



Structural similarity-based rate control algorithm for 3D video

Y. Harshalatha¹ · Prabir Kumar Biswas²

Received: 5 February 2020 / Revised: 21 February 2021 / Accepted: 5 April 2021 /

Published online: 24 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Human vision easily identifies the structural degradation in a video and thus perceptual quality improvement is necessary. To adjust the video quality to match with the human vision we proposed a structural similarity (SSIM) based rate control algorithm for the 3D video. We incorporated a structural similarity index (SSIM) as the quality metric in rate control algorithm. The rate-distortion model and Lagrange multiplier are derived considering the structural dissimilarity ($dSSIM$) as the distortion metric to achieve the rate control with perceptual quality improvement. Furthermore, the optimal joint bit allocation scheme at texture video/depth map level, frame level, and basic unit (BU) level is modified to incorporate $dSSIM$ for measuring distortion. The proposed algorithm is implemented in HEVC compression standard HTM-16.2. The performance of the proposed algorithm is compared using RD curves, BD-Rate comparison, and subjective evaluation. Besides, rate accuracy is also computed to measure the bit rate mismatch. Compared to the original λ -domain rate control algorithm, the proposed algorithm achieves a better SSIM along with a reduction in bit rate.

Keywords 3D video · Virtual view synthesis · Bit allocation · Perceptual quality · SSIM

1 Introduction

The popularity of 3D technology is increasing day by day as the demand for 3D services is increasing in many application areas like television, surveillance, and so on. To support these demands, vast technological improvements are being made at each stage of the 3D video system. Different stages of a 3D video system are acquisition, representation, coding,

✉ Y. Harshalatha
latha_rupesh2002@sit.ac.in; harshalathasit@gmail.com

Prabir Kumar Biswas
pkb@ece.iitkgp.ernet.in

¹ Department of Electronics and Communication Engineering, Siddaganga Institute of Technology, Tumakuru, 572103, India

² Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India

view synthesis, and display. The multiview video acquisition enables the autostereoscopic display while increasing the bandwidth requirement and also the task of encoding. The multiview plus depth (MVD) representation format enables the transmission of few views along with respective depth maps. The remaining views are synthesized using view synthesis.

View synthesis process uses decoded texture videos and depth maps to generate new views at different viewpoints. Synthesized views are generated using 3D warping, which is performed by projecting reference images to 3D world coordinates and converting 3D coordinates back to the 2D target view. The texture video, as well as depth data, are distorted due to lossy compression resulting in synthesis distortion. The relationship of view synthesis distortion (D_{vs}) with texture distortion (D_{tex}) and depth distortion (D_{dep}) is given by the (1) where both texture and depth distortions were considered to have a linear relationship with synthesis distortion.

$$D_{vs} = \alpha D_{tex} + \beta D_{dep} + C \quad (1)$$

where α , β and C are model parameters. However, the error in the depth value leads to a positional error in the target image. The synthesis distortion with the same amount of depth distortion varies according to texture details. As these synthesized views are displayed, it is necessary to reduce the view synthesis distortion. An approach to reduce synthesis distortion is efficient encoding through the rate-distortion optimization (RDO) technique by reducing distortion at the available target rate.

For 3D video, synthesis distortion is to be minimized through proper rate allocation between texture video and depth map. This ensures the better quality of synthesized views by reducing the distortion. The first, bit allocation algorithm was based on the details in depth map and texture video. Bit allocation criterion proposed in [5] allocates bit ratio of 5:1 between texture video and depth map. The fixed ratio does not ensure optimal bit allocation and leads to increased distortion in synthesized views. A search algorithm was proposed in [9] which discusses hierarchical optimization algorithm and finds a pair of quantization steps for texture video and depth map. However, this algorithm has computational complexity due to an extensive search. A view synthesis distortion model was proposed by Yuan *et al.* [18] along with an optimization technique to find optimal values of quantization steps for texture video and depth map. These bit allocation algorithms model the synthesis distortion and try to reduce the same at the available rate. However, the total bandwidth is not considered which is an essential factor for 3D video.

