

# Efficient Region-of-Interest Based Adaptive Bit Allocation for 3D-TV Video Transmission over Networks

Pham Thanh Nam, Vu Duy Khuong, Dinh Trieu Duong<sup>\*</sup>, Le Thanh Ha

*VNU University of Engineering and Technology, Hanoi, Vietnam*

---

## Abstract

Due to characteristics of human visual system (HVS), people usually focus more on a specific region named region-of-interest (ROI) of a video frame, rather than watch the whole frame. In addition, ROI-based video coding can also help to effectively reduce the number of encoding bitrates required for video transmission over networks, especially for the 3D-TV transmissions. Therefore, in this work, we propose a novel ROI-based bit allocation (BA) method which can adaptively extract and increase the visual quality of ROI while saving a huge number of encoding bitrates for video data. In the proposed method, we first detect and extract ROI based on the depth information obtained from 3D-TV video coding sequences. Then, based on the extracted ROI, a novel BA scheme is performed to solve the rate-distortion (R-D) optimization problem, in which the higher priority bitrates are adaptively assigned to ROI while the total encoding bitrates of video frames are kept satisfying all constraints required by the R-D optimization. Experimental results show that the proposed method provides much better higher peak signal-to-noise ratio (PSNR) as compared to other conventional BA methods.

Received 05 December 2015, revised 25 December 2015, accepted 31 December 2015

*Keywords:* ROI detection, Bit allocation, Rate-Distortion Optimization.

---

## 1. Introduction

BA or rate control (RC) are important schemes that help to deal with bitrate and compressed video quality fluctuations. Therefore, BA algorithms have been widely studied and proposed for efficient video transmission over networks [1]. This problem is also related to challenging issues such as resource optimization, computational complexity, and real-time video processing [2]. In this work, we consider BA for a specific class of applications, namely 3D television (3D-TV), in which one of the most interesting issues to focus on is the quality enhancement of ROI.

Relating to the ROI, several studies have shown that human eyes do not treat the content equally in a whole video frame, but usually

focus more on a specific region, ROI [3], [4]. Therefore, based on ROI and HVS, how to improve the performance of video coding has important theoretical and practical value. In [5], Hu *et al.* used a macroblock (MB) classification based on R-D characteristics to generate three kinds of ROIs (called basic units). Then, a weighted BA per region is performed with predetermined factors in heuristic ways. Lee and Bovik *et al.* [5] proposed to use an eye tracker to obtain the fixation points as ROI regions, for the earlier H.263 standard. However, it is impractical to have the eye tracker available during the video encoding process. Intuitively, the important cue for the perception model in conversational video coding is extracting faces as ROI regions. Then, a perceptual BA scheme [6] was proposed to reduce the quantization parameter (QP) values of skin regions.

---

\* Corresponding author. E-mail.: duongdt@vnu.edu.vn

Recently, 3D-TV has emerged as an attractive video coding framework for giving users more immersive experience by allowing users to view 3D scenes. 3D-TV is based on 3D-HEVC which is a standardized extensions of High efficiency video coding (HEVC) or H.265/HEVC standard [7]. Like HEVC, 3D-TV has eminent compression performance, much better than that based on the preceding H.264/AVC [8]. However, in order to meet the requirements of low bit-rate video transmission of 3D-TVs or mobile devices, 3D-HEVC still poses the great challenging problem of compression efficiency for HEVC. In fact, there still remains much perceptual redundancy in HEVC, since human attentions do not focus on the whole scene, but only a small region of ROIs. Therefore, ROI based BA scheme can be considered as a key solution to improve the coding efficiency for 3D-HEVC. Unfortunately, to our best knowledge, the existing BA approaches have yet to be sophisticatedly developed for the latest 3D-HEVC standard.

