Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

An Efficient Region-of-Interest (ROI) based Scalable Framework for Free Viewpoint Video Application

H. Roodaki *

Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran.

Article Info

Article History:
Received 24 August 2023
Reviewed 02 October 2023
Revised 06 November 2023
Accepted 12 December 2023

Keywords:
Tile-based scalability
Region of interest
λ-domain rate control algorithm
MV-HEVC
Parallel processing

*Corresponding Author's Email Address: *hroodaki@kntu.ac.ir*

Abstract

Background and Objectives: From the multiview recorded video, free viewpoint video provides flexible viewpoint navigation. Thus, a lot of views need to be sent to the receivers in an encoded format. The scalable nature of the coded bitstream is one method of lowering the volume of data. However, adhering to the limitations of the free viewpoint application heavily relies on the kind of scalable modality chosen. The perceptual quality of the received sequences and the efficiency of the compression technique are significantly impacted by the scalable modality that was chosen.

Methods: In order to address the primary issues with free-viewpoint video, such as high bandwidth requirements and computational complexity, this paper suggests a scalable framework. The two components of the suggested framework are as follows: 1) introducing appropriate scalable modality and data assignment to the base and enhancement layers; and 2) bit budget allocation to the base and enhancement layers using a rate control algorithm. In our novel scalable modality, termed Tile-based scalability, the idea of Region of Interest (ROI) is employed, and the region of interest is extracted using the tile coding concept first presented in the MV-HEVC.

Results: When compared to the state-of-the-art techniques, our approach's computational complexity can be reduced by an average of 44% thanks to the concept of tile-coding with parallel processing capabilities. Furthermore, in comparison to standard MV-HEVC, our suggested rate control achieves an average 17.7 reduction in bandwidth and 1.2 improvement in video quality in the Bjøntegaard-Bitrate and Bjøntegaard-PSNR scales.

Conclusion: Using new tile-based scalability, a novel scalable framework for free-viewpoint video applications is proposed. It assigns appropriate regions to the base and enhancement layers based on the unique features of free viewpoint scalability. Next, a rate control strategy is put forth to allocate a suitable bitrate to both the base and enhancement layers. According to experimental results, the suggested method can achieve a good coding efficiency with significantly less computational complexity than state-of-the-art techniques that used the λ-domain rate control method.

Introduction

Free viewpoint video is a system allowing the users to observe the scene from various view point and freely change the views. It can be used in multiple applications such as immersive teleconference, 3DTV, and sporting

events, enabling the viewers to move freely around the scene. But, to provide this capability, some main challenges must be overcome.

- **High Bandwidth:** To provide high quality and rich interactive experience, the resolution of the free viewpoint video should be 4K or even higher. However, streaming numerous views at such resolution requires extraordinary bandwidth, which is not reachable in low bit rate communication channels such as the wireless mobile networks. Even with the compression offered by Multiview Video Coding (MVC), which uses Intra and Inter-view prediction to extract statistical dependencies between views, free viewpoint video is currently beyond the capabilities of most wireless networks. Intra-view prediction uses temporally adjacent frames, and inter-view prediction uses corresponding frames in adjacent views as reference views in the prediction process.
- **Computational Complexity:** As mentioned before, an essential characteristic of free viewpoint video is increasing the control over view angle and direction, according to the perceptual preferences of users. A secondary effect of increasing the viewing controllability is higher computational complexity. Since, there are numerous numbers of views that should be encoded and send to the viewers, while the human eye can only focus on a particular area of the scene and small numbers of view at any time. Hence, the main challenge in this application is controlling the high computational complexity of encoder side [1].

Scalable Multiview Video Coding (SMVC) is one of the main techniques to address the high bandwidth requirement and reduce computational complexity by scaling down the video. Usually, a scalable bitstream consists of a “Base” layer that carries the minimum amount of video data necessary for all receivers. Then, one or more “Enhancement” layers can be built on top of the base layer to improve the overall video perceptual quality.

In this paper, we propose an efficient scalable framework for free viewpoint video applications to overcome the mentioned challenges. Our proposed framework has two distinct parts:

1) Selecting proper scalable modality and data assignment to base and enhancement layers: Selecting the proper scalable modality can help satisfying the main requirements of free-viewpoint application such as the required bandwidth. In particular, the selected scalable modality can affect the perceived quality and the compression efficiency.

To clarify the meaning of scalable modality, suppose that, a mobile client has an available bandwidth that is less than the bitrate of the original free viewpoint video. A video adaption should be performed to extract base and enhancement layers to match the bitrate of the adapted video to the bandwidth of the mobile client. This type of video adaptation is named scalable modality. It is more efficient that the scalable modality is extracted according to the specific characteristics of the application at hand, such as free viewpoint application. After selecting a proper scalable modality, the available data should be assigned to various layers according to this particular scalable modality. In the free viewpoint application, usually, the receivers are more interested in specific regions of the scene, and the perceptual quality of those regions is more important to them. This concept is typically referred to as Region of Interest (ROI). Our proposed framework in this paper proposes to use the concept of ROI to define the appropriate scalable modality for free viewpoint application. The user's favorite regions, ROIs, are located in the base layer and the other areas are located in enhancement layers.

But, extracting the ROI is not so trivial. Several methods are presented in the literature to extract ROI based on some specific features such as texture perceptual map and the motion perceptual map [2], visual attention model, and gaze tracking data [3] and so on.

In this paper, we suggest using the idea of tile coding in the MV-HEVC standard to extract the ROI and data assignment to the base and enhancement layers.

In MV-HEVC standard, the pictures can be divided into independently decodable rectangular regions with approximately equal numbers of CTUs that are entitled as tiles. The main goal of partitioning the frames into tiles is to increase the capability for parallel processing and provide error resilience [4].