The video coding standards incorporate a rate control algorithm to meet the bandwidth constraint. Bit allocation along with the good quality of the video is possible with a rate control algorithm. The rate control (RC) algorithm mainly tries to balance the bit rate while maintaining the quality of the views. Rate control algorithms are proposed for 2D video in different coding standards. They include TM5 for MPEG2, VM8 for MPEG-4, and the quadratic model for H.264. The bit allocation through rate control was proposed in [11] in which the encoder used was H.264. Similarly, there are rate control algorithms proposed for HEVC, which was replaced later by the lambda-based rate control algorithm. For 3D video, rate control algorithms are proposed by many researchers. Adaptive frame-level rate control is proposed in [14] where the initial QP decision scheme is discussed along with smoothing the bit rate fluctuation. An adaptive bit allocation algorithm with rate control was proposed by [16] using 3D-HEVC. Authors in [3] proposed a rate control algorithm for depth map with rate-lambda model with an adaptive clipping algorithm to exploit depth characteristics. Inter-view dependency-based rate control for 3D-HEVC is proposed in [13]. In this work, based on the synthesis distortion model, bit allocation is done at the texture/depth level, the

view level, and frame level. In these works, the video quality is measured through objective metrics that do not consider the human visual characteristics.

In the literature, the rate control algorithms are based on objective metrics, and to the best of our knowledge, there are very few or no works on subjective quality based rate control for 3D video. However, fewer rate control algorithms for 2D video used SSIM as metric [4, 10, 19] to improve the subjective quality. Motivated by the fact that subjective quality improvement gives a better visual experience to the viewer, in this paper we proposed a structural similarity-based rate control algorithm for 3D video. To improve perceptual quality, rate-distortion analysis is performed using $dSSIM$ as a distortion metric to derive a new RD model. Using this RD model, the relationship between rate and Lagrange multiplier is derived. The bit allocation is performed at the texture/depth level, frame level, and basic unit (BU) level. The proposed algorithm is implemented using HEVC reference software HTM-16.2 [7].

A review of Lambda-domain rate control is briefly explained in Section 2. The proposed SSIM-based rate control algorithm is discussed in Section 3 that includes rate-distortion analysis of texture video and depth map in Section 3.1. The bit allocation at different levels is presented in Section 4. Finally, rate-distortion performance, subjective evaluation and rate accuracy comparison are elaborated in Section 5 which is followed by conclusion in Section 6.

2 Review of lambda-domain rate control

Rate control is necessary to regulate the varying bit rate by modifying the necessary encoder parameters. The two important functions of the rate control algorithm are, allocation of bits at different coding levels and monitor allocated bit levels at different stages. The coding levels in any encoder are the group of pictures (GOP) level, frame level, and basic unit level. Rate-distortion optimization (RDO) is a part of the rate control algorithm where an optimum set of rate and distortion is obtained.

The rate control method adopted in the HEVC encoder that relates rate (R) and quantization step (Q) is a unified rate-quantization (URQ) model [14]. In this scheme, a quadratic RD model is used, and R is related to Q using the (2).

$$R = aQ^{-1} + bQ^{-2} \quad (2)$$

where values of a and b depend on the content of the video sequence. ρ -domain rate control algorithm relates quantization parameter Q with the percentage of zeros in DCT coefficients ρ . With this model also, a relation is obtained between R and Q. According to URQ and ρ -domain method, the rate achieved entirely depends on Q value. Thus, there is a need to obtain a proper relation between R and Q. In [8] drawbacks of R-Q methods are mentioned, and the authors proposed λ -domain rate control algorithm. In λ -domain rate control method, a relation between R and Lagrange multiplier, λ is established to achieve the target bit rate. λ is also the slope of the RD curve. λ -domain RD function is a hyperbolic function given in (3).

$$D(R) = CR^{-K} \quad (3)$$

where K is the model parameter. The relation between R and λ is obtained by minimizing the RD function as in (4)

$$R = \alpha\lambda^\beta \quad (4)$$

where α and β are the parameters. The current GOP level bit allocation always depends on the number of bits allocated for the previous GOP to balance the total bit target. The bit allocation at the GOP level is given by (5).

$$Target_{GOP} = \frac{R_{PicAvg} \times (N_{coded} + SW) - R_{coded}}{SW} \times N_{GOP} \tag{5}$$

where R_{PicAvg} is the average bits per picture, SW is the size of the sliding window, N_{coded} is the number of pictures coded, R_{coded} is the bit cost of all coded pictures and N_{GOP} is the total number of pictures in a GOP. The picture level bit allocation is done using (6) and bit cost assigned to each picture depends on the value of ω_{pic} .