In [9], coding units (CUs) are classified referring to their depth in the quad tree and their coding type. Texture-based RC models for HEVC have been developed according to signal characteristics in different CU depths and coding types. In this method, the BA scheme for three types of CUs of different texture levels have been constructed to deal with more complex content and to ensure more accurate RC at the CU level. More efficient BA scheme applied for 3D-HEVC was proposed in [10] which is based on ROIs detection and extraction. In [10], Meddeb *et al.* proposed an approach to allocate a higher bitrate to the ROI while keeping the global bitrate close to the assigned target value. The ROIs, typically faces in this application, are automatically detected and each coding tree unit (CTU) is classified in a ROI map. This approach therefore can achieve high performance compared with that of BA applied for conventional H.264/AVC and provides an improvement in ROI quality. However, approaches mentioned above merely focus on color or texture information of video frames, and they do not take into account the depth information. In other words, since the characteristics of depth information introduced

in 3D-HEVC and the high correlations between depth and ROIs are not effectively employed in the previous schemes, the accuracy and effectiveness of ROI detection algorithm can be reduced in these schemes.

In this paper, we propose a novel ROI-based BA method (ROI-BA) which can adaptively extract and increase the visual quality of ROI while saving a huge number of encoding bitrates for video data. In the proposed ROI-BA method, we first detect and extract ROI based on the depth information obtained from 3D-TV video coding sequences. Then, based on the extracted ROI, a novel BA scheme is performed to solve the R-D optimization problem, in which the higher priority bitrates are adaptively assigned to ROI while the total encoding bitrates of video frames are kept satisfying all constraints required by the R-D optimization. Experimental results show that the proposed method can provide higher PSNR compared to other conventional methods.

The rest of this paper is organized as follows. Section 2 describes the proposed method in detail. Experimental results are discussed in section 3. Finally, section 4 concludes this paper.

## 2. Proposed method

Figure 1 shows a general 3D-TV video streaming framework of the proposed ROI-BA method. In Figure 1, input video frames consist of multiple color frames, associated depth maps, and corresponding camera parameters of each frame. The 3D-TV coder encodes input video frames into color and associated depth-map packets, respectively, and these packets are then transmitted over network paths. At the sender, based on the ROI and non-ROI regions extracted from color frames and the available bandwidth estimated for network paths, the proposed ROI-BA method performs an optimal BA algorithm to minimize total distortion achieved over the system. Then, at the receiver, video frames are reconstructed and finally fed into the 3D-TV decoder where they are decoded, virtual view synthesized, and displayed.

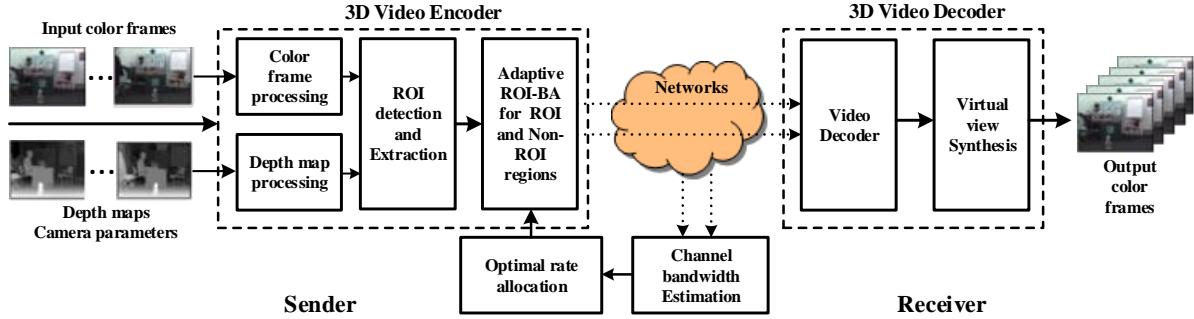


Figure 1. 3D-TV video coding using adaptive ROI-BA scheme.

## 2.1. Depth based ROI detection

Generally, in conventional methods, only texture information introduced in color video frames are employed to detect and extract ROI/Non-ROI regions. However, in our proposed method, we employ both texture and depth information to detect ROIs. Specifically, we propose to use the object detector algorithm (ODA) introduced in [11] for ROI detection. ODA is a famous algorithm and has been successfully applied for many applications performed on the colors frames for ROI detection such as text, faces, eyes detections, etc. In addition, to improve more on the accuracy of ROI detection for 3D-TV video frames, in our method, we also employ the high correlation between the ROI located in a color frame and its associated depth map.