Using this concept, we can code the tiles corresponding to ROI as the base layer and the remaining ones as the enhancement layers. The tiles can be coded and decoded in parallel to improve the computational complexity of the encoder and decodes side.

2) Rate control algorithm for bit budget allocation to base and enhancement layers: For the second part of the proposed framework, we propose a rate control algorithm for scalable coding in free viewpoint application. We have used the λ-domain rate-distortion model [5] in the HEVC video coding standard as our reference rate-distortion model. Then we will use specific features of the Tile-based scalable modality to extract coding parameters for free viewpoint video application efficiently.

The main innovations of the paper are as follows:

1. Introduce a proper scalable modality for free-viewpoint video application based on Region of Interest, named Tile-Based scalability.
   - In tile-based scalability, the concept of tile coding in the MV-HEVC video coding standard is used to extract the interested and non-interested regions

and allocate them to the base and enhancement layers, respectively.
- In addition, using the concept of tile coding in the MV-HEVC video coding standard, we can benefit from parallel processing to reduce the complexity of the encoder and improve efficiency.

2. Propose a rate control algorithm for rate assignment to base and enhancement layers for Tile-Based scalability in free-viewpoint video application.
   - The concept of inter-view disparity is used to find the appropriate relationship between the quantization parameters of various tiles

Related Work

In this section, the scalable coding for multiview video and free viewpoint application and the rate control algorithms for scalable coding are reviewed.

A. Scalable Coding in Multiview Video and Free Viewpoint Application

Several scalable modalities have been mentioned in the literature for single view and multiview video. For instance, temporal, spatial, and quality scalability and various combinations of them [6] and Region-Of-Interest (ROI) and object-based scalability [7] are introduced for single view video coding. Besides, view scalability [8] and free viewpoint scalability [9] are presented for multiview video. In [10] an efficient scalable multiview video coding method is proposed in which high-quality depth maps are coded as a piece of scene information. Then, the view-dependent depth map is generated from this information at the decoder side. So the free viewpoint scalability and coarse granular SNR scalability are achieved using these synthesized depth maps. In [11], a scalable approach is presented for immersive video streaming to support different receivers. This method limits the number of views in the base layer and uses view scalability and free viewpoint scalability in the enhancement layers to synthesize more views at the receiver side to improve the quality of free viewpoint views for the user. In [12] an encoding configuration for scalable multiview video coding is proposed that realizes higher compression efficiency and provides view switching for the users. In the proposed approach, the base layer uses inter-view prediction and produces a video sequence with the acceptable quality. Then, the enhancement layers use the corresponding base layer without interview prediction.

As we mentioned before, complying the constraint of free viewpoint application is strongly dependent on the type of selected scalable modality that has a significant effect on the effectiveness of the compression method and perceptual quality of received sequences. The above mentioned scalable modalities are not fitted precisely to the particular characteristics of free viewpoint application. Hence, in this paper, we propose a new scalable modality according to the specific features of this application.

B. Rate Control Algorithms in Multiview Scalable Video Coding

In this section, the most recent rate control algorithms for multiview scalable video coding are presented. In [13] a ρ-domain rate control algorithm for multiview high efficiency video coding is proposed. First, the prediction structure of MV-HEVC is optimized. Then, the ρ domain rate control model based on multi-objective optimization is used. In this study, the image similarity is considered to analyze the correlation between viewpoints. Then, this correlation and the frame complexity are used for the rate allocation process. The method proposed in [14] uses the analysis of the characteristics of multi-view video coding and the requirements of its bit rate control to improve the traditional quadratic rate-distortion model. In [15], a bit allocation and rate control approach for multiview video coding is proposed that uses the frame complexity and human visual characteristics. These characteristics are used to improve the quadratic rate-distortion (R-D) model. Then, the proposed algorithm reasonably allocates bit-rate among views based on frame complexity and human visual characteristics. The bit allocation process among various views is done by solving a multi-objective optimization problem. In [16], a rate control algorithm for MV-HEVC based on scene detection is proposed. A ρ-domain rate control model based on multi-objective optimization that uses the image similarity is used. This image similarity is suggested to make a reasonable bit allocation among viewpoints. So, by switching the video scene, the image similarity is recalculated. In this paper, the frame layer rate control considers the layer B-frame and other factors in allocating the code rate. Then, the basic unit layer rate is done according to the content complexity of the CTU. In [17], a three levels quadratic rate-distortion model for multiview video coding is proposed. A λ-domain rate control algorithm for the scalable extension of HEVC video codec is presented in [18] that includes temporal, spatial, and quality scalability. The proposed algorithm introduces an initial target bits and encoding parameters determination algorithm for the first frame of each layer. Then, considering the inter frames, a bit allocation method is suggested using intra and inter layer dependencies. In [19] a rate control method is proposed that takes the human visual system into account to allocate the bit budget to various vision perceptual regions. The proposed model in [20] takes the QP values of B frames into account since some views in the multiview video are consisted of B frames only. In [21] a rate-distortion model is introduced that considers the dependency between frames made by motion compensation and depth image-based rendering. Finally,

[22] proposes a method for key frames that encouraged R-λ model using the depth map characteristics.

A novel method for rate assignment of free-viewpoint video is presented in [23] that uses the distance between view directions to allocate the appropriate rate for each view and provide a broader field of view.

In [24] rate control algorithm for MV-HEVC based on scene detection is presented. The motivation is that the rate control algorithm for multiview in (MV-HEVC) does not have the capability of bit allocation efficiently at the CTU level. So, the video quality varies greatly for sequences with sudden scene changes or large motions.

As we can see, most of the above mentioned methods are suggested for multiview video and cannot be extended to scalable multiview video easily. Also, besides considering the main features of scalable modality in the rate allocation process may lead to much better performance in the compression process. To the best of our knowledge, none of the presented rate allocation methods addresses this issue.

