

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

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

1. Problem and its impact

The subject of the review is the doctoral dissertation of Mr. Jakub Stankowski, M.Sc., entitled "Analysis, extensions, and optimizations of high-efficiency video compression techniques." The reviewed doctoral dissertation is the summary of PhD candidate's research in the area of video compression. In particular, the submitted dissertation addresses a pivotal area of modern telecommunications and computer science: the efficient representation and protection of visual data. This research is situated within the context of the rapid growth of multimedia systems, where video content now accounts for approximately 80% of global network traffic. Consequently, developing methods to reduce data volume while maintaining high image quality and low computational complexity is a fundamental challenge for the evolution of modern digital infrastructure.

The objectives of the research, set in the doctoral dissertation, were formulated clearly and rigorously. The problems raised are important. The goals have been correctly defined and consistently meet the expectations of the scientific community dealing with video compression. At the time of publication, they were in line with engineering and technical sciences challenges in the discipline of Technical Information Technology and Telecommunications. Furthermore, while the research primarily focuses on H.265/HEVC, the proposed methodologies and conclusions are universal to hybrid coding schemes and have been actively applied to the development of next-generation technologies, such as the Enhanced Compression Model (ECM)—the successor to Versatile Video Coding (H.266/VVC)—demonstrating the continued relevance and impact of this work on the future of video compression standards.

The primary scientific problem of this work is the optimization of hybrid video coding techniques, with a specific focus on the H.265/HEVC (High Efficiency Video Coding) standard. While the standard offers significant compression gains—roughly doubling the efficiency of its predecessor—it introduces immense computational complexity on the encoder side. This challenge is two-fold: there is a theoretical need to establish the maximum performance limits of these algorithms and a practical requirement to develop encoder control methods that allow for implementation in devices with limited processing power, such as mobile hardware.

The significance of this dissertation extends beyond standard 2D video by addressing critical issues in 3D and multiview (FTV) video representation. In these systems, a key scientific challenge involves exploiting correlations between multiple views and depth maps to enable seamless user navigation in 3D space without exceeding acceptable transmission bandwidth. Equally relevant is the problem of copyright protection through digital watermarking. The author correctly identifies the need for methods



that embed watermarks directly into the compressed bitstream, thereby avoiding the prohibitive energy and time costs of full decoding while maintaining bitstream consistency and perceptual invisibility.

2. Contribution

The dissertation is divided into six chapters numbered from 0 to 5. The author's contributions are described in Chapters 1-4. The original contribution of the dissertation is the development and experimental verification of a comprehensive set of methods for the analysis, optimization, and extension of hybrid video coding techniques, with a focus on the H.265/HEVC standard. The PhD candidate presents a multifaceted research approach that encompasses statistical modelling of bitstreams, non-normative encoder optimizations, and architectural extensions for specialized video formats. Specifically, the candidate's statistical analysis of H.265/HEVC data provided the foundation for high-performance encoder control algorithms.

The most significant and innovative element of this work is the candidate's contribution to 3D and multiview video coding (3D-HEVC). These contributions include the original Nested Prediction (NP) algorithm for motion and disparity vector prediction, for which the candidate had a primary role (80% contribution in publication [A12]), conceptualizing the method and developing the software from scratch. These innovations were integral to a team-developed codec recognized as one of the top two globally in an international MPEG/VCEG competition [A13]. While this was a collaborative effort, the candidate was specifically responsible for the dedicated distortion estimation algorithm for depth map coding, which effectively reduced visual artifacts in synthesized views by incorporating edge energy measures [Section 3.3.1]. Additionally, he contributed to the extension of 3D-HEVC for circular camera arrangements (25% share in [A11]), achieving a 6% bitrate reduction for non-linear setups.

