Grzegorz Pastuszak, Ph.D., D.Sc., university prof.
Institute of Radioelectronics and Multimedia Technology
Faculty of Electronics and Information Technology
Warsaw University of Technology
grzegorz.pastuszak@pw.edu.pl

22.08.2023

Reviewer's opinion

on Ph.D. dissertation authored by

Yasir Abdulraheem Ahmed Al-Obladi, M.Sc.,

entitled:

Modeling of Codecs for 3-Dimensional video

1. Problem and its impact

The subject of the review is the doctoral dissertation of Mr. Yasir Abdulraheem Ahmed Al-Obladi, M.Sc., entitled "Modeling of Codecs for 3-Dimensional Video." The reviewed doctoral dissertation is experimental. In his doctoral dissertation, the Ph.D. candidate raised the problem of effective bit allocation and bit rate control for 3-Dimesional (3D) video. In particular, the author's goal was to find a model that combines bit rates with quantization steps in multi-view coding with depth maps (MVD). The author set the thesis that there exist general models which are very similar for several video coding standards.

The objectives of the research, set in the doctoral dissertation, were formulated clearly and understandably. The problems raised are important and topical in the field of the problem, as they concern effective methods of controlling 3D vision encoders. Thus, the goals have been correctly defined and consistently meet the expectations of the scientific community dealing with video compression. They are ambitious and in line with current engineering and technical sciences challenges in the discipline of Technical Information Technology and Telecommunications.

2. Contribution

The dissertation is divided into nine chapters and 15 appendixes including the extension of experimental results for all test sequences. Author’s contributions are described in Chapters 4-8. The dissertation addresses the bit-rate control problem in multi-view video compression with depth maps. Three video coding standards (AVC, HEVC, and VVC) are used to find general models. Developing models for the latest standards points to the novelty of the described achievements. The reviewer agrees with the author regarding main and secondary contributions. The main, original contributions of the dissertation are as follows:

1. In Chapter 5, the author describes the developed bit-rate allocation model. The bit-rate allocation relates to the selection of quantization parameters for depth maps (QD) based on that for video (QP). To this end, statistically optimal QP-QD pairs are found and applied in all considered standards. The pairs maximize the quality of virtual views synthesized from two coded views and corresponding depth maps.
2. In Chapter 7, the author presents the development of a relatively simple unified encoder model. The simplification relates to reducing the number of model parameters describing rate-quantization (R-Q) curves. In particular, one or two of three parameters are fixed by averaging of their best values for test sequences. The model accuracy is evaluated for GOP-level and

frame-level coding. The encoder model provides acceptable accuracy of the estimated bit rate as a function of the quantization step.

3. In Chapter 8, the two models (allocation and encoder) are combined to verify their efficiency for the bit-rate control. Results show that the QP/QD prediction accuracy is sufficient to be applied in video coding.

The above contributions are practical as they can be applied in MVD video encoders to achieve better quality and bit-rate accuracy. The contributions were also published in five conference papers during the Ph.D. course.

Apart from the main contributions, the dissertation has less important ones:

1. In Chapter 4, the author shows that encoder models for H.264/AVC, H.265/HEVC, and H.266/VVC are closely related to each other. This allows the estimation of a single model parameter for the two newer standards to obtain the three-parameter model knowing the H.264/AVC model for the same content. It is possible save computations for newer standards when coding for H.264/AVC is performed first.
2. In Chapter 6, the author analyses unexpected rate-distortion (RD) dependence for high bit rates and some videos. In particular, the increased bit rate involves small decreases in virtual view quality. It is explained and proved that this effect originates from inaccuracies of depth maps introduced by imperfect depth estimation algorithms. For high bit rates, quantization noise becomes smaller than the inaccuracies. Additionally, smaller quantization steps cannot smooth the inaccuracies. As a consequence, the quality of the virtual view slightly decreases with a significant increase in bit rate.

3. Correctness

The doctoral dissertation is prepared correctly from the editorial point of view. The edition of the work is careful, i.e., the drawings showing the results of the analyses are clear. The same applies to the tables. The language of the dissertation (English) is correct. There are only minor typos or style errors. These defects are not significant from the point of view of the substantive assessment of the work.