$$Target_{pic} = \frac{Target_{GOP} - Coded_{GOP}}{\sum_{\{AllNotCodedPictures\}} \omega_{Pic}} \times \omega_{PicCurr} \tag{6}$$

where $Target_{GOP}$ is target bits of GOP, $Coded_{GOP}$ is number of bits in the current GOP. The target bits computed for the BU level is given by (7).

$$Target_{BU} = \frac{Target_{pic} - Bit_H - Coded_{pic}}{\sum_{\{AllNotCodedBUs\}} \omega_{BU}} \times \omega_{BUCurr} \tag{7}$$

where Bit_H is the estimated header bits, $Target_{pic}$ is the target bits of a picture, ω_{BU} is the weight of BU level bit allocation.

3 SSIM-based rate control algorithm

To improve the visual quality, it is necessary to incorporate perceptual quality metrics for distortion measurement. We use *SSIM* as perceptual quality metric and distortion is measured using *dSSIM*. The Lagrange multiplier is modified to incorporate *dSSIM* and the steps followed are: (i) derive rate-distortion (RD) model where distortion is measured using *dSSIM* (ii) minimize the RD model to obtain Lagrange multiplier. These steps are explained in the following subsections.

3.1 Rate-distortion analysis

View synthesis distortion is dependent on texture video distortion and depth distortion. Thus, the relationship between rate and distortion is to be analyzed for texture video, depth map, as well as the synthesized view. To obtain the relationship, texture and depth sequences are encoded at different rates. Distortion is computed between the original and encoded sequences. Distortion is measured using both objective quality metric, MSE, and subjective quality metric, *dSSIM*. The RD graph for texture views and depth map of Kendo sequence is shown in Fig. 1a and b respectively where distortion is measured using MSE. The RD graph in Fig. 2 shows the relation between rate and *dSSIM* for Kendo sequence. In both cases, we can observe that the rate and distortion are related using the power model and corresponding R^2 values for different sequences are listed in Table 1.

RD analysis is also done for synthesized views of Kendo sequence as shown in Fig. 3. Original texture video and depth maps are used to generate synthesized views that have no distortion. Decoded texture views and depth maps are used to generate synthesized views that are distorted. Distortion is computed between original synthesized views and distorted

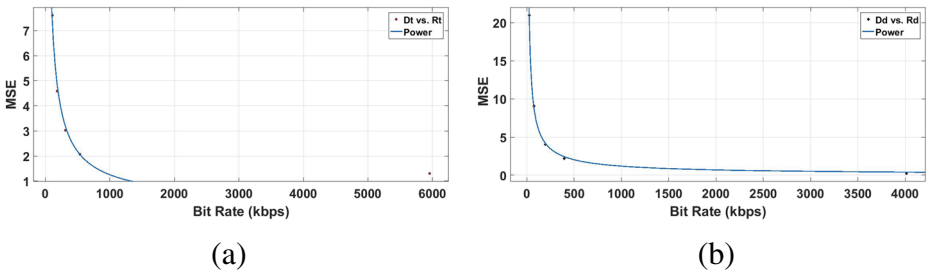


Fig. 1 RD curve for **a** Texture video and **b** Depth map of Kendo sequence (Distortion metric: MSE)

synthesized views using *dSSIM* as distortion metric. For synthesized views also, rate and distortion are related using the power model.

According to RD analysis with *dSSIM* as distortion metric, texture video and depth maps follow hyperbolic RD function as in (8).

$$dSSIM(R) = CR^{-K} \tag{8}$$

where *C* and *K* are the parameters that depend on characteristics of the source. View synthesis distortion $dSSIM_v$, is related to texture video rate R_t and depth map rate R_d as in (9)

$$dSSIM_v = aR_t^{-K_t} + bR_d^{-K_d} \tag{9}$$

where *a*, K_t , *b* and K_d are the model parameters.

3.2 Rate control based on structural similarity

The rate control algorithm achieves better quality videos, along with maintaining the target bit rate. To improve the subjective quality, we are incorporating *dSSIM* for distortion measurement. The works related to perceptual quality improvement through rate control are found in the literature for 2D video [2, 4]. For 3D video, view synthesis distortion is modeled as $dSSIM_v$ as in (9) and is related to texture video rate R_t and depth map rate R_d .