The second significant and innovative element of the research is a unified framework for compressed-domain watermarking for both H.264/AVC and H.265/HEVC. The candidate was the lead researcher for the H.265/HEVC watermarking framework, holding a 90% contribution share [A14]. He independently developed the extremely complex H.265/HEVC entropy decoder/encoder required for bitstream-level processing and conceptualized the advanced mechanisms for adaptive drift avoidance and Sign Data Hiding (SDH) tool interaction to maintain bitstream consistency. For the H.264/AVC suite, the candidate was primarily responsible for the scaling matrices watermarking technique (70% share in [A15]) and the development of the unified watermark detection algorithm used across all proposed methods. He also co-developed the local strength adjustment mechanism for H.264/AVC (40% share in [A16]), which ensures the watermark remains robust against severe attacks like camcording while being perceptually invisible.

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

1. Chapter 1 discusses the statistical analysis and control of high-speed encoders. The candidate's contribution in this area is foundational, characterized by a very high degree of individual work (an 35-80% share) in the core statistical studies.
 - a) In Section 1.2-1.4 and [A1-A3], the candidate performed a complex study on the frequency of selection for Coding Unit (CU), Prediction Unit (PU), and Transform Unit (TU) sizes in H.265/HEVC. He was primarily responsible (80% contribution) for the research conceptualization and the development of a dedicated bitstream analyzer software from scratch, which consists of an H.265/HEVC-compliant entropy decoder and a modified "entropy counter". His independent experiments provided direct insights into which combinations of units can be omitted to accelerate the encoder's mode selection process.



- b) Section 1.5 and [A4] presents the analysis on entropy coding efficiency. With an 80% share, the candidate developed the methodology and software tools to evaluate all 154 CABAC context models. He revealed that a significant portion of binary symbols is encoded with suboptimal efficiency, leading to potential data expansion, which serves as a baseline for future entropy coding improvements.
 - c) Section 1.6 and [A5, A6] analyses transcoding and multi-generation encoding. The candidate's input was significant (60%) in the award-winning study on multiple H.265/HEVC encoding-decoding cycles. He independently developed the software and conducted experiments involving 80 cycles, determining that quality loss saturates over time. In the broader transcoding loss analysis [A6] (20% share), he focused on the implementation of research scenarios and experimental verification.
 - d) Section 1.7 and [A7, AL0198, AK0173] presents complexity reduction algorithms. Practical application of the gathered statistical knowledge led to a fast mode selection algorithm that prunes statistically rare encoding modes. While this was a team effort (candidate's 35% contribution), his role focused on the implementation, testing, and result analysis for high-resolution sequences. He also contributed to improvements for the ECM model (successor to H.266/VVC), specifically in optimizing probability tracking in CABAC and proposing the adaptive "VERMIR" encoding direction.
2. In Chapter 2, the candidate shifted from broad statistical analysis to the targeted mathematical refinement of coding tools, with a contribution reaching up to 85% in motion estimation research.
- a) Section 2.1 and [A8] presents the exact cost RDOQ implementation for H.265/HEVC. The candidate conceptualized the research idea and supervised the development of software to evaluate an RDOQ variant using exact calculations of SSD distortion and CABAC bit-cost. Although this was a collaborative effort (30% contribution), he was instrumental in establishing the theoretical "performance limit" of this technique, achieving an approximately 1% compression gain over the standard H.265/HEVC implementation.
 - b) Section 2.2 and [A9-A10] discusses the reduction of motion estimation complexity and analyses tradeoffs. The candidate had a dominant role (85% contribution) in this research area. He was solely responsible for creating highly optimized software implementations of motion estimation algorithms (Full Search, Log2, TZ Search) from scratch. Using SIMD (SSE/AVX) instructions, he demonstrated that the more accurate SSD metric can be computed with a complexity comparable to the simplified SAD metric. Furthermore, he proposed adaptive search range algorithms based on PU size, allowing for significant time savings during encoding.