The author of the work correctly presented the starting point of the research, which is the origin, justification of the subject, and motivation. In general, the research methodology described in the dissertation is correct. However, the evaluation of the developed models is insufficient:

1. Tests for whole sequences are not the best way to evaluate the accuracy of the rate control. It would be better to see results for each GOP since the bit rate should be adjusted to the transmission channel bandwidth in periods of about 1 second or less.
2. The difference between QP_{goal} and $QP_{calulated}$ is not the best measure to evaluate the efficiency of the model. It would be better to measure the difference in bit rates since it is the goal of the rate control. As QP values are discrete, their differences do not fully reveal the accuracy. For example, the difference in bit rate may be noticeable, but QP_{goal} and $QP_{calulated}$ are equal in some cases.

Average model parameter values included in the last row of Tab. 5.2 and in Tab. 5.3 are incorrect. The results are expected to be better than those reported in the dissertation. It would be nice to see the correct results.

More concerns and comments on correctness are included in Section 5 of the review.

4. Knowledge of the candidate

Chapters 2 and 3 are tutorial parts of the dissertation. The state of the knowledge is described in Chapter 2. The author review video compression methods, rate control algorithms, rate-distortion modeling, bit allocation, immersive video, multi-view video coding, depth map coding, point clouds, ray spaces, view synthesis, and the dependence of its quality on depth map quality. Common elements of the methodology in MVD video coding are given in Chapter 3.
The literature on problems addressed in the dissertation list is rich, and the author referred to many publications. In particular, the list of references includes above 300 items. Their number and importance seem sufficient to show the background of the work and the current state of the art.
The author presents the logical sequence of the research cited in the work in a clear way.
The tutorial chapters (2 and 3) and the numerous references confirm a general knowledge of the Ph.D. candidate in the field of video compression, especially in 3D video coding.

5. Other remarks[1]

The dissertation describes models evaluated for stereoscopic videos. It would be beneficial to see how they can be extended to more views. Is it sufficient to apply the same QP-QD pairs?
Concerns and comments to successive parts of the dissertations are described below.

Abstract:
Page 5: "…to estimate the bit-rate or frame size of stereoscopic video …" – frame size is known and is not estimated. The same misleading statement is in the Polish version of the abstract on page 8.

Page 8: “mona” -> “można”

Chapter 2:

Page 22: In equation (2.2), MAD is computed for residual data rather than as the difference between original and reconstructed samples.

Page 24: N is the total number of pixels in the frame rather than a total number of the frame.

Page 24: In [Wang_13a, Wang_13b], there is no statement that "ρ-R model cannot be used with modern video coding standards." The approach proposed in the papers is based on ρ-R model.

Page 24: “… was modeled …” should be “… is modeled …”

Pages 35 and 37: The author does not explain what linear and nonlinear Multi-view sequences are.

Chapter 3:

The number of frames in each sequence should be provided in Tab. 3.1. The author should state if all frames are used in experiments.

Chapter 4:

A model is useful when it can be applied to predict modeled parameters before processing, e.g., in a video encoder, we would like to know the bit rate for QP. Although the author states that "the goal is to use such parameters in rate control," it is unclear how to do it if someone wants to compress video only with the VVC format. The author silently assumes the video is coded with AVC and one or two newer formats. What practical use cases are considered for such a scenario?

Is the model sensitive to GOP structure, i.e., the number of B frames? What is the practical sense of using the AVC model in favor of HEVC and VVC in the wide QP range? Is it sufficient to limit to subranges/regions?

It would be instructive to see point corresponding to GOPs in Fig. 4.5 and 4.6. Points would form clouds. Different colors could distinguish sequences.

The author states that the practical QP range is either 15-50 or 25-50. The claims are not consistent. In the reviewer's opinion, the upper bound is too high.

Chapter 5:

Average parameter values included in the last row of Tab. 5.2 are incorrect. Correct values should be: 1.27, -13.64, 1.26, -12.57, 1.26, -11.51, 1.14, -4.83. As a consequence, values in Tab. 5.3 should also be corrected to: 1.23 and -10.64.

Results presented in Tab. 5.4 and 5.5. are obtained for incorrect model parameters. In most cases, the model with global parameters involves worse results than those specific to different codecs. However, global parameters provide better results for 3D-HEVC, which seems to be surprising. The reason could be incorrect parameter values. It is expected that the results will be better than those reported in the dissertation.

Page 75: Expression on SSE should include the square function of each difference, i.e., 2 should be placed as the power.

It is unclear what the goal of the modeling of view-bitrate to total-bitrate in dependence on QP is. Although the knowledge of this dependence is instructive and informative, the modeling should provide more benefits.