The distortion *dSSIM* can be related to MSE [17] and the resulting expression is given in (10).

$$1 + \frac{MSE(R)}{2\sigma_x^2 + C_2} = CR^{-K} \tag{10}$$

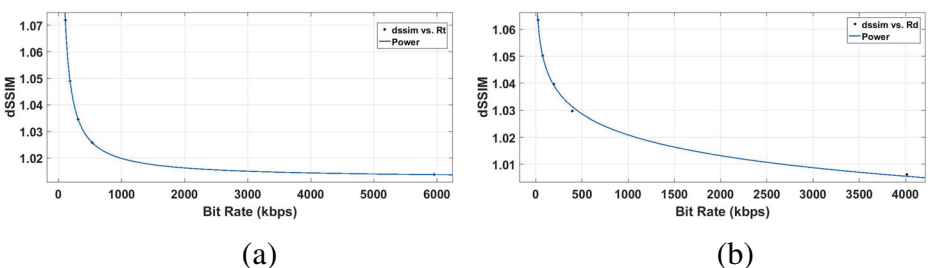


Fig. 2 RD curve for **a** Texture video and **b** Depth map of Kendo sequence (Distortion metric: *dSSIM*)

Table 1 Coefficient of determination measured for Synthesized Views

| Sequence | R^2 |
|---------------|--------|
| Kendo | 0.9985 |
| Balloons | 0.9803 |
| Breakdancer | 0.9761 |
| Shark | 0.9855 |
| Poznan Street | 0.9582 |
| Ballet | 0.9720 |

where σ_x^2 is the variance of the block to be coded and C_2 is constant. Using (10), synthesis distortion of (9) can be expressed in terms of MSE as

$$d_v = a_t(2\sigma_{xt}^2 + C_2)R_t^{-K_t} + a_d(2\sigma_{xd}^2 + C_2)R_d^{-K_d} \tag{11}$$

Lagrange multiplier λ is the slope of the RD curve and is given by (12)

$$\lambda = -\frac{\partial D}{\partial R} = CK(2\sigma_x^2 + C_2)R^{-(k-1)} \tag{12}$$

This Lagrange multiplier is used to achieve better visual quality using the relationship of *dSSIM* with MSE.

4 Bit allocation

Bit allocation for 3D video is very important to achieve distortion-free synthesized views. Thus, we considered bit allocation at three stages: (i) texture/depth level, (ii) frame level and (iii) BU level bit allocation. In each stage, we aimed to allocate the bits to improve quality in the visual sense.

The bit allocation at texture/depth level is solved as an RDO problem. Another approach to allocating bits between texture video and depth map is to set an optimum bit ratio between them. The bit ratio is computed by minimizing the RD model for view synthesis. In this

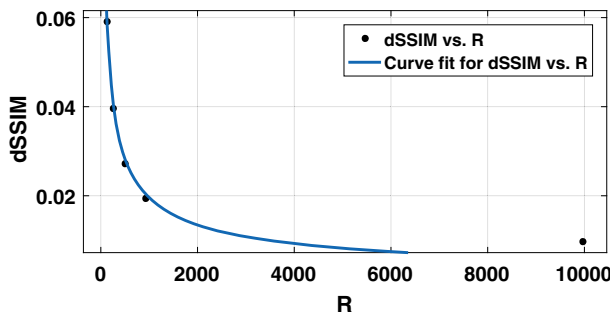


Fig. 3 RD curve for synthesized view of Kendo sequence

paper, we consider the RD models with distortion being measured using $dSSIM$ as in (9). Thus, the RDO is formulated as given in (13).

$$\begin{aligned} & \min_{(R_t, R_d)} dSSIM_v \\ & s.t. \quad R_t + R_d \leq R_c \end{aligned} \tag{13}$$

where $dSSIM_v = a_t(2\sigma_{xt}^2 + C_2)R_t^{-K_t} + a_d(2\sigma_{xd}^2 + C_2)R_d^{-K_d}$ and σ_{xt}^2 and σ_{xd}^2 are the variance of texture video and depth map respectively. The expression for bit ratio, η is calculated [11, 13] as in (14).

$$\eta = \frac{(2\sigma_{xt}^2 + C_2)R_t}{(2\sigma_{xd}^2 + C_2)R_d} \tag{14}$$

The variance of texture video and depth map are to be calculated along with the other model parameters. The ratio of texture to depth bit is expressed as in (15) and (16).

$$R_t = R_c \cdot \frac{\eta}{\eta + 1} \tag{15}$$

$$R_d = R_c \cdot \frac{1}{\eta + 1} \tag{16}$$

For the frame-level bit allocation, we used the concept of entropy. For each frame in both texture video and depth map, we compute the entropy. The entropy gives the amount of information present in a frame. As human vision is sensitive to structural information, the frames with higher entropy must be given more importance. Target bit allocation at the frame level is calculated as in (6). The weight $\omega_{PicCurr}$ is considered to be constant in HEVC, and this makes an equal distribution of the target bits between the frames. However, the amount of information in each frame is different and requires bit allocation accordingly. Thus, we measured the entropy of each frame and assigned to weight $\omega_{PicCurr}$.