The proposed solutions are characterized by their high degree of practicality and industrial relevance. All algorithms were implemented within established reference software (HM and ECM) and are compatible with standard-compliant decoders. The candidate's ability to conduct independent scientific work is further proven by his creation of a dedicated bitstream analyzer software from scratch and highly optimized implementations of motion estimation algorithms using SIMD (SSE/AVX) instructions [A9, A10], which demonstrate his advanced software engineering skills applied to technical problems. The high quality of the candidate's work is further evidenced by indicators related to publications, patents, the award, and citations:

- Most significant research findings appeared in prestigious journals, most notably IEEE Transactions on Image Processing (IF=13.7) and IEEE Access (IF=3.6).
- The solutions are protected by a significant number of international patents granted by the USPTO (e.g., [US8761527], [US9582859]) and the EPO.
- The candidate received the Kalman Fazekas special award for research on multiple encoding cycles.

- The scientific impact is demonstrated by substantial citation counts, such as 58 citations (Google Scholar) for the work on 3D view synthesis [A13].

A defining strength of this dissertation is the candidate's adherence to a rigorous scientific methodology, characterized by a clear transition from the formulation of hypotheses to their systematic experimental validation. Unlike many works that present a fragmented list of engineering improvements, Jakub Stankowski structured his research around three fundamental PhD theses (defined in Section 0.1.1), which serve as the scientific backbone for the entire publication cycle. The candidate effectively navigated a 'logical research sequence' where each major chapter acts as a proof-of-concept for a specific thesis. Specifically, for Thesis 1, he moved from foundational statistical discovery to the creation of high-performance control algorithms, proving that deep bitstream modelling is a prerequisite for effective complexity reduction. Thesis 2 was actively proven through the implementation of geometry-aware tools in Chapter 3, where the candidate demonstrated that utilizing 3D scene information directly results in substantial bitrate gains. Finally, Thesis 3 was validated by the development of a unified watermarking framework in Chapter 4, where the author provided empirical evidence that robust protection can be achieved in the compressed domain with computational efforts magnitudes lower than traditional methods. The explicit recapitulation of these findings in Chapter 5 confirms that the work is not merely a summary of software developments, but a goal-oriented scientific inquiry that successfully bridges the gap between theoretical assumptions and experimental verification.

3. Correctness

The findings and methodologies presented in this dissertation are highly credible and scientifically rigorous. The trust in the results is primarily established by the fact that the core of the work consists of a collection of peer-reviewed scientific articles published in prestigious, high-impact journals such as IEEE Transactions on Image Processing and IEEE Access. Furthermore, the candidate's arguments are consistently supported by a sound mathematical framework and a transparent experimental methodology. Several aspects regarding the correctness and technical quality of the work are particularly noteworthy:

1. **Adherence to International Standards:** All experiments were conducted following the Common Test Conditions (CTC) and reference configurations defined by international expert groups (ISO/IEC MPEG and ITU-T VCEG). This ensures that the results are not only accurate but also directly comparable to established global benchmarks.
2. **Scale and Quality of Empirical Data:** The analysis is based on extensive datasets. For instance, the statistical study of syntax elements involved the examination of 320 encoded bitstreams, and the multi-generation encoding study encompassed 3,840 coding points over 80 cycles. Such a large scale of empirical data provides strong statistical significance to the conclusions.
3. **Quality of Prototype Software:** The candidate demonstrated advanced software engineering skills by developing specialized tools from scratch, including a dedicated H.265/HEVC bitstream analyzer and a complex H.265/HEVC entropy decoder/encoder for watermarking. The implementation of these tools within reference software like HM, HTM, and ECM further validates their technical correctness.
4. **Experimental Validation:** The correctness of the proposed 3D coding tools was externally verified through the MPEG Call for Proposals, where the team's codec (incorporating the candidate's algorithms) was recognized as one of the two best in the world based on both objective (PSNR) and extensive subjective tests (MOS) involving 62 subjects.



The mathematical approximations used in the dissertation, such as those for frame type contribution, exhibit extremely high coefficients of determination (R^2 values up to 0.9988), indicating an almost perfect fit between the proposed models and the experimental data. No significant flaws were identified; the candidate honestly discusses the trade-offs of his solutions, such as the increased computational complexity of the "Exact cost RDOQ" or the bitstream expansion inherent in certain watermarking scenarios [Chapter 2, 529].

A significant discrepancy is identified between the descriptive conclusions in the text and the actual numerical data presented in the results tables. In Section 1.7.1 (page 28) and in the abstract of publication [A7], the candidate claims that for 1920x1080 video sequences (Class B), the proposed fast mode selection algorithm achieves a 'threefold and twofold reduction' in encoding time with BD-rate increases of 4% and 1.5% respectively. However, according to the data in Table 1 (page 28), the actual 'Encoding time reduction' for Class B sequences is 33.32% and 49.82% for a BD-rate increase of $\sim 1.5\%$ ($\text{th}=1.5$) and $\sim 4\%$ ($\text{th}=2.5$), respectively.

From a mathematical standpoint, a 'twofold' (2x) reduction implies a new execution time of 50% of the original (a 50% reduction), while a 'threefold' (3x) reduction implies a new execution time of 33.3% of the original (a 66.6% reduction). The experimental results show that the algorithm actually achieves a twofold reduction (49.82%) where a threefold one is claimed, and a 1.5-fold reduction (33.32%) where a twofold one is claimed. Furthermore, across all test classes in Table 1, the time reduction values range from 18.63% to 59.41%, which does not support the claim of a 'threefold' (66.6%) reduction in any scenario. While the achieved speedup is still practically significant, the candidate's descriptive summary overstates the performance compared to the verified empirical data.

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 in general. There are some minor typos and style errors listed in Section 5. In summary, these defects are not significant from the point of view of the substantive assessment of the work.

4. Knowledge of the candidate

The candidate demonstrates a comprehensive and profound knowledge of the discipline of Information and Communication Technology, with a specialized focus on video compression systems. This expertise is evident throughout the dissertation, particularly in the following aspects:

- **Tutorial quality of the Introduction:** Chapter 0 serves as a technical tutorial, covering the evolution of Hybrid Video Coding (HVC) from MPEG-2 to the latest H.266/VVC and ECM models. The author provides a clear distinction between normative standards and encoder implementations, which is crucial for understanding current research challenges. The quality of these introductory sections is high, reflecting the candidate's ability to synthesize complex technical concepts into a coherent educational narrative.
- **Wide range of addressed problems:** The candidate's research is not limited to a narrow niche. It spans a logical research sequence starting from fundamental statistical analysis, through low-level DSP optimizations (RDOQ, Motion Estimation), to high-level system extensions such as 3D/multiview video and Free-Viewpoint Television (FTV), and finally to security aspects in the form of compressed domain watermarking. This wide scope proves that the candidate has a holistic understanding of the video processing pipeline.
- **Advanced Implementation Skills:** A significant indicator of the candidate's expertise is his ability to work with and modify extremely complex reference software models, such as HM, VTM, and ECM.



The candidate developed an H.265/HEVC bitstream analyzer from scratch and implemented original tools (e.g., 'Exact cost RDOQ' or 'VERMIR') within these models, which often consist of hundreds of thousands of lines of code. This demonstrates high-level engineering skills and the ability to bridge the gap between theoretical algorithms and practical, standard-compliant implementations.

- International scientific engagement: The candidate's participation in international standardization activities (MPEG, JVET) and his contributions to the Enhanced Compression Model (ECM) development indicate that his knowledge is at the forefront of global ICT research. This is further validated by the co-authorship of the significant number of patents and high-ranking publications in the cycle.
- Depth and completeness of references: The list of over 230 references is comprehensive and remarkably up-to-date, including foundational papers, current international standards, and the latest research from top-tier journals (e.g., IEEE Transactions on Image Processing). This reflects a diligent and systematic approach to literature review and a solid grounding in the scientific heritage of the field.

In summary, the dissertation provides strong evidence that Jakub Stankowski possesses not only a broad theoretical background in ICT but also the specialized technical mastery required to conduct and lead advanced research in the field of multimedia telecommunications.