Depth map is an 8-bit gray image that can be captured by depth camera or computed by stereo matching [12]. Each pixel in the depth map represents a relative distance between the video object and the camera. The depth data are usually stored as inverted real-world depth data  $d$ , according to

$$d(z) = \text{round} \left[ 255 \left( \frac{1}{z} - \frac{1}{z_{\max}} \right) / \left( \frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right) \right], \quad (1)$$

where  $z$  is the real-world depth value for the image,  $z_{\min}$  and  $z_{\max}$  are the minimum and the maximum values for  $z$ , respectively.

It is worth noticing that the ROI located in a color frame and its associated depth map are highly correlated, and two points belong to the same object in ROIs have the same or approximate depth values associated with them.

As illustrated in Figure 2, pixels  $d_1$  and  $d_2$  located in the region  $\Omega$ , which is the associated depth map of ROI region  $\Psi$ , have closed pixel values together and these values are quite different from pixel  $d_3$  which is not belong to region  $\Omega$ . Therefore, by determining exactly the region  $\Omega$  in the depth map,  $F_{Depth}$ , the mapped region  $\Psi$  of  $\Omega$  in the color frame,  $F_{Color}$ , can be accordingly determined as shown in Figure 2.

It is also noted that depth maps generated for 3D-TV are often noisy with irregular changes on the same object in color frames, which may cause unnatural-looking pixels in synthesized views as well as reduce the accuracy of ROI detection algorithms applied for color frames [13]. Smoothing the depth map with a low-pass filter can suppress the noises and improve the rendering quality. However, low-pass filtering will blur the sharp depth edges along object boundaries which are critical for high-quality view synthesis. Therefore, in the proposed ROI-BA method, we utilize a bilateral filter introduced in [14] for effectively smoothing plain regions while preserving discontinuities occurred along edge regions. The new filtered depth value,  $\tilde{Z}_s$ , obtained using the bilateral filter is then defined by:

$$\tilde{Z}_s = \frac{1}{k(s)} \sum_{p \in N} f(p-s).g(Z_p - Z_s).Z_p, \quad (2)$$

where  $N$  is the neighborhood around pixel location  $s(u, v)$  under the convolution kernel, and  $k(s)$  is a normalization term.

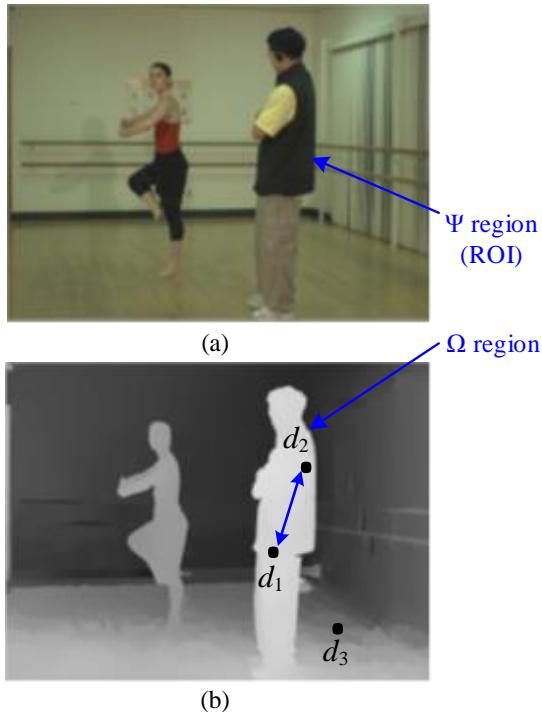


Figure 2. Depth based ROI/Non-ROI detection.

## 2.2. ROI based adaptive bit allocation

The objective of optimal BA scheme is to achieve a target bitrate as close as possible to a given constant while ensuring minimum quality distortion. Knowing that quantization consists in reducing the bitrate of the compressed video signal, the major role of BA algorithms is thus to find for each transform coefficient the appropriate QP under the constraint

$$R(QP) \leq R^{\max}, \quad (3)$$

where  $R(QP)$  and  $R_{\max}$  are the number of coding bits for source samples and the fixed target bit budget, respectively. Let  $D$  denotes the distortion measure between the original and the constructed samples, then the optimal BA problem can be formulated as follows:

$$\min_{QP} D(QP) \text{ subject to } R(QP) \leq R^{\max}. \quad (4)$$

In (4), at frame level, the expected distortion for a frame  $f$  of a video sequence can be measured using the average mean-square error (MSE) as

$$D_f = E_i \left\{ \left( x_f^i - y_f^i \right)^2 \right\} = \frac{1}{XY} \sum_{i=1}^{XY} \left( x_f^i - y_f^i \right)^2, \quad (5)$$

where  $x_f^i$  and  $y_f^i$  denote the original and the reconstructed pixel values of the  $i$ th pixel in the frame  $f$  at the encoder and the decoder, respectively;  $E_i \{\cdot\}$  denotes the expected MSE over all pixels in the frame  $f$ , and  $x$  and  $y$  respectively denote the frame width and height in pixels.

In the conventional BA methods, QP parameter is generally adopted as a global QP applied for all regions in a video frame without considering the different perceiving characteristics of different regions and depths. However, in our proposed ROI-BA method, we propose to use an adaptive BA scheme which adaptively adjusts QP based on visual attention region (ROI) without sacrificing the reconstructed video quality. Specifically, in our proposed method, the lowest QP is assigned to the highest priority region, ROI, and the higher QPs are assigned to the non-ROI regions such as background or transition regions between ROI and non-ROI.

In the proposed ROI-BA, the BA scheme is performed at two levels including frame and CTU levels. Frame level is to initialize a target amount of bits for each region, and CTU level is to make independent BA of CTUs of different regions. At the frame level, let  $R_r$  and  $R_{nr}$  denote the ROI and non-ROI bitrates, respectively. The relation between  $R_r$  and  $R_{nr}$  can be formulated as

$$R_r = \alpha \cdot R_{nr}, \quad (6)$$

where positive constant  $\alpha$  represents the desired ratio between the ROI and non-ROI bitrates. Then, the bitrate of the color video can be represented as a function of other bitrates that are applied for particular regions of the video:  $R = f(R_r, R_{nr})$ . This is a linear function; its coefficients are determined according to the area of those above regions. The parameters of coding process applied for all the CTUs in each region,  $R_r$  and  $R_{nr}$  need to be determined. Based on the importance of those regions to the HVS, it can be set as  $R_r > R_{nr}$ . The problem is to figure out their specific values and how they affect the quality of compressed video. To do

this, we calculate based on the constraints among the area of examined regions, how the capacity of the internet can satisfy to transmit the video.

Assume that  $R^{\max}$  is the maximum bitrate that the network can adapt

$$R^{\max} = S_r \cdot R_r^{\max} + S_{nr} \cdot R_{nr}^{\max}, \quad (7)$$

where  $S_r$  and  $S_{nr}$  are the number of CTUs represented for ROI and non-ROI regions, respectively.

As assumed in (6), the bitrate budget spent for non-ROI coding region in a color frame is then given by:

$$R_{nr}^{\max} = \frac{R^{\max}}{\alpha \cdot S_r + S_{nr}}. \quad (8)$$

Similarly, the bitrate budget spent for ROI coding region is

$$R_r^{\max} = \alpha \cdot R_{nr}^{\max} = \frac{\alpha \cdot R^{\max}}{\alpha \cdot S_r + S_{nr}}. \quad (9)$$