For bit allocation at BU level, a combination of just noticeable depth difference (JNDD) threshold with texture gradient is used considering the effect of depth on synthesis distortion. Human perception varies with change in depth distance and is sensitive to change in depth value. JNDD model explains that human vision is highly sensitive to closer objects compared to distant objects [12]. A JNDD threshold is set such that, the depth value below the threshold cannot be perceived. Another fact that affects synthesis distortion is its non-linear relation with depth distortion. This implies that, with the same depth distortion, synthesis distortion varies with the texture details. Thus, a new weight factor is proposed for BU level bit allocation (7) as given in (17).

$$\omega_{BUCurr} = dJND_{avg} \cdot t_G \tag{17}$$

where $dJND_{avg}$ is the average JND and t_G is the texture gradient. $dJND$ is computed using (18).

$$dJND = \left\{ 1 - \frac{(1 - D_{jnd})^2}{(1 + D_{jnd})^2} \right\} \cdot 255 \tag{18}$$

where D_{jnd} is determined by subjective perception experiments in [1] and is given by (19).

$$D_{jnd} = \begin{cases} 21 & \text{if } (0 \leq A(i, j) < 64) \\ 19 & \text{if } (64 \leq A(i, j) < 128) \\ 18 & \text{if } (128 \leq A(i, j) < 192) \\ 20 & \text{if } (192 \leq A(i, j) < 255) \end{cases} \tag{19}$$

5 Results

For testing and verification of the proposed algorithm, we used 3D-HEVC reference software HTM-16.2. The algorithm is tested with different input sequences, namely Kendo, Balloons, Breakdancer, Ballet, Poznan Street and Shark [6, 20]. For each sequence, views are rendered using the rendering algorithm available in HTM-16.2. The GOP size is set to 8 with an intra period of 24. The main objective of the proposed algorithm is to improve the perceptual quality. To the best of our knowledge, SSIM-based rate control algorithms for 3D video are not found in the literature. Hence, all the sequences were encoded with the proposed rate control algorithm and compared with the original λ -based rate control algorithm of HTM-16.2. We also compared the proposed algorithm with the SSIM-based bit allocation algorithm that uses model parameters to determine the quantization parameter.

5.1 Rate-distortion performance evaluation

The rate-distortion performance of the proposed rate control algorithm is compared with the original rate control algorithm of HTM-16.2 and SSIM-based bit allocation algorithm. For comparison, the original RC algorithm is considered as Anchor1, and the SSIM-based