5. Other remarks¹

While the proposed method in Section 1.7.1 successfully reduces computational complexity, the resulting efficiency losses appear disproportionately high for many practical applications. The candidate describes a 4.03% and 8.41% increase in BD-rate (for Class B and Class C respectively) as a 'slight increase'. In the context of modern video coding standards, where even a 1% gain is considered significant, such losses are substantial. This is particularly evident in Table 1, where for Class F, a moderate 39.87% reduction in encoding time comes at the cost of a prohibitive 25.19% increase in BD-rate. Similarly, for Class E, a 54.67% speedup results in an 8.92% loss. While the method offers a welcome scalability through the 'th' threshold, the practical utility of operation points with BD-rate losses exceeding 2-3% is questionable for high-quality distribution. The candidate's descriptive assessment of these losses as 'slight' may therefore be overly optimistic from an industrial and standard-compliant perspective.

A significant concern arises regarding the proportionality of the declared contributions in publication [A7] ("Fast mode selection..."). The candidate, listed as the third author among six, declares a 35% share of the work. From a logical and mathematical standpoint, this declaration is problematic. In academic practice, the first and second authors are presumed to have contributed at least as much as, if not more than, the third author. If the first two authors were to claim a share equal to or greater than the candidate's 35%, the total contribution for just the top three authors would exceed 100%, without even accounting for the remaining three co-authors. This high percentage (35%) for a third-tier position suggests either an improper ordering of authors or an overestimation of the candidate's actual individual contribution. Given that Section 1.7.1—which is based on this specific paper—already contains identified mathematical discrepancies regarding encoding time reduction, this additional



inconsistency further weakens the reliability of this part of the dissertation. The candidate should be asked to provide a detailed justification for this specific percentage and clarify the exact division of labor among all six co-authors during the public defense.

In Section 1.7.2, the PhD candidate presents an interesting two-pass encoder control mechanism designed to assist mode selection by using noise reduction. The core idea is to perform a full mode search on a 'noise-reduced version' of the sequence to prevent the Rate-Distortion Optimization (RDopt) algorithm from making sub-optimal, 'greedy' decisions caused by random pixel fluctuations. While the architectural benefits of this approach are clearly stated—achieving a 1.5% bitrate reduction—the dissertation summary does not specify the actual denoising algorithm or filter used to create the reference sequence. Although the author provides a citation to a specific publication [Stan18a], for the sake of the completeness of the self-report, it would be appropriate to identify the class of the filter (e.g., a simple low-pass FIR filter, Gaussian blur, or more advanced denoising). The specific characteristics of the chosen denoising method (such as how it handles edge energy vs. texture) are crucial, as they directly influence the 'chain of decisions' made by the encoder in the first pass. Without this information, it is difficult to fully assess the conditions under which the proposed mechanism remains effective across different types of source noise.

A significant inconsistency is between the evaluation of watermarking for the H.265/HEVC and H.264/AVC standards in Chapter 4. For the H.265/HEVC implementation (Section 4.2), the candidate provides clear objective metrics, reporting PSNR drops of 0.3–1.1 dB and bitrate increases of 0.3–1.8%. However, for the H.264/AVC-based methods (Section 4.1), similar quantitative data regarding objective quality degradation is largely missing, with the author relying on the qualitative claim that distortions are 'perceptually unnoticeable'. Furthermore, the reported average bitrate expansion of 3.8% for H.264/AVC is explicitly linked to the 'special mode' that converts the entropy coding from CABAC to CAVLC to reduce decoding complexity. Because these two processes (scaling-based modification and entropy coder switching) are bundled in the results, it is impossible to evaluate the 'pure' impact of the scaling algorithm itself on compression efficiency. To provide a complete scientific assessment, the dissertation summary should include objective results (BD-rate and PSNR) for the H.264/AVC implementation where CABAC is maintained. Such data is necessary to isolate the cost of the watermarking scaling modification from the overhead introduced by the CAVLC conversion, allowing for a fair comparison of the proposed watermarking techniques across both standards.