The proposed ROI-BA scheme is then stated as follows: Given  $R^{\max}$ , the proposed BA finds the optimal set of  $QP_i = \{QP_{r,i}^*, QP_{nr,j}^*\}$  ( $i = 0, 1, \dots, S_r$ ;  $j = 0, 1, \dots, S_{nr}$ ), where  $QP_{r,i}^*$  and  $QP_{nr,j}^*$  are the optimal QP chosen for the  $i$ th CTU of ROI and non-ROI coding regions, respectively. This optimal set of  $QP_i = \{QP_{r,i}^*, QP_{nr,j}^*\}$  should be derived to minimize the total distortion  $D(QP_i)$  at the receiver of the 3D-TV system (10)

$$\begin{aligned} & \underset{QP_{r,i}, QP_{nr,i}}{\text{Min}} D(QP_{r,i}, QP_{nr,i}) \\ & \text{subject to } R(QP_{r,i}) \leq R_r^{\max} \\ & \text{and } R(QP_{nr,i}) \leq R_{nr}^{\max} \end{aligned} \quad (10)$$

At the sender, the ROI-BA scheme presented in (10) is processed to get the optimal bitrates assigned to ROI and non-ROI regions to transmit over networks. The proposed adaptive ROI-BA scheme takes all possible combinations of  $QP_i = \{QP_{r,i}, QP_{nr,j}\}$  that satisfy the constraints in (10) and chooses the best one that minimizes the total expected distortion  $D$ .

### 3. Experimental results

Several experiments have been performed to illustrate the effectiveness of the proposed ROI-BA method. The experiment results are reported for several video sequences using 3D test model (3DTM) reference software [15] of the 3D-HEVC extension of H.265/HEVC standard at 30 frames/s. The four main test sequences used in our experiments are *Ballet*, *Breakdancers*, *Alt Moabit*, and *Book Arrival* with resolution is XGA  $1024 \times 768$ , and each sequence consists of 8/16 color views captured

from different cameras (100 frames per view). Along with color views are correlative depth maps generated from stereo. The former two test sequences come from [16] by Microsoft, while the latter are provided by [17] from Heinrich Hertz Institute. In our experiments, the value of  $\alpha$  is set to 1.3 for *Alt Moabit* test sequence and 1.25 for three remaining samples. The first test sequence *Ballet* contains a dancing-ballet woman and a watching-man in a room. The second, *Breakdancers*, contains a dancing man and four other men are watching him in a practicing room. The third test sequence, *Alt Moabit* is a traffic scene in Berlin with some cars parked down near the pavement while other cars are moving. The final one is *Book Arrival* with a man sits in the room before another man coming in and they have a talk.

The ROI detection was applied to the monoscopic 2D sequences. Table I shows results of the proposed ROI detection and tracking method, which is implemented in several situations with the camera is set up indoor and the location of the camera can be fixed or changeable. In these cases, specific ROIs chosen by users are moving objects. And, to evaluate the effectiveness of our proposed ROI detection method, we utilize a *success ratio*, which is measured by:

$$P_{\text{succ}} = 1 - \frac{|N_1 - N_2|}{N_2}, \quad (11)$$

where  $N_1$  and  $N_2$  are the areas of ROI extracted by our proposed method and manually measured method, respectively. After

Table 1. Results of ROI detection and tracking

Video sequence	Environment	Depth structure	ROI's velocity	ROI's position	ROI	Detection result	Tracking result
Ballet	Indoor	Simple	Fast	Almost stable	Ballet dancer	99.3%	Good
Break dancers	Indoor	Complex	Fast	Almost stable	Break dancer	98.5%	Good
Alt Moabit	Outdoor	Simple	Fast	Unstable	Car	99.1 %	Good
Book Arrival	Indoor	Complex	Slow	Unstable	Moving man	97.9 %	Good

ROI extracting, the number of CUs presented for ROI regions are counted for  $N_1$  and  $N_2$ . As reported in Table I, our proposed method achieves a high successful ratio of ROI detection for ROI regions. Specifically, in Table I, compared to the exactly results obtained by the manually measured method, our proposed method always achieves a high successful ratio with the lowest value of 97.9%. As mentioned in Section 2, these results can help to improve efficiently the performance of the proposed ROI-BA scheme. In addition, for subjective evaluation, Figures 3 and 4 show the results of ROI regions extracted by using our method. As can be seen in Figures 3 and 4, ROI regions can be exactly detected and extracted from any frame of input video sequences, *Ballet* or *Breakdancers*.

We also compare the distortion or PSNR performance of the proposed method with that of the conventional 3D-HEVC [7] and ROI-BA scheme introduced in [18]. In [7], the BA scheme is performed without considering the

ROI detection and ROI based BA. The QPs values in [7] therefore are equally assigned to all CTUs encoded in a color frame. Lei *et al.* [18] introduce a multilevel ROIs based BA strategy, in which the MB saliency is derived from depth information of the video sequence, and then the multilevel ROI segmentation is conducted based on the MB saliency distribution.

For fair comparisons between PSNR performance of the proposed ROI-BA with that of the conventional 3D-HEVC and Lei *et al.* [18] methods, we calculate the average

distortion or PSNR of the ROI for  $m$  consecutive frames as follows:

$$PSNR_{ROI} = \frac{1}{m} \sum_{i=1}^m 10 \log_{10} \frac{255^2}{MSE_{ROI}^{(i)}}, \quad (12)$$

where  $MSE_{ROI}^{(i)}$  is the  $MSE$  of the ROI region at the  $i$ th frame,  $MSE$  is given by:

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2. \quad (13)$$

In (13),  $N$  denotes the size of each encoded block in conventional 3D-HEVC video coding, and  $C_{ij}$  and  $R_{ij}$  are the current and reconstructed pixel values, respectively.

It is worth noticing that given the same target bit budget assigned to the same encoded video sequence, the more accurate ROI regions are extracted, the more bitrates need to be allocated to these regions, and thus the higher

PSNR performances can be achieved. The PSNR performances of video coders are also improved if the ROI-BA scheme is adaptively and effectively performed at the sender of video coding system as mentioned in Section 2. In this work, the effectiveness of both ROI detection and adaptive BA scheme obtained from the proposed ROI-BA, 3D-HEVC, and Lei *et al.* [18] methods are compared and verified using different tested input sequences, and different experimental conditions.

Figure 5 shows the PSNR performance of the proposed ROI-BA, the conventional 3D-HEVC, and Lei *et al.* [18] methods corresponding to a wide range of encoding bitrates. As seen in Figure 5, the proposed method outperforms the conventional methods by a large margin of performance. For example, at the bitrate of 6 Mbps, the proposed ROI-BA



(a)



(b)



(c)

Figure 3. ROI detection performed on *Ballet* sequence.

provides up to 0.84 dB better performance than the conventional 3D-HEVC coder. The proposed method also provides higher PSNR performance than the multiple ROI-BA [18] coder. With the same target bit budget assigned to the proposed ROI-BA, however the multiple ROI-BA coder yields worse performances than the proposed method at all values of bitrates as shown in Figure 5. The reason lies in the fact that the ROI based BA scheme is not supported in the conventional 3D-HEVC for adaptive BA, and thus, all CTUs are encoded using equal QPs without assigning more bitrates for ROI regions. In Lei *et al.* [18] method, low-pass filters are not applied for depth maps to smooth and suppress noises on the depths. Therefore, as



(a)



(b)



(c)

Figure 4. ROI detection performed on *Breakdancers* sequence.

confirmed from the experimental results of this method that there are often noisy with irregular changes on the extracted ROI regions, which make confusing on the choice of threshold and thus reduce the accuracy of ROI detection algorithms proposed by this method.