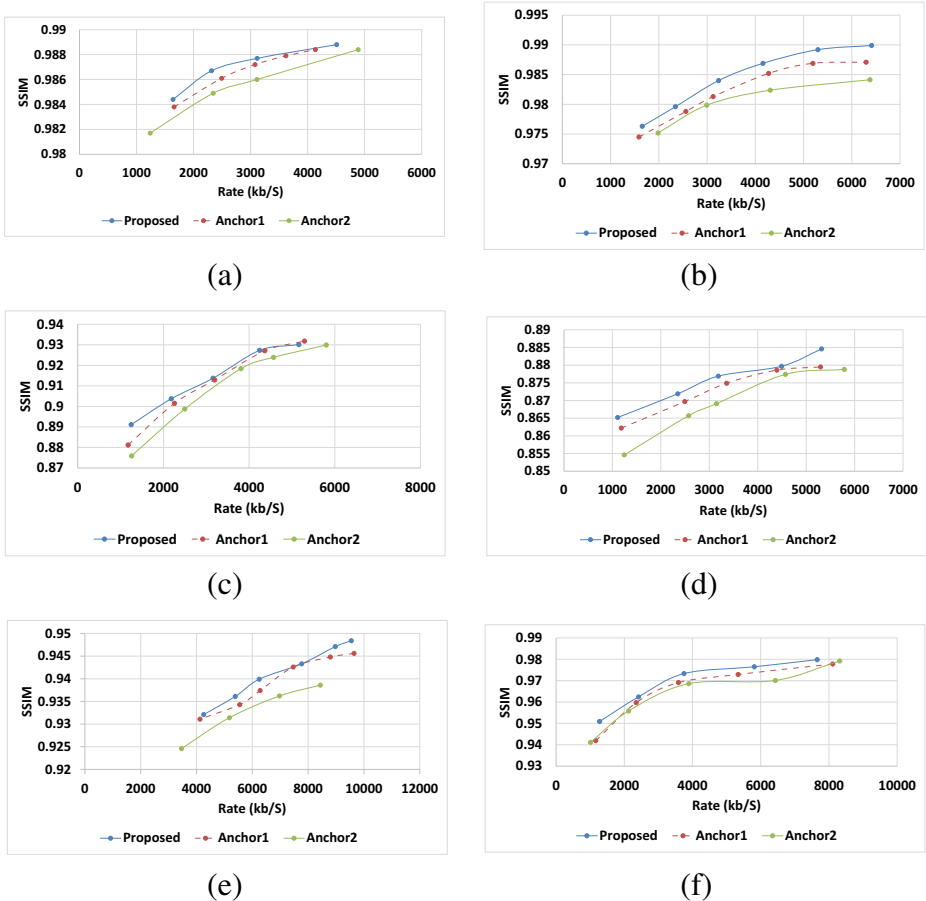


Fig. 4 RD curves for a Kendo and b Balloons c Breakdancer d Ballet and e Poznan Street f Shark

Table 2 BD-rate comparison: proposed vs Anchor1

| Sequence | Δ SSIM | Δ Rate |
|---------------|---------------|---------------|
| Kendo | 0.0007 | −17.4121 |
| Balloons | 0.0023 | −18.2283 |
| Breakdancer | 0.0029 | −8.751 |
| Ballet | 0.0039 | −20.256 |
| Poznan Street | 0.0018 | −7.6113 |
| Shark | 0.0031 | −15.6328 |

bit allocation method [15] is considered as Anchor2. The RD graph is plotted as in Fig. 4. The proposed algorithm shows significant improvement in perceptual quality for Balloons, Ballet, PoznanStreet, and Shark. However, for Kendo and Breakdancer sequences, the proposed algorithm shows better improvement in SSIM at lower rates compared to Anchor1 and Anchor2.

The BD-Rate calculation is presented in Tables 2 and 3 to verify the proposed algorithm with Anchor1 and Anchor2 respectively. The BD-Rate shows the reduction in rate and improved SSIM for the proposed algorithm when compared to the other two algorithms. The rate is reduced by 14.6% and 27.34% with respect to Anchor1 and Anchor2 respectively. The rate control accuracy is summarized in Tables 4 and 5 for Kendo and Balloons respectively. The difference between the actual bit rate and the target bit rate is the error. The computed error is much lesser compared to Anchor1 and Anchor2. The average mismatch is less in the case of the proposed algorithm.

As video quality is decided by both spatial and temporal information, we computed Temporal SSIM between the frames. The spatio-temporal structural gradient is computed in three directions x-t, y-t and x-y directions [15]. The temporal SSIM is computed for synthesized views of the proposed algorithm and compared with anchor1 and anchor2. The RD graph is plotted for Kendo and Ballet as shown in Fig. 5.

5.2 Subjective evaluation

The proposed rate control algorithm improves the perceptual quality and to support this claim, we conducted a subjective evaluation to compare the proposed algorithm with the original rate control algorithm. Twelve participants viewed the sequences which are encoded with SSIM-based rate control algorithm and lambda-based rate control algorithm. For assessing the subjective quality, participants were asked to grade the sequences on a five-point scale: 5 - Excellent, 4 - Good, 3 - Fair, 2 - Poor, 1 - Very poor. Table 6 shows the

Table 3 BD-rate comparison: proposed vs Anchor2

| Sequence | Δ SSIM | Δ Rate |
|---------------|---------------|---------------|
| Kendo | 0.0019 | −31.697 |
| Balloons | 0.0045 | −28.271 |
| Breakdancer | 0.0089 | −18.198 |
| Ballet | 0.0085 | −36.263 |
| Poznan Street | 0.0049 | −23.313 |
| Shark | 0.0048 | −26.3276 |

Table 4 Rate accuracy for the sequence *Kendo*

| Target rate (Mbps) | View | Anchor1 | | Error | Proposed | | Error |
|--------------------|------|---------|--------|-------|----------|----------|-------|
| | | Rate | | | Rate | | |
| | | T | D | | T | D | |
| 1.5 | V1 | 522.26 | 385.62 | 10.61 | 637.02 | 221.24 | 9.438 |
| | V3 | 503.49 | 247.89 | | 603.57 | 179.74 | |
| 2.5 | V1 | 763.79 | 610.75 | −0.25 | 927.808 | 273.88 | −7.35 |
| | V3 | 743.69 | 375.43 | | 900.01 | 214.33 | |
| 3 | V1 | 1021.84 | 668.72 | 2.66 | 1216.026 | 393.8496 | 3.89 |
| | V3 | 1000.34 | 389.09 | | 1200.874 | 306.0216 | |

Table 5 Rate accuracy for the sequence *Balloons*

| Target rate (Mbps) | View | Anchor1 | | Error | Proposed | | Error |
|--------------------|------|---------|--------|-------|----------|--------|-------|
| | | Rate | | | Rate | | |
| | | T | D | | T | D | |
| 2.5 | V1 | 746.26 | 625.04 | 2.43 | 945.82 | 316.49 | −6.22 |
| | V3 | 718.34 | 471.25 | | 804.22 | 277.93 | |
| 4 | V1 | 1552.81 | 736.85 | 6.93 | 1796.48 | 446.94 | 3.95 |
| | V3 | 1496.41 | 491.5 | | 1596.46 | 318.29 | |
| 5 | V1 | 1792.77 | 936.45 | 3.88 | 2044.87 | 658.8 | 5.97 |
| | V3 | 1701.58 | 763.24 | | 1817.67 | 777.26 | |

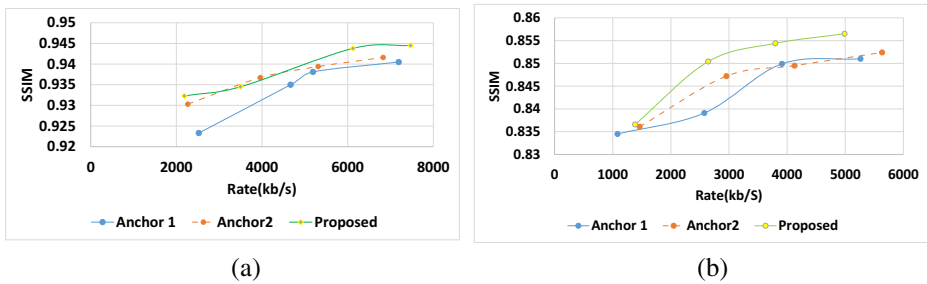


Fig. 5 Temporal SSIM comparison for **a** Kendo and **b** Ballet sequence

Table 6 Subjective evaluation: mean opinion score and standard deviation

| Sequence | SSIM-based rate control | | Lambda-based rate control | |
|--------------|-------------------------|------|---------------------------|------|
| | MOS | SD | MOS | SD |
| Kendo | 4.25 | 0.75 | 4.02 | 0.79 |
| Balloons | 4.08 | 0.81 | 4.00 | 0.79 |
| Breakdancer | 3.52 | 0.80 | 3.33 | 0.71 |
| Ballet | 4.00 | 0.63 | 3.9 | 0.68 |
| Pozan Street | 4.45 | 0.82 | 4.13 | 0.85 |
| Shark | 3.85 | 0.65 | 3.6 | 0.72 |

mean opinion score (MOS) and standard deviation (SD), which implies that the proposed algorithm performs better than the lambda-based algorithm.

6 Conclusion

In this paper, we proposed a rate control algorithm for 3D-HEVC using which optimum bit allocation is performed for texture video and depth map along with improving the visual quality. The relation between rate and distortion of synthesized views is derived experimentally, and distortion is measured using *dSSIM*. The λ value is modified to achieve structural similarity-based rate control. In addition, bit allocation is performed at the texture/depth level, frame level, and basic unit level. At texture/depth level, bit ratio is derived between texture and depth map, whereas the entropy level decides the weight factor at frame level. At basic unit level, a product of DJND and texture gradient is used as a weight factor. Subjective analysis, RD performance, and BD-Rate evaluation show that the proposed algorithm performs better compared to the original RC algorithm.