There is a notable omission in the technical summary of the watermarking research presented in Chapter 4. The candidate explicitly claims the development of the watermark detection algorithm as a major original achievement of the dissertation. This is further supported by the author's declarations in Appendix 1, where a significant contribution (70% for [A15] and 40% for [A16]) is claimed specifically for the 'development of the watermark detection algorithm'. Despite being a central component that ensures the compatibility of the proposed system across both H.264/AVC and H.265/HEVC standards, the actual detection mechanism is only briefly mentioned in Section 4.1 (page 51). The summary provides only a high-level conceptual description, stating that detection is based on 'the difference in energy of transform coefficients' and 'statistical analysis of DCT-transformed frames'. Given that the detection algorithm is a claimed authorial contribution and serves as the foundation for the entire 'coherent system,' a technical outline—including the fundamental mathematical formulas or a block diagram of the detector's response—should have been included in the self-report chapters (specifically in Section 4.1). The current level of detail in the summary chapters does not allow the reader to assess



the scientific merit or technical complexity of the detection process without referring to the full text of the attached publications.

The dissertation is based on a cycle of 16 publications, many of which involve a large number of co-authors (ranging from 4 to 9). While the candidate is the lead author on 13 of these papers, reporting contributions of 80-90% for the core research, there are concerns regarding the separability of findings in the remaining works. Specifically, publications [A6], [A11], and [A13] list the candidate's contribution at only 20-25%. Given that these papers (especially [A13] published in IEEE Transactions on Image Processing) represent some of the most significant scientific results of the group, it is challenging to distinguish the candidate's individual breakthroughs from the collective output of the research team. Although the author attempted to clarify his specific roles in the summary chapters (e.g., claiming sole responsibility for the 'Exact cost RDOQ' or 'Nested Prediction' algorithms), the high degree of collaborative synergy makes it difficult to assess whether the primary scientific novelty of certain sections should be credited to the candidate or to the team as a whole.

Minor typos and style errors are listed below in three groups:

1. Typos and Spelling Errors:

- "reducte" instead of reduce (p. 23: "avoid processing of I frames to reducte computational complexity").
- "af" instead of of (p. 47: "a development af scaling matrices").
- "ompresed" instead of compressed (this error is repeated twice on p. 47: "AVC-ompresed" and "HEVC-ompresed").
- "testes" instead of tested (p. 52: "for all four testes sequences").
- "CONCLUSSIONS" instead of CONCLUSIONS (in the header of section V of article A4).
- "RESRUCTS" instead of RESULTS (in the header of section III of article A15).
- "ot" instead of of (p. 23: "Consequently, the author ot this thesis decided to...").

2. Grammar and Syntax:

- Subject-Verb Agreement: In the abstract of article A4, there is an agreement error: "for each and every binary symbols" - should be the singular "symbol".
- Tense/Form Consistency: In section IV of article A4: "those particular bins was encoded" – should be "were".
- Repeated Grammatical Error (Polonism): A recurring grammatical error is identified involving the incorrect use of the verb 'allow'. On page 50 (and similarly on pages 11 and 42), the author uses the construction 'allowing to reduce' or 'allowing to avoid'. In English, 'allow' requires an object (e.g., 'allowing the user to reduce') or should be followed by a prepositional phrase, such as 'allowing for the reduction of'.

3. Style and Consistency:

- Colloquialisms: Article A4 contains the phrase: "The things look better for...", which is somewhat too informal for a scientific publication. A phrase such as "More favorable results were observed for..." would be more appropriate.
- Nomenclature: While generally consistent, the author interchangeably uses standard names (e.g., "AVC" vs. "H.264/AVC" and "HEVC" vs. "H.265/HEVC"). Though both are correct, a unified notation is usually recommended for the final version of a dissertation.
- Mathematical Notations: There are occasional errors in the rendering of symbols or variable names, such as "img_resg" instead of the likely intended img_res.



- On page 15, the author refers to the first major part of the work as 'Section 1'. However, this part is explicitly designated as 