Similar results are obtained from *Breakdancers*, *Alt Moabit*, and *Book Arrival* sequences as shown in Figures 6-8, respectively. For the *Breakdancers* sequence where the motion activities are high and complexity, however, as can be seen in Figure 6, the proposed method also introduces much higher PSNR performance than the 3D-HEVC and multiple ROI-BA [18]. More specifically, at the rate of 7.5 Mbps, the proposed provides

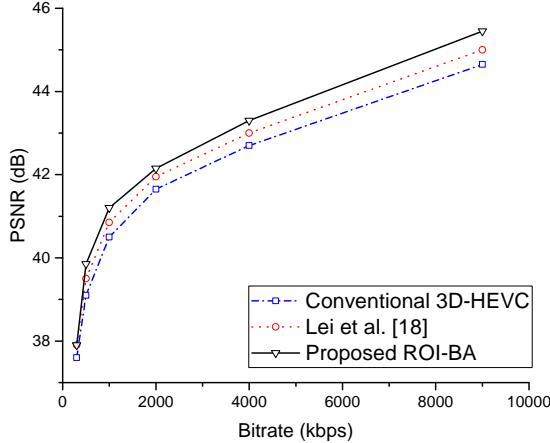


Figure 5. Rate-Distortion of the proposed ROI-BA method as compared with that of conventional 3D-HEVC and Lei *et al.* [18] performed on *Ballet* sequence.

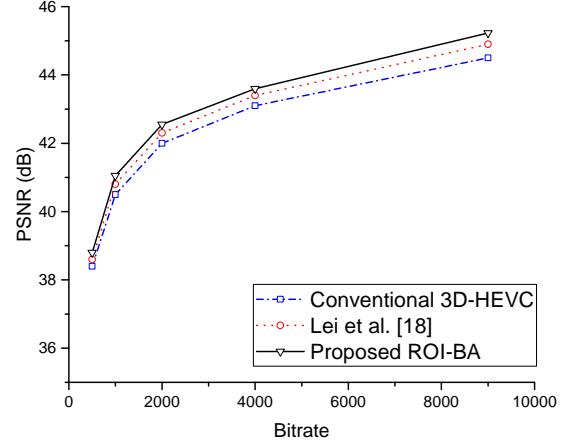


Figure 7. Rate-Distortion of the proposed ROI-BA method as compared with that of conventional 3D-HEVC and Lei *et al.* [18] performed on *Alt Moabit* sequence.

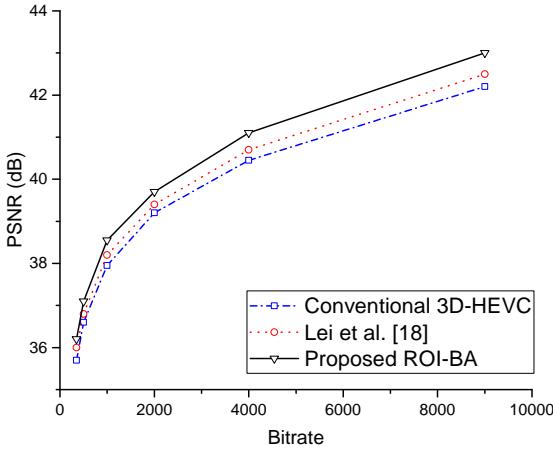


Figure 6. Rate-Distortion of the proposed ROI-BA method as compared with that of conventional 3D-HEVC and Lei *et al.* [18] performed on *Breakdancers* sequence.

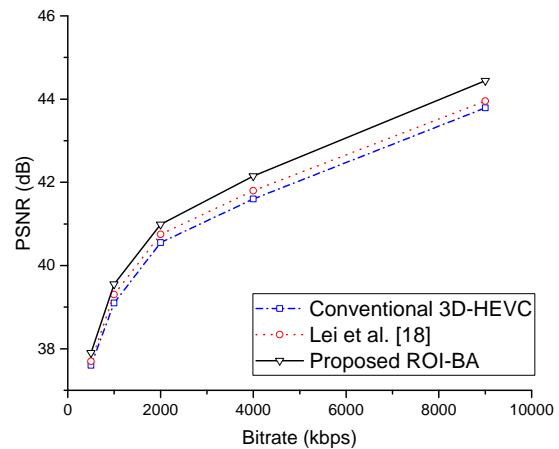


Figure 8. Rate-Distortion of the proposed ROI-BA method as compared with that of conventional 3D-HEVC and Lei *et al.* [18] performed on *Book Arrival* sequence.

about 0.96 dB and 0.71 dB better performances than the 3D-HEVC and multiple ROI-BA coders, respectively as shown in Figure 6.