References

1. Bai Y, Zhang Y, Li Z (2015) 3D Video Coding using Just Noticeable Depth Difference based on H. 265/HEVC. In: 11Th international conference on computational intelligence and security (CIS). IEEE, pp 142–145
2. Chen X, Gu D (2014) Macroblock Layer Rate Control based on Structural Similarity and Mean Absolute Difference for H. 264. Adv Multimed:5
3. Cordina M, Debono CJ (2016) A Depth Map Rate Control Algorithm for HEVC Multi-View Video plus Depth. In: IEEE International conference on multimedia & expo workshops (ICMEW), pp 1–6
4. Cui Z, Gan Z, Zhu X (2011) Structural Similarity Optimal MB Layer Rate Control for H. 264. In: Proceedings of IEEE International Conference on Wireless Communications and Signal Processing (WCSP), pp 1–5
5. Fehn C (2004) Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV. In: Electronic imaging 2004. International society for optics and photonics, pp 93–104
6. Fujii Laboratory, Nagoya University. Available at: <http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/>. [Online; accessed on 29-July-2018]
7. HM Reference Software HTM-16.2. Available at: <https://hevc.hhi.fraunhofer.de/trac/3d-hevc/browser/3DVCSsoftware/tags/HTM-16.2>. [Online; accessed on 21-September-2018]
8. Li B, Li H, Li L, Zhang J (2014) Lambda Domain Rate Control Algorithm for High Efficiency Video Coding. IEEE Trans Image Process 23:3841–3854
9. Morvan Y, Farin D (2007) Joint Depth/Texture Bit-Allocation for Multi-View Video Compression. In: Picture coding symposium (PCS)
10. Ou TS, Huang YH, Chen HH (2010) A Perceptual-based Approach to Bit Allocation for H. 264 Encoder. In: Visual communications and image processing 2010. International society for optics and photonics, vol 7744, pp 77441b
11. Shao F, Jiang G, Lin W, Yu M, Dai Q (2013) Joint Bit Allocation and Rate Control for Coding Multi-View Video plus Depth Based 3D Video. IEEE Trans Multimed 15(8):1843–1854
12. De Silva DVS, Fernando WAC, Nur G, Ekmekcioglu E, Worrall ST (2010) 3D Video Assessment with Just Noticeable Difference in Depth Evaluation. In: IEEE International conference on image processing. IEEE, pp 4013–4016
13. Tan S, Ma S, Wang S, Wang S, Gao W (2017) Inter-View Dependency-Based Rate Control for 3D-HEVC. IEEE Trans Circ Syst Video Technol 27(2):337–351. <https://doi.org/10.1109/TCSVT.2015.2511878>
14. Tan S, Si J, Ma S, Wang S, Gao W (2014) Adaptive Frame Level Rate Control in 3D-HEVC. In: Proceedings of IEEE visual communications and image processing conf, pp 382–385. <https://doi.org/10.1109/VCIP.2014.7051586>

15. Harshalatha Y, Biswas PK (2018) SSIM-based Joint-Bit Allocation for 3D Video Coding. *Multimed Tools Appl* 77(15):19,051. <https://doi.org/10.1007/s11042-017-5327-0>
16. Yang C, An P, Shen L (2016) Adaptive Bit Allocation for 3D Video Coding. *Circ Syst Signal Process*:1–23
17. Yeo C, Tan HL, Tan YH (2013) On Rate Distortion Optimization using SSIM. *IEEE Trans Circ Syst Video Technol* 23(7):1170–1181
18. Yuan H, Chang Y, Li M, Yang F (2010) Model Based Bit Allocation Between Texture Images and Depth maps. In: International conference on computer and communication technologies in agriculture engineering (CCTAE). IEEE, vol 3, pp 380–383
19. Zhao H, Xie W, Zhang Y, Yu L, Men A (2013) An SSIM-Motivated LCU-Level Rate Control Algorithm for HEVC. In: Proceedings of Picture coding symp. (PCS), pp 85–88. <https://doi.org/10.1109/PCS.2013.6737689>
20. Zitnick CL, Kang SB, Uyttendaele M, Winder S, Szeliski R (2004) High-quality video view interpolation using a layered representation. In: *ACM Transactions on graphics (TOG)*. ACM, vol 23, pp 600–608

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.