Chapter 6:

Page 79: "This increasing part of the R-D curve …" should be "This decreasing part of the R-D curve …"

What QP is used for experiments illustrated in Fig. 6.10 and 6.11? 22?

In Subsection 6.5 (the first paragraph), the conclusion that noised depth maps involve lower qualities of virtual views seems obvious and typical. The author should stress that this is expected. In this context, unexpected dependencies should be concluded.

Chapter 7:

Model parameter values in Tab. 7.1. are obtained by averaging values optimal for each of the eight sequences used/listed in Appendix J. Tables in the appendix should include the additional row with averaged values. The same concerns relate to the following appendixes (K, L, M, N, and O) and the tables in the chapter.

Frame size denotes the width and height of video frames. The author also uses the term to compressed bit-rate of a frame. This involves some ambiguities for a reader.

It is not necessary to repeat what GOP pattern is used in experiments. It would be sufficient to state it in Chapter 3 and refer to this information at the beginning of Chapter 7.

Results for simplified models are provided for two test sequences. Errors for different QPs have some correlations. It would be leveraged to decrease them by introducing non-parametrized correction functions depending on QP (look-up tables). Is it the case? Could it be benefitable taking into account other sequences?

The frame-level rate control is evaluated only for HEVC and VVC, whereas the GOP-level rate control is also evaluated for MV-HEVC and 3D-HEVC. Why?

Chapter 8:

Bit-rate allocation is not only the relation between QP and QD. It is also important to predict the relation of QPs for different frame types in a GOP. This problem was skipped by the author.

In Section 8.2, Step 4 of the algorithm is unclear as it is not explained in the section how to obtain QP_{goal} and QD_{goal} to calculate $Q_{calculated}$. The author should write that it is explained in the following section. It would be clearer to distinguish MVD coding with QP_{goal} and QD_{goal} as a preceding step.

In section 8.3, the dependence between QP_{goal} and $QP_{initial}$ is not clarified. The sentence before Equation 8.3 suggests that the initial value should be close to the goal value, and the difference is a certain error 'delta'. Generally, the error should be minimized. However, in Step 1, 'delta' is specified as a value from the set {2, 3, 4, 5}. The author forces the reader to guess what the algorithm is.

The phrase "… the simulation of the relevant estimation procedure for estimation of …" is clumsy due to the repetition.

At the end of Section 8.3, the author states that QP used in experiments ranges from 25 to 50. In the following Section, he states that "the experiments were performed for all QP_{goal} in the range from 25 to 45". The upper bound is different. Why?

In Section 8.4, the author should explain why the total number of tests is 42. The reviewer guess there are 21 QP_{goal} values (the range is from 25 to 45), and two $QP_{initial}$ values are computed for each QP_{goal}. Tests for whole sequences are not the best way to evaluate the accuracy of the rate control. It would be better to see results for each GOP since the bit rate should be adjusted to the transmission channel bandwidth in periods of about 1 second or less.

The difference between QP_{goal} and $QP_{calulated}$ is not the best measure to evaluate the model's efficiency. It would be better to measure the difference in bit rates since it is the goal of the rate control. As QP values are discrete, their differences do not fully reveal the accuracy. For example, the difference in bit rate may be noticeable, but QP_{goal} and $QP_{calulated}$ are equal in some cases.

Experiments reported in Sections 8.4 and 8.5 are related to procedures specified in Sections 8.2 and 8.3. The reader must guess how to combine the procedures into a single one due to entangled dependencies and skipping important details. It would be better to provide the combined procedure in the dissertation directly. The reviewer guessed that results reported in Sections 8.4 and 8.5 give the percentage of QP_{goal} and $QP_{calulated}$ differences when using the model anchored in $QP_{initial}$.

6. Conclusion

Taking into account what I have presented above and the requirements imposed by *Article 13* of *the Act of 14 March 2003 of the Polish Parliament on the Academic Degrees and the Academic Title* (with amendments)[2] and *Article 187 of the Act of 20 July 2018 – The Law on Higher Education and Science*, my evaluation of the dissertation according to the three basic criteria is the following:

A. Does the dissertation present an original solution to a scientific problem? (the selected option is marked with **X**)

X				
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

B. After reading the dissertation, would you agree that the candidate has general theoretical knowledge and understanding of the discipline of **Information and Communication Technology**, and particularly the area of multimedia technology?

X				
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

C. Does the dissertation support the claim that the candidate is able to conduct scientific work?

X				
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

Grzegorz Pastuszak

Signature

---

[2] http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf