

Reviewer's opinion
on Ph.D. dissertation authored by
Jakub Stankowski
entitled:

Analysis, extensions, and optimizations of high-efficiency video compression techniques

1. Problem and its impact

This thesis addresses a collection of problems from different relevant fields of scientific research in the area of video compression and video transmission. These include the problem of optimum mode decision for video encoding algorithms, efficient and generalized multiview video coding, and robust watermarking processes. The work is based on substantial analysis of existing video coding technologies, such as the reference implementation of the High Efficiency Video Coding standard ISO/IEC 23008-2 | ITU-T H.265 HEVC, based on which transcoding approaches are also developed. All of these topics have substantial practical meaning. Efficient encoding algorithms are the key for cost effective compression of video signals (which cover a substantial amount of traffic on the Internet). Multiview video coding is relevant for immersive video applications such as video transmission to head-mounted displays. The presented research enables extension towards more flexible camera setups which broadens the usability of such technologies. Finally, light-weight and robust watermarking gains increasing relevance in the era of AI-generated or -manipulated video.

2. Contribution

The publications compiled as the basis for this cumulative dissertation span a time frame of about 10 years (ranging from 2012 to 2022). Given the broadness of the research areas addressed, it includes various original contributions in the different areas. In general, it seems that the author has often been contributing as a team player, substantially supporting his team in a broad range of research work. Chapter 1 “Analysis of video coding efficiency and statistical properties of data in the encoder” starts of with detailed analyses of HEVC encoder decisions and studies of the efficiency of the representation of non-binary symbols and the applied contexts for adaptive binary arithmetic coding (CABAC). This knowledge provides the foundation for relevant contributions to transcoding and multi-generation encoding of video sequences. The presented work provides valuable insight into effects and limits of multi-generation encoding and specifically points out conditions for optimum decisions w.r.t single bitstream transcoding and provision of multiple bitstreams. Furthermore, the collected statistical insights are successfully applied for enhanced encoder decision making where the author reports substantial encoder time reduction at partially minor loss of compression efficiency. Chapter 2 “Video encoder optimizations” presents a method for improved rate-distortion optimized quantization and an approach for improved motion estimation for recent video coding schemes. Here, the approach for improved rate cost estimation based on quasi-exact numbers for the entropy coding is considered to be most relevant.

In Chapter 3 “Extensions for encoding of multiview and FTV content”, the major contribution is seen in the development of an improved prediction of motion vectors and disparity vectors. The presented approach makes use of the relation between the pictures of different views of a given Multiview scenario where information of the reference views (base or lower-level dependent views) is exploited to improve the precision of the applicable displacement vectors. These were part of an overall standardization proposal by Poznan University which induced a number of patent filings with contributions from the author. Further work in this chapter deals with the generalization of the applicable camera arrangements for Multiview video coding, moving from a strictly linear camera setup as specified in HEVC to more flexible camera arrangements, including circular arrangements or even 2D camera positioning.

Chapter 4 “Compressed bitstream watermarking” presents contributions not related to compression but to modifications of the coded video sequence for marking purposes. The author presents two approaches for application in AVC and a transfer of these concepts to the HEVC standard. The impact of applying watermarking via high-level syntax modifications in the header in comparison to approaches working in the transform domain is discussed. The author claims original contribution of the concept of using scaling matrices and implementation of the framework for both, AVC and HEVC, specifically providing a single detector mechanism operable for both bitstream types.

3. Correctness

In general, the work presented in this thesis appears to be well-founded, solid and trustworthy. The author spent a lot of effort in rigorous analysis and significant implementation work. The general claims of the thesis are considered to be correct.

Some aspects of slight misconception or impreciseness are listed below. In Chapter 0, an overview of the basic concepts of video coding schemes is provided, using HEVC as the focus of the presentation. It is noted that the concept of an Access Unit in HEVC is different from what is presented in the thesis. Conceptually, an HEVC Access Unit is the set of all NAL units associated to one picture (definition slightly broadened when including multiview or scalable video), hence, the NAL units of the video coding layer plus potential non-video coding layer NAL units like parameter sets, SEI messages, picture or stream delimiters etc. The presentation around Fig. 3 suggests that an Access Unit would be an independently decodable group of pictures. It is noted that the presentation in this chapter only focuses on the aspects of the video coding scheme which are later addressed by the research work. E.g., intra prediction and the loop filters are not further detailed.

In Chapter 1, the state-of-the-art in fast motion estimation at the time might not be fully covered. The chapter specifically claims an exhaustive search to be performed in the HEVC reference software, where already fast methods might have been applied. The chapter sometimes is slightly imprecise in usage of terms “block” and “unit” (CTB/CTU, CB/CU, PB/PU, TB/TU). It should be noted that the “block” refers to a sample array of a colour component while the “unit” refers to all components plus motion vectors, intra modes etc. Towards the end of the chapter, future video coding standardization is briefly addressed. It should be noted that the development of a potential successor of VVC is just about to be started. The work in JVET on tools in the ECM software was pure exploration work and the standard will most certainly not be based on the ECM (as claimed in the text).

In Chapter 2, the sum of squared transformed differences (SSTD) is mentioned. The value of this metric might have been discussed a bit more, specifically since (based on the application of orthogonal transforms and appropriate normalization) the computation of the squared error in the spatial and in the transform domain might induce the same results (Parseval theorem).

A minor imprecision is observed in Chapter 3 where the same acronym “NP” is used for “Nested Prediction” and “Neighboring Prediction” in one paragraph.

4. Knowledge of the candidate

Throughout the thesis, the author demonstrates a broad and deep knowledge in the discipline of Information and Communication Technology, specializing in state-of-the-art video coding, embracing 2D video, Multiview and free viewpoint TV scenarios. Chapter 0 provides a concise presentation of the relevant components of the fundamental technology which are needed to dive into the details of the various research directions presented in this thesis. The information provided here serves as a solid basis for the following chapters. It is noted that, by the cumulative character of the dissertation, concise information on the specific foundations and application scenarios related to the respective research topics is provided in the corresponding papers.

No arguments against the authors claimed general knowledge and understanding of the Information and Communication Technology discipline are raised.

5. Other remarks¹

n/a.

6. Conclusion

Taking into account what I have presented above and the requirements imposed by Article 187 of *the Act on Higher Education and Science of the Polish Parliament* (Dz. U. 2018 poz. 1668 with amendments)², 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)

☐	☒	☐	☐	☐
<i>Definitely YES</i>	<i>Rather yes</i>	<i>Hard to say</i>	<i>Rather no</i>	<i>Definitely NO</i>

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 **video coding technology**?

☒	☐	☐	☐	☐
<i>Definitely YES</i>	<i>Rather yes</i>	<i>Hard to say</i>	<i>Rather no</i>	<i>Definitely NO</i>

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

☒	☐	☐	☐	☐
<i>Definitely YES</i>	<i>Rather yes</i>	<i>Hard to say</i>	<i>Rather no</i>	<i>Definitely NO</i>



Digital unterschrieben von
Mathias Wien
Datum: 2026.08.01 14:24:35
+02'00'

Signature

¹ Optional

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