Proposed Method

In this section, we introduce our proposed scalable framework for free viewpoint application. Our proposed framework has two distinct steps, 1) introducing a new scalable modality for free viewpoint application and 2) presenting a rate allocation method to assign reasonable rate to base and enhancement layers.

A. Introduce A New Scalable Modality for Free Viewpoint Application

Free viewpoint video often uses in situations that want to give the viewers the higher coverage of Field of View. Hence, the number of cameras and the corresponding setup should provide the most coverage from the scene. But, it is usually more important to the viewers to see the most interesting parts of the scene than the events around the boundaries. For instance, a football game often relies on the movement of participants within a specific playing area. Where the interesting events will take place, are more important to the viewers. Missing these areas will impose a harmful effect on the user's quality of experience [25].

Our proposed scalable modality for free viewpoint application suggests using this criterion to assign data to the base and enhancement layers. The most interesting parts of the scene that are more important to the users are located in the base and the other parts are located in one or more enhancement layers. But, the main challenge is how to extract the interesting parts of the scene.

The proposed ROI selection methods are either feature-based or object-based. Feature-based methods find pixels that share significant optical features with the target and aggregate them to form ROIs [26], [27]. These methods can capture most of the target pixels based on the optical feature similarity. However, not all target pixels have strong optical features, so the detected ROI usually fails to encompass the entire target. In addition, feature-based methods cannot distinguish between targets, which can cause confusion in subsequent stages of processing.

Object-based methods, on the other hand, detect ROIs at a higher level than the pixel-by-pixel approach of feature-based systems using information such as target shape and structure [28], [29]. Typical approaches include template matching and matched filters. Although these methods can assign a single ROI to one target, they are limited because they require many calculations, have difficulty detecting multiple target types, and are not reliable when applied to low-quality images.

According to the above discussion, finding the ROI using feature-based and object-based methods is computationally complex. In these approaches, various parts of the scene should be traversed pixel-by-pixel or some other information such as shape and structures should be considered.

The concept of tile in MV-HEVC offers an alternative partitioning of a picture into rectangular parts that are encoded and decoded independently of another tile [26]-[30]. The main goal of using tiles coding is to enable the use of parallel processing for the encoding and decoding process [4]. In addition, tile coding facilitates video coding based on ROI. For instance, the tile containing the ROI can be extracted quickly and processed more efficiently [30].

We have used this concept in our proposed framework to introduce a new scalable modality for free viewpoint video, named "Tile-based Scalability". In the proposed scalable modality, in various views, the tiles corresponding to the ROI will be coded independently as the base layer and the other tiles of views corresponding to non-interesting parts are encoded as one or more enhancement layers. This way, selecting the tiles related to the region of interest is simple, since it is not required to detect specific objects or areas. Just finding the approximate coordinates of the desired area is enough. So, this method is much less complex than the feature-based and object-based ROI detection methods. As discussed, this new scalable modality is precisely matched to the free viewpoint application and its requirements.

As discussed before, in free view-point application, numerous numbers of views should be encoded and send to the viewers that increase the computational complexity of the encoder side. In addition, MV-HEVC has developed from previous standards adding some advanced features to increase compression ratios. This higher coding efficiency is obtained at the expense of substantial growth in computational complexity [31]. Besides, the motion compensation and loop filtering functions are the most complex functions at the decoder

side [32].

The suggested method to overcome the high complexity is supporting parallelism at the encoders and the decoders using tile coding [33]. Motion vector prediction, intra prediction, entropy coding, and reconstruction dependencies are not allowed across a tile boundary [33].

So the tiles can be coded end decoded independently from each other and the encoder or the decoder can process a tile in parallel with the other ones.

In tile-based scalability, we generalize this concept to scalable video coding and the tiles of each layer should be encoded and decoded independently and without using the tiles of the other layers. For instance, suppose that our free viewpoint video has three views, each of them has three tiles as shown in Fig. 2.

Assume that, considering the region of interest in this video, we decided to allocate three middle tiles to the base and the other six tiles to enhancement layers1 and 2, respectively.

Hence, the tiles of the base layer, Tile #12, Tile #22, and Tile #32 should be coded from each other using inter-view coding. Tile #11, Tile #21, and Tile #31 form enhancement layer 1 also can use inter-view prediction for much efficient compression. The tiles of different layers, for instance, the Tile #12 from the base layer and Tile #11 of the enhancement layer 1, cannot use for inter-view prediction. Coding the tiles of each layer independently from the other layer can lead to parallel processing of the tiles of each layer that can improve the computational complexity of the encoder and decoder side.

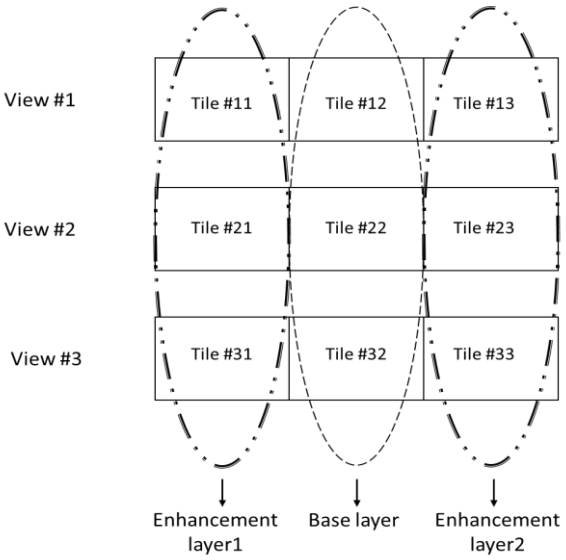

Fig. 2: Tiles allocation to base and enhancement layers in Tile-based scalability in free viewpoint video.

B. Rate Control Algorithms for Tile-Based Scalability in Free Viewpoint Application

Efficient rate allocation to the tile-based scalable video should consist of two steps as shown in Fig. 2. First, assigning the required bitrate to base and enhancement layers according to its importance and then, giving efficient rate to the corresponding tiles of each layer. Our proposed approach uses the main features of tile-based scalability to allocate the proper bitrate to various layers and the corresponding tiles efficiently, as explain in the following sub-sections.

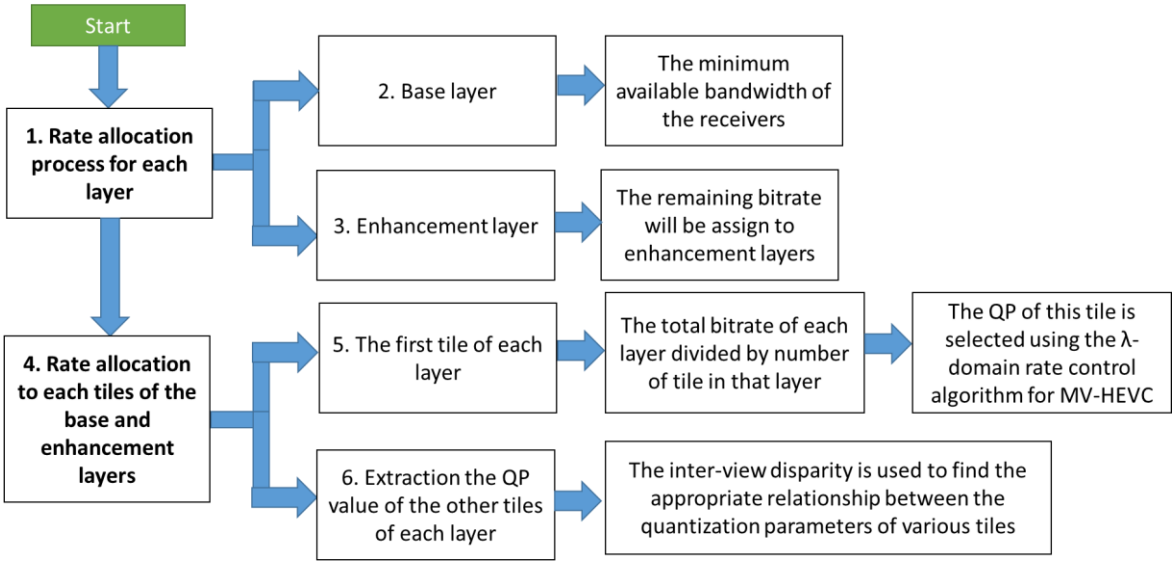

Fig. 1: The block diagram of the proposed method for rate control algorithms.

I) Rate Allocation for Each Layer

For the first step, rate assignment is performed considering the main characteristic of tile-based scalability and free viewpoint application.

In tile-based scalability, the base layer contains the minimum number of required tiles according to the minimum bandwidth of all receivers. Then each specific receiver can request one or more enhancement layers to increase the number of received tiles and to cover the required viewing angle according to its additional bandwidth. Hence, the total bitrate of the base layer should be selected according to the minimum available bandwidth of the receivers. Since, all the receivers should be able to receive this layer. For instance, when a scene is being captured by multiple cameras and should be sent to a mobile phone with limited bandwidth and also a portable tablet with more resources, then, the total bitrate of the base layer should be selected according to the available bandwidth of the mobile phone. Then, the remaining bitrate will be assigned to enhancement layers.

II) Rate Allocation to Each Tiles of the Base and Enhancement Layers

In order to find the bitrate of tiles in each layer, we will use the following steps.

First, an initial bitrate is chosen for the first tile of each layer. This initial bitrate can be considered as the total bitrate of each layer divided by the number of tiles in that layer. Then, the quantization parameter (QP) of this tile is selected using the λ-domain rate control algorithm for MV-HEVC [34].

In λ-domain, the λ parameter is defined as the slope of the R-D curve, which can be expressed as [5]:

$$\lambda = -\frac{\partial D}{\partial R} \triangleq \alpha R^{\beta} \quad (1)$$

Hence,

$$R = \left(\frac{\lambda}{\alpha}\right)^{\frac{1}{\beta}} \quad (2)$$

where α and β are parameters related to the video content and using some pre-encoded video, these parameters values can be extracted via the fitted R –λ curve.

According to (2), the rate-distortion analysis can be carried out in the λ domain and the λ can be determined according to the target bitrate. We have used the initial bitrate for the first tile to extract the λ parameter. Then, the other coding parameters such as QP can be determined using the following equation as suggested in [5].

$$Qp = c1 \times \ln(\lambda) + c2 \quad (3)$$

where c1 and c2 are set to 4.2005 and 13.7122, respectively [34]. The extracted QP value from (3) should be rounded to the nearest integer value [5].

The benefit of using the λ-domain model is that, λ is the slope of the operational convex rate-distortion curve. So, there is a one-to-one correspondence between rate value and λ. Consequently, the λ parameter can be determined in a specific range according to the target bitrate, without interference from the rest of the coding parameters. In addition, adjusting λ parameter can be precise enough since it can take any continuous positive value. Finally, since λ is a parameter that is not needed by the decoder, it shouldn't be sent to the receiver side. Therefore, the higher precision of λ will not raise the bitrate [5].

After specifying the QP value using initial bitrate and (2) and (3), the QP value of the other tiles of each layer should be extracted. Since the various tiles of each layer are captured from a common scene from various viewpoints, they may be similar to each other. Our proposed approach uses this similarity in order to select the QP value of the other tiles of each layer. For this purpose, we generalize the approach suggested for view scalability in [35] to tile-based scalability as follows. As we explained in the features of tile-base scalability, the tiles inside each layer can be predicted from each other using inter-view prediction.

Hence, if the disparity between the first and second tiles of each view, i.e., Tile #12, and Tile #22 in Fig. 2, is low, the Tile #22 can be predicted better from Tile #21 and consequently can be compressed more efficiently. So the bitrate of this tile can be much lower without affecting the overall quality. Hence, the inter-view disparities among various tiles of each layer have a direct impact on prediction efficiency and the total rate of tiles inside the layer.

As such, we generalize the method proposed in [35] and use the concept of inter-view disparity to find the appropriate relationship between the quantization parameters of various tiles using the following equation:

$$QP_{Tn} = QP_{T0} + \frac{1}{inter{-}view\ disparity\ between\ Tile\#0\ and\ Tile\#n} \quad (4)$$

where QP_{T0} is the QP value of the first tile in each layer that is extracted from the previous step and QP_{Tn} is the QP value of tile #n in that layer.

Results and Discussion

In this section, several experiments were performed to show the efficiency of our proposed scalable framework. The results show the effectiveness of the proposed tile-based scalability in terms of computational complexity and the effectiveness of our method in allocating reasonable rate to each base and enhancement layer and their corresponding tiles. Results have been obtained using the MV-HEVC reference software [36]. Table 1 summarizes the properties of our test sequences [37]-[39].

Table 1: Properties of the sequences

Sequence	Original Resolution	Number of Views
Lincoln	2048 × 2048	107
Pantomime	1280 × 960	80
Tower	1280 × 960	80
Vassar	640 × 480	7

For each sequence, we have used three views for our experiments. Each view is partitioned into two or three different tiles. Then, we have considered one base layer and one or two corresponding enhancement layers. Table 2 shows the detail of layers and tiles selection for our test sequences. For instance, for the Lincoln sequence, we have used three views for our experiments, view#00, view#40, and view#60. Each view partitioned to three different tiles, the left tile, the middle tile, and the right tile. Then, we have considered one base layer and two corresponding enhancement layers. The middle tiles are placed in the base layer, and the other ones are placed in enhancement layers, respectively.

It should be noted that, we tried to have selected the views as far away as possible in each video sequence to cover the wide-ranging field of view at the receiver side. Then, the base and enhancement layers are selected based on the ROI parts of the video. The more attractive parts are considered as the base layer.

For instance, if the important objects of the scene are located in the right part of the scene, the tiles of this part are selected as the base layer. Fig. 3 shows the base and enhancement layers of the Tower sequence for more clarification. As we can see, the most important part of the scene is assigned to the base layer and the other parts are assigned to the enhancement layer.

Table 2: Base and enhancement layers and tiles selection for our test sequences

Seq.	View#	Number of layers	Base layer tiles	Enh. layer tiles
Lincoln	3	View#00 View#40 View#60	Middle tiles	Left tiles Right tiles
Pantomime	3	View#00 View#30 View#50	Left tiles	Right tiles
Tower	3	View#30 View#35 View#55	Right tiles	Left tiles
Vassar	3	View#00 View#04 View#07	Right tiles	Left tiles

We also have used the Bjøntegaard-Delta bitrate (BD-bitrate) and Bjøntegaard-PSNR (BD-PSNR) as the measure for Rate-Distortion (RD) performance [40]. BD-bitrate indicates the average difference in bitrate for the same the quality evaluation in PSNR, and BD-PSNR shows the average PSNR difference in dB over the whole range of bitrates.

Fig. 3: The enhancement (left) and base (right) layers of Tower sequence in Tile-based scalability.

In the reported results, negative BD-bitrate means bitrate savings, and positive BD-PSNR means the bitrate increase compared to the anchor case and positive BD-PSNR means the quality improvement over the anchor case.

For the anchor case, we have used the “current” methods that use the λ-domain rate control [5] to find the QP values for all layers and corresponding tiles.

The anchor method extracts the proper QP values of each tile of base and enhancement layers using λ-domain rate control method as follows. The proper QP value for each tile should be extracted using (3). Hence, first, the λ parameter for each tile should be exploited using (2) and the total bitrate of each tile. As discussed before, the total bitrate of the base layer should be assigned according to the minimum bandwidth requirements of all receivers. Since the base layer is the most important part of the scene, the highest portion of the total bitrate should be assigned to this layer. Without loss of generality, we assume that the 2/3 and 1/3 of the total bitrate are assigned to the base and the enhancement layers, respectively.

We have considered four initial QP values, 30, 25, 20, and 15 and encode all of our test sequences using these QP values. Then, 2/3 and 1/3 of the total extracted bitrate are assigned to base and enhancement layers, as suggested.

Since, we have encoded three views of each video sequence, each base and enhancement layer has three tiles. So, the bitrate of each tile is the total bitrate of each layer divided by 3.

In addition, in order to use (2), the α and β parameters

should be exploited. As we have discussed before, these parameters are related to the content of the video. We pre-encoded our test sequences with four different initial QP values, 30, 25, 20, and 15. Then we have used these initial QP values and the total extracted bitrate in (2) to find α and β parameters.

Finally, using the extracted α and β parameters and the bitrate of each tile, the λ value of each tile can be extracted and this λ value can be used in (3) to find the proper QP of each layer.

For our proposed approach, we have used the λ-domain rate control method to find the proper QP value just for the first tile of each layer. The QP values of the other tiles are extracted using the concept of inter-view disparity as explained before. To calculate the inter-view disparity between two tiles, a step which is needed for QP value extraction in (4), we have used the full search approach to estimate disparity between the tiles, accurately.

Table 3 shows the inter-view disparity values for our test sequences measured by this method. As mentioned before, we have considered two enhancement layers for Lincoln sequence and just one enhancement layer for the other ones.

Table 3: The inter-view disparity between Tile0 and the other tiles of each layer for our test sequences

Seq.	Base layer		Enh. layer1		Enh. Layer2	
Lincoln	1.61	1.022	0.60	0.55	0.55	3.06
Pantomime	18.05	18.10	0.09	0.04		
Tower	0.61	5.22	0.48	21.29		
Vassar	0.26	0.52	1	0.65		

We have calculated the QP values of each tile in base and enhancement layers according to (4) and using the mentioned inter-view disparity estimation. The corresponding extracted QP values for the anchor and the proposed method for initial QP=15 is illustrated in Table 4. As the same way, the QP are exploited for anchor and proposed method for initial QP, 20, 25 and 30.

At the next step, each tile of the base and enhancement layers has been coded by the extracted QP values for the anchor and the proposed method using MV-HEVC.

Table 5 provides the coding performance analysis of the proposed approach against the anchor method, where the negative BD-bitrate means bitrate savings compared to the anchor method. As we can see, our method can achieve mush better performance compared to the anchor method in terms of bitrate saving and quality.

Table 4: The extracted QP values for anchor and proposed method for our test sequences for initial QP=15

	Seq.		Lincoln	Pantomime	Tower	Vassar
			QP			
Anchor Method	Base Layer	Tile 0	14	15	15	15
		Tile 1	15	15	15	15
		Tile 2	15	15	15	15
	Enh. Layer1	Tile 0	16	15	15	16
		Tile 1	16	15	15	16
		Tile2	16	15	15	16
	Enh. Layer2	Tile 0	16			
		Tile 1	16			
		Tile 2	16			
Proposed method	Base Layer	Tile 0	14	15	15	15
		Tile 1	15	15	17	20
		Tile 2	15	15	15	17
	Enh. Layer1	Tile 0	16	15	15	16
		Tile 1	18	26	17	51
		Tile2	18	44	15	18
	Enh. Layer2	Tile 0	16			
		Tile 1	18			
		Tile 2	16			

Table 5: Coding performance comparison of the proposed Method against anchor method

Seq.	Base layer		Enh. Layers		Base + Enh. Layers	
	BD-rate	BD-PSNR	BD-rate	BD-PSNR	BD-rate	BD-PSNR
Lincoln	-0.1	0.1	-0.3	0.1	0.2	-0.1
Pantomime	0	0	-30.2	3.7	-26.4	2.3
Tower	-22.7	1.03	-2.7	0.2	-22.6	1.3
Vassar	-1.2	0.3	-24.3	1.83	-21.8	1.3
Avg.	-6	0.3	-14.4	1.5	-17.7	1.2

The corresponding RD curves for the total base and enhancement layers bitrate and PSNR for the Vassar sequence and for the anchor and the proposed method are illustrated in Fig. 4 for visual clarification. As seen, the bitrate is reduced significantly using our proposed method.

We have compared the results of the proposed method with papers as follows.

In [1] a regional bit allocation and rate-distortion optimization method for multiview with depth video coding is proposed that allocate more bits color texture area corresponding depth region and fewer bits to the

color smooth area corresponding depth region. The results of Table 6 compare the coding performance of our proposed method and the method proposed in [41] over the same anchor method where negative BD-bitrate means bitrate savings with respect to the anchor method. As we can see, our method can achieve much better performance compared to the anchor method in terms of bitrate saving and quality.

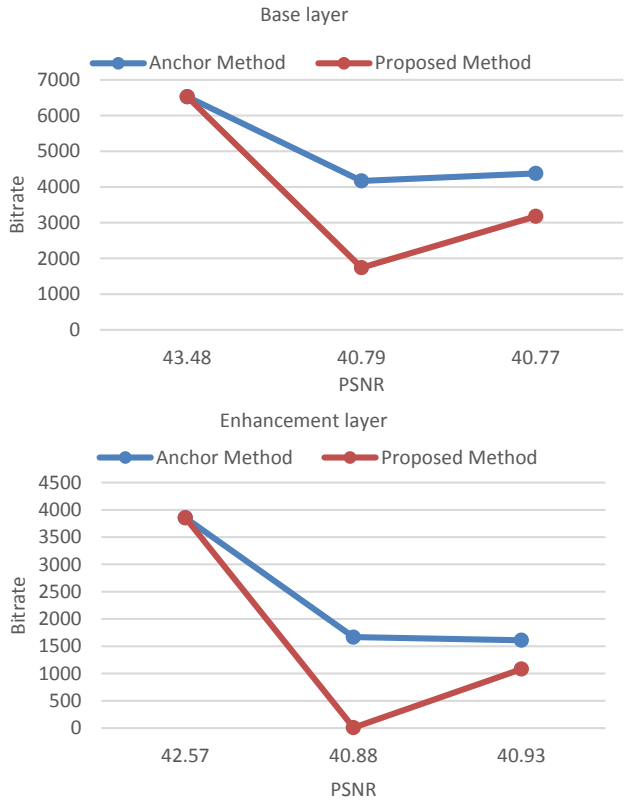

Fig. 4: Rate-Distortion curves for base and enhancement layers for anchor and proposed method of Vassar sequence.

Table 6: The comparison of coding performance of our proposed method against the method proposed in [41]

	Our proposed Method		The method proposed in [41]	
	BD-rate	BD-PSNR	BD-rate	BD-PSNR
Tower	-22.58%	1.35 dB	-19.78%	0.82 dB
Pantomime	-26.40%	2.29 dB	-21.98%	0.19 dB

In [42] a bit allocation optimization method for Multiview Video Coding (MVC) based on stereoscopic visual attention that exploits the visual redundancies from human perceptions. The results of Table 7 compare the coding performance of our proposed method and the method proposed in [2] over the same anchor method where negative BD-bitrate means bitrate savings and negative BD-PSNR shows the performance degradation with respect to the anchor method. The results show that our method can achieve much better performance compared to the anchor method in terms of bitrate saving and quality.

Table 7: The comparison of coding performance of our proposed method against the method proposed in [42]

	Our proposed Method		The method proposed in [42]	
	BD-rate	BD-PSNR	BD-rate	BD-PSNR
Tower	-8.58%	-0.40 dB	-19.78%	0.82 dB
Pantomime	-8.19%	-0.14 dB	-21.98%	0.19 dB

The main advantage of our proposed scalable modality over the previous ones is the low computational complexity of our method. In tile-based scalability, the tiles of base and enhancement layers can be coded independently from each other in parallel. It can improve the total computational time of our method over the previous scalable modalities such as view scalability [35].

We have implemented view scalability for our test sequences and coded the data of base end enhancement layers in both tile-based scalability and view scalability. Table 8 shows the processing time of our proposed algorithm compared to the method proposed in [35] which uses view scalability that cannot benefit from parallel processing. Both methods are run on an Intel i7-4790 CPU @ 3.6 GHz. As we can see, the processing time of our proposed that can uses parallel processing is much better.

Table 8: Comparison between the computational complexity of tile-based scalability and view scalability [35]

Sequences	Processing time of tile-base scalability (sec)	Processing time of view scalability (sec)
Lincoln	3240.267	6007
Pantomime	156.6	256.99
Tower	199.93	242.14
Vassar	50.3	103
Average	911.77	1652.28

Conclusion

In this paper, we presented a novel scalable framework for free-viewpoint video application. This framework proposes new tile-based scalability to assign proper regions to base and enhancement layers according to the features of free viewpoint scalability.

Then, a rate control approach is proposed to assign an appropriate bitrate to base and enhancement layers.

Besides, using tile coding may lead to reduce computational complexity. Experimental results showed that the proposed method could achieve a good coding efficiency over the anchor method that used the λ-domain rate control method with much less computational complexity.

Author Contributions

H. Roodaki, designed the experiments, collected the data, carried out the data analysis, interpreted the results, and wrote the manuscript.

Acknowledgment

The author received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

There is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

Abbreviations

3DTV	3D television
MVC	Multiview video coding
SMVC	Scalable multiview video coding
MV-HEVC	Multiview video-High efficiency video coding
QP	Quantization parameter

References

[1] C. C. Lee, A. Tabatabai, K. Tashiro, “Free viewpoint video (FVV) survey and future research direction,” APSIPA Trans. Signal Inf. Process., 4: 1-10, 2015.

[2] Y. Zheng, J. Feng, H. Ma, Y. Chen, “H.264 ROI coding based on visual perception,” in Proc. 5th International Conference on Visual Information Engineering. Xian China, China: 829-834, 2008.

[3] J. Zhang, L. Zhuo, Y. Zhao, “Region of interest detection based on visual perception model,” Int. J. Pattern Recognit. Artif. Intell., 26(02), 2012.

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

[5] B. Li, H. Li, L. Li, J. Zhang, ”λ domain rate control algorithm for high efficiency video coding,” IEEE Trans. Image Process., 23(9): 3841-3854, 2014.

[6] H. Schwarz, D. Marpe, T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. Circuits Syst. Video Technol., 17(9): 1103-1120, 2007.

[7] D. Grois, E. Kaminsky, O. Hadar, “Dynamically adjustable and scalable ROI video coding,” in Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB): 1-5, 2010.

[8] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, Y. Yashima, “View scalable multiview video coding using 3-D warping with depth map,” IEEE Trans. Circuits Syst. Video Technol. 17(11): 1485-1495, 2007.

[9] H. Yo Sung, O. Kwan Jung, “Overview of multi-view video coding. 14th International Workshop on Systems Signals and Image,” in Proc. 6th EURASIP Conference focused on Speech and Image Processing Multimedia Communications and Services, 2007.

[10] S. Shimizu, H. Kimata, K. Kamikura, Y. Yashima, “Free-viewpoint scalable multi-view video coding using panoramic mosaic depth maps,” in Proc. 16th European Signal Processing Conference, 2008.

[11] H. Roodaki, S. Shirmohammadi, “Scalable multiview video coding for immersive video streaming systems,” in Proc. Visual Communications and Image Processing (VCIP), 2016.

[12] N. Ozbek, A. Murat Tekalp, E. Turhan Tunali, “A new scalable multi-view video coding configuration for robust selective streaming of free-viewpoint TV,” in Proc. IEEE International Conference on Multimedia and Expo, Beijing: 1155-1158, 2007.

[13] T. Yan, I. H. Ra, Q. Zhang, H. Xu, L. Huang, “A novel rate control algorithm based on ρ model for multiview high efficiency video coding, “Electronics, 9(1): 166, 2020.

[14] T. Yan, I. H. Ra, D. Liu, D. Chen, Y. Youhao, S. Hou, “Rate control based on similarity analysis in multi-view video coding,” in Proc. 9th International Conference on Information Technology Convergence and Services (ITCSE 2020), 2020.

[15] T. Yan, I. H. Ra, Q. Zhang, H. Wen, H. Xu, S. Chen, “Rate control algorithm for multiview video coding based on human visual characteristics,” Int. J. Performability Eng., 14(8): 1913-1921, 2018.

[16] T. Yan, I. H. Ra, H. Wen, M. H. Weng, Q. Zhang, Y. Chen, “CTU layer rate control algorithm in scene change video for free-viewpoint video,” IEEE Access, 8: 24549-24560, 2020.

[17] T. Yan, P. An, L. Shen, Q. Zhang, Z. Zhang, “Rate control algorithm for multi-view video coding based on correlation analysis,” in Proc. Symposium on Photonics and Optoelectronics: 1-4, 2009.

[18] L. Li, B. Li, D. Liu, H. Li “λ-domain rate control algorithm for HEVC scalable extension,” IEEE Trans. Multimedia, 18(10): 2023-2039, 2016.

[19] P. J. Lee, Y. C. Lai, “Vision perceptual based rate control algorithm for multi-view video coding,” in Proc. International conference on system science and engineering (ICSSE), 2011.

[20] S. Park, D. Sim, “An efficient rate-control algorithm for multi-view video coding,” in Proc. IEEE 13th International Symposium on Consumer Electronics: 115-118, 2009.

[21] A. Fiengo, G. Chierchia, M. Cagnazzo, B. Pesquet-Popescu, “Convex optimization for frame-level rate allocation in MV-HEVC,” in Proc. IEEE International Conference on Image Processing (ICIP), 2016.

[22] M. Cordina, C.J. Debono, “A depth map rate control algorithm for HEVC Multi-View Video plus Depth,” in proc. IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2016.

[23] R. Abolfathi, H. Roodaki, S. Shirmohammadi, "A novel rate control method for free-viewpoint video in MV-HEVC," in Proc. 2019 International Conference on Computing, Networking and Communications (ICNC): 582-587, 2019.

[24] T. Yan, I. Ra, H. Wen, M. Weng, Q. Zhang, Y. Che, "CTU layer rate control algorithm in scene change video for free-viewpoint video," IEEE Access, 8: 24549-24560, 2020.

[25] J. Kilner, J. Starck, A. Hilton, “A comparative study of free-viewpoint video techniques for sports events,” in proc. 3rd European Conference on Visual Media Production, 2006.

[26] C. M. Privitera, L. W. Stark, “Algorithms for defining visual regions-of-interest: comparison with eye fixations,” IEEE Trans. Pattern Anal. Mach. Intell., 22(9): 970-982, 2000.

[27] J. L. Solka, D. J. Marchette, B. C. Wallet, V. L Irwin, G. W. Rogers, “Identification of man-made regions in unmanned aerial vehicle imagery and videos,” IEEE Trans. Pattern Anal. Mach. Intell., 20(8): 852-857, 1998.

[28] A. Mohan, C. Papageorgiou, T. Poggio, "Example-based object detection in images by components," IEEE Trans. Pattern Anal. Mach. Intell., 23(4): 349-361, 2001.

[29] T. M. Stough, C. E. Brodley, "Focusing attention on objects of interest using multiple matched filters," IEEE Trans. Image Process., 10(3): 419-426, 2001.

[30] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth, M. Zhou, "An overview of tiles in HEVC," IEEE J. Sel. Top. Signal Process., 7(6): 969-977, 2013.

[31] G. Corrêa, P. Assunção, L. Agostini, A. L. da Silva Cruz, "Performance and computational complexity assessment of high-efficiency video encoders," IEEE Trans. Circuits Syst. Video Technol., 22(12): 1899- 1909, 2012.

[32] M. Viitanen, J. Vanne, T. D. Hämäläinen, M. Gabbouj, J. Lainema, "Complexity analysis of next-generation HEVC decoder," in Proc. IEEE International Symposium on Circuits and Systems: 882-885, 2012.

[33] A. Fuldseth, M. Horowitz, S. Xu, K. Misra, A. Segall, M. Zhou, "Tiles for managing computational complexity of video encoding and decoding," in Proc. Picture Coding Symposium (PCS): 389-392, 2012.

[34] J. Wen, M. Fang, M. Tang, K. Wu, "R-(lambda) model based improved rate control for HEVC with Pre-Encoding," in Proc. Data Compression Conference (DCC): 53-62, 2015.

[35] H. Roodaki, M. R. Hashemi, S. Shirmohammadi, "Rate-distortion optimization for scalable multi-view video coding," in Proc. IEEE International Conference on Multimedia and Expo (ICME), 2014.

[36] G. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall, A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," IEEE J. Sel. Top. Signal Process., 7(6): 1001-1016, 2013.

[37] Tanimoto Laboratory, http://www.fujii.nuee.nagoya-u.ac.jp/~fukushima/mpegftv/, last ACCESS on. March, 2021.

[38] Merl, http://www.merl.com/pub/avetro/mvc-testseq/orig-yuv/vassar/, last ACCESS on March, 2021.

[39] ftp://ftp.research.microsoft.com/users/fvv/, last ACCESS on March, 2021.

[40] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," ITU T SG16/Q6, Doc. VCEG-M33, 2001.

[41] Y. Zhang, S. Kwong, L. Xu, S. Hu, G. Jiang , C. J. Kuo, "Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model," IEEE Trans. Image Process., 22(9): 3497-3512, 2013.

[42] Y. Zhang, G. Jiang, M. Yu, et al., "Stereoscopic visual attention-based regional bit allocation optimization for multiview video coding," EURASIP J. Adv. Signal Process. 848713, 2010.

Biographies

Hoda Roodaki was born in Tehran, Iran in 1982. She received the B.S. and M.S. degrees in Computer engineering from the University of Tehran, and Sharif University of Technology, in 2005 and 2007 and the Ph.D. degree in Computer Architecture from university of Tehran, Iran in 2014. Since 2015, she has been an Assistant Professor with the Computer Engineering Department, K. N. Toosi University, Tehran, Iran. Her research interests include Multi-view/3D and 360-degree video coding, Scalable video coding, video quality assessment and cloud gaming.

- Email: hroodaki@kntu.ac.ir
- ORCID: 0000-0002-3575-0587
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: https://wp.kntu.ac.ir/hroodaki/

How to cite this paper:
H. Roodaki, "An efficient Region-of-Interest (ROI) based scalable framework for free viewpoint video application," J. Electr. Comput. Eng. Innovations, 12(1): 283-293, 2024.

DOI: 10.22061/jecei.2023.7934.455

URL: https://jecei.sru.ac.ir/article_2017.html