantly more bits compared to inter coded pictures, the bitrate increases suddenly at the IDR picture, which increases the end-to-end delay of the video conferencing session.
- If it is required to keep the delay small and hence use a similar amount of bits for the IDR pictures as the bit count for the inter-predicted pictures, the IDR pictures will have a significantly lower visual quality than that of the inter-predicted pictures. This will cause a blip in quality in the video conferencing session and reduce the user experience.

In this paper, we propose to utilize the draft scalable H.265/HEVC (SHVC) standard for supporting adaptive resolution applications efficiently. The proposed method utilizes inter-layer prediction from the base layer to the enhancement layer at the resolution switching point. Because inter-layer redundancies are exploited, a smaller number of bits is used compared to the bit count of the IDR pictures having the same quality and coding efficiency is therefore increased.

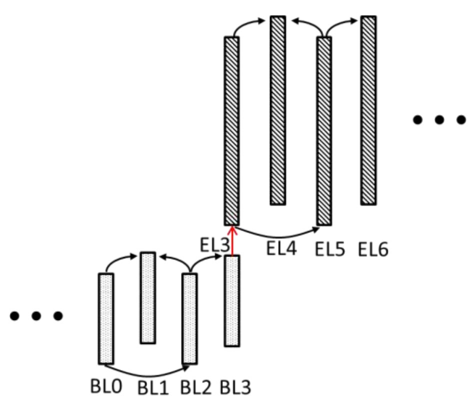
In order to support adaptively changing resolution with SHVC, it is proposed that the following two indications are included in SHVC bitstreams. Firstly, it needs to be indicated that inter-layer prediction is used only for resolution switching purposes, but not for scalability purposes. In addition, it needs to be indicated that each access unit in the sequence, except for the resolution switch points, contains only a single picture from a single layer (which may or may not be a base-layer picture).

Indicating these restrictions allows the decoder to know that scalability is not used except for achieving resolution change, so that the decoder resources can be allocated accordingly upon session start. In addition, an SHVC decoder needs to know that some access units do not necessarily contain base layer pictures. This is useful for the following case. Let's assume an SHVC decoder has the capability to decode a lower resolution bitstream (i.e. 720p). If adaptive resolution switching is used and switching happens from 720p to 1080p, there should be a mechanism to indicate the decoder that it won't be able to decode and display all the pictures.

Figure 2 (a) and (b) shows the adaptive resolution change approach using the IDR picture and the proposed method with scalable coding respectively. In this Figure, switching happens at picture 3. Hence, using IDR picture for resolution switching, as shown in Figure 2 (a), the decoder receives the bitstream with pictures 0, 1, 2 and 3 of Resolution 1 following by pictures 4, 5, 6,…. of Resolution 2. But in the case of proposed method, as shown in Figure 2 (b), the following bitstream is transmitted by the encoder: BL0-BL1-BL2-BL3-EL3-EL4-EL5-EL6 and so on. Because no IDR picture is transmitted, there is no immediate bitrate increase in the proposed method, hence improving coding efficiency of the system.



(a)



(b)

Figure 2. Resolution switching using (a) IDR picture and (b) our proposed method with scalable coding

4. EXPERIMENTAL RESULTS

In order to measure the coding efficiency improvements of the proposed method for adaptive resolution change use cases, several experiments were performed, both using the proposed method and using IDR pictures for resolution switching. For this purpose, several video sequences were encoded using the scalable H.265/HEVC software version 0.1.1 [10]. Encoding options are based on common conditions described in [13] with some modifications. Since the primary use of dynamically changing the resolution is low-delay applications such as video-conferencing, low-delay IPP coding configuration with one reference picture was used to simulate a common configuration of these applications. Table 1 summarizes the properties of our test sequences.