#### 4. Conclusion

This paper presents a novel and efficient method of allocating bit for ROI and non-ROI regions for robust video transmission. Based on the depth information, which has been smoothed by bilateral filter, the proposed method detects and extracts ROI effectively.

Given the constraint of network bandwidth, the extracted ROI is then allocated more bits than other regions to keep ROI at high visual quality and minimize the overall distortion. Experimental results show that the proposed method achieves better PSNR performances than both conventional 3D-HEVC and Lei *et al.* in various testing sequences and conditions. In future works, multi-levels ROI detections and classifications would be taken into account for further extending our frameworks. Furthermore, it is our belief that by employing additional information from channel feedback reports and unequal error protection

(UEP) scheme applied for ROI regions, the performance of the proposed ROI-BA method can be more improved to provide an optimal end-to-end rate-distortion optimization.

## Acknowledgement

This work was supported by the basic research projects in natural science in 2012 of the National Foundation for Science & Technology Development (Nafosted), Vietnam (102.01-2012.36, Coding and communication of multiview video plus depth for 3D Television Systems).

## References

- [1] Z. He and S. Mitra, "Optimum bit allocation and accurate rate control for video coding via  $\rho$ -domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840-849, Oct. 2002.
- [2] B. Li, H. Li, and L. Li, "Adaptive bit allocation for R-lambda model rate control in HM," *JCT-VC M0036*, 13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG16WP3 and ISO/IEC JTC1/SC29/WG11, Incheon, Kr, 2013.
- [3] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 1, pp. 185-207, Jan. 2013.
- [4] R.A. Khan, A. Meyer, H. Konik, and S. Bouakaz, "Exploring human visual system: Study to aid the development of automatic facial expression recognition framework," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 49-54, 2012.
- [5] H. Hu, B. Li, W. Lin, W. Li, and M. -T. Sun, "Region-based rate control for H.264/AVC for low bit-rate applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1564-1576, Oct. 2012.
- [6] X. Yang, W. Lin, Z. Lu, X. Lin, S. Rahardja, E. Ong, and S. Yao, "Rate control for video phone using local perceptual cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 496-507, Apr. 2005.
- [7] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001-1016, Dec. 2013.
- [8] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.
- [9] B. Lee, M. Kim, and T. Nguyen, "A frame-level rate control scheme based on texture and non-texture rate models for high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 1-14, Mar. 2014.
- [10] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-interest-based rate control scheme for high efficiency video coding," *APSIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1-18, Dec. 2014.
- [11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.
- [12] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. IEEE 99*, vol. 4, pp. 643-656, 2011.
- [13] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Sig Processing: Image Comm.*, vol. 24, no. 1-2, pp. 65-72, 2009.
- [14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proceedings of IEEE international conference computer vision*, pp 839-846, 1998.
- [15] Test Model 6 of 3D-HEVC and MV-HEVC. Available:  
<http://mpeg.chiariglione.org/standards/mpeg-h/high-efficiency-video-coding/test-model-6-3d-hevc-and-mv-hevc>.
- [16] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High quality video view interpolation using a layered representation," *ACM Transactions on Graphics (TOG)*, vol. 23, pp. 600-608, 2004.
- [17] I. Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, and T. Wiegand, "HHI test material for 3D video" ISO/IEC JTC1/SC29/WG11, vol. 15413 Apr. 2008.
- [18] J. Lei, M. Wu, K. Feng, C. Hu, and C. Hou, "Multilevel region of interest guided bit allocation for multiview video coding," *International Journal for Light and Electron Optics*, vol. 125, no. 1, pp. 39-43, Jan. 2014.