Table 1. Properties of the Test Sequences

Sequence	Original Resolution	Number
Kimono	1920x1080	240
ParkScene	1920x1080	240
Cactus	1920x1080	500
BasketballDrive	1920x1080	500
BQTerrace	1920x1080	600

Figure 3 illustrates the number of bits used for each picture for both the proposed method and the anchor where the switching happens from 720p (1280x720) resolution to 1080p (1920x1080) resolution at the middle of the sequence. For the proposed method, the bit count used at the switching point was calculated using the base layer picture at low resolution and enhancement layer picture at high resolution (i.e. BL3+EL3 in Figure 2.b). For the anchor case, where the switching happens by transmitting an IDR picture, the number of bits used for switching point is the number of bits used for the IDR picture. As shown in Figure 3, the proposed method reduced the bitrate at the switching point for the Parkscene sequence by 64%.

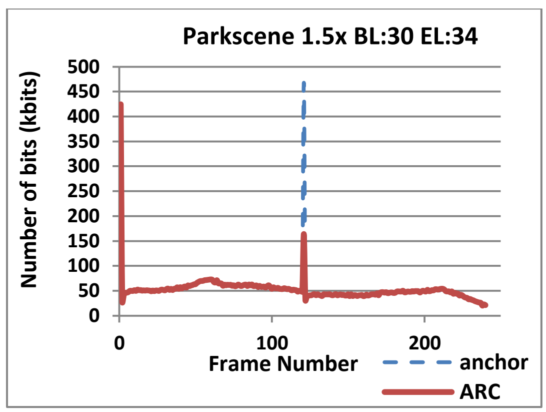


Figure 3. Comparison of bitrate increase at switching point for proposed method and anchor (using IDR picture) for Parkscene.

The coding efficiency gains of the proposed method were measured extensively for different sequences coded at different quantization parameters (QPs) for the base and enhancement resolutions. The reason to test different quantization parameters is that the QP could be different for high resolution video, as there could be different quality optimizations depending on the content type. In Figure 4, the additional bitrate required at switching point for both the proposed method and anchor are shown for Kimono sequence for different QP's. As seen in this Figure the bitrate at the switching point is reduced using the proposed method. Indeed, due to better inter-layer prediction at higher base layer quality, the bitrate reduction is more visible when the difference between the QPs of base and enhancement layers is increased.

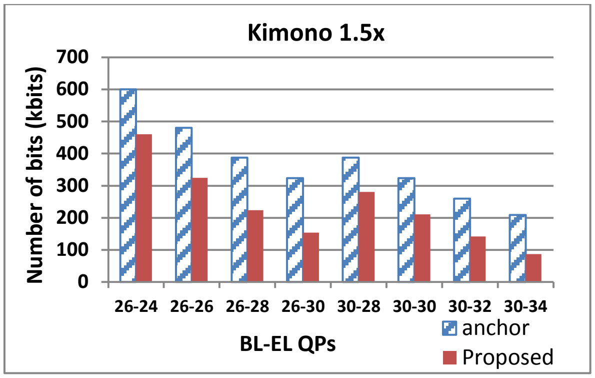


Figure 4. Bitrate required at switching point for the proposed method and anchor case for Kimono sequence with different QP settings

Table 2 shows the average bitrate increase at switching picture both for anchor and proposed method for different QPs for the smaller resolutions. The proposed method used approximately half the bitrate compared to the anchor method for switching pictures. For sequences that contain a lot of high frequency content (such as BQTerrace), the gains of the proposed method are reduced. Supposedly, in such type of sequences, the redundancy between layers is reduced and inter-layer prediction is not as efficient when compared to sequences with less high frequency content.

The gain of the proposed method was further analyzed for different delta QPs (i.e., different QP value differences between the enhancement and the base layer) in

Table 3, where the average percentage bitrate increase at switching point with different delta QPs (-2, 0, 2, 4) is shown for anchor and proposed method. As seen in this table, the bitrate at the switching point is reduced significantly using the proposed method, and the gains become more visible when the QP difference is increased for high resolution pictures. This is due to the fact that when base layer quality is relatively high, inter-layer prediction improves and consequently the coding efficiency of the proposed method is also improved.

Table 2. Average increase in bitrate at switching point for anchor (using IDR picture) and proposed method for base layer quality.

Sequence	BL QP	Anchor	Proposed
Kimono	26	280%	134.8%
	30	331.8%	148.3%
ParkScene	26	1052.1%	592.3%
	30	1353.3%	727.9%
Cactus	26	660.8%	374.8%
	30	874.6%	466.3%
BasketballDrive	26	322.1%	237.6%
	30	356%	246.1%
BQTerrace	26	496.5%	325.6%
	30	936.1%	555.1%

Table 3. Average increase in bitrate at switching point for anchor (using IDR picture) and proposed method with different **ΔQPs** for all tested sequences

ΔQPs	Anchor	Proposed method
-2	653.4%	491.2%
0	673.7%	463.7%
2	681%	355.1%
4	628.6%	213.5%

5. CONCLUSION

In this paper, a novel method was presented to support adaptive resolution change effectively using the draft scalable H.265/HEVC standard. The proposed approach utilizes inter-layer prediction at pictures where switching to a higher resolution is needed, instead of using IDR pictures. The experimental results showed that by using the proposed method the bitrate at the switching point was reduced significantly. It was also shown that, when the same bitrate is used for switching pictures, the proposed method improved the visual quality.

(a)

(b)

Figure 5. The reconstructed picture at switching point for (a) anchor method using IDR picture and (b) the proposed method

REFERENCES

[1] T. Davies, "Resolution Switching for Coding Efficiency and Resilience", *Technical Report JCTVC-F158*, July 2011.

[2] T. Davies, P. Topiwala, "Adaptive Resolution Coding (ARC)", *Technical Report JCTVC-G264*, November, 2011.

[3] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, T. Wiegand (editors), "High Efficiency Video Coding (HEVC) text specification draft 10", *JCT-VC output document JCTVC-L1003*, March 2013.

[4] "Advanced Video Coding for Generic Audiovisual Services", *ITU-T Recommendation H.264*, January 2012.

[5] J. Chen, J. Boyce, Y. Ye, M. M. Hannuksela (editors), "SHVC Test Model 1 (SHM 1)", *JCT-VC output document JCTVC-L1007*, February 2013.

[6] J. Chen, J. Boyce, Y. Ye, M. M. Hannuksela (editors), "SHVC Working Draft 1", *JCT-VC output document JCTVC-L1008*, March 2013.

[7] K. Ugur, H. Roodaki, M. M. Hannuksela, "AHG9: Using SHVC for adaptive resolution change and efficient trick mode", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, *Technical Report JCTVC-M0040*, April 2013.

[8] G. J. Sullivan, J.-R. Ohm, F. Bossen, T. Wiegand, "JCT-VC AHG Report: HM Subjective Quality Investigation (AHG22)", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC 1/SC 29/WG 11, *Technical Report JCTVC-H0022*, February. 2012.

[9] G.J. Sullivan, J.R. Ohm, Woo-Jin Han, Woo-Jin Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649 - 1668, 2012.

[10]M. M. Hannuksela, K. Ugur, J. Lainema, D. Rusanovskyy, J. Chen, V. Seregin, Y. Wang, Y. Chen, L. Guo, M. Karczewicz, Y. Ye, J. Boyce, "Test Model for Scalable Extensions of High Efficiency Video Coding (HEVC)", *Technical Report JCTVC-L0453*, January 2013.

[11] M. M. Hannuksela, K. Ugur, J. Lainema, D. Rusanovskyy, J. Chen, V. Seregin, Y.-K. Wang, Y. Chen, L. Guo, M. Karczewicz, Y. Ye, J. Boyce, "Common Specification Text for Scalable and Multi-view Extensions", *Technical Report JCTVC-L0452,* January 2013.

[12] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103 – 1120, 2007.

[13] F. Bossen, "Common Test Conditions and Software Reference Configurations", *Technical Report JCTVC-E700*, January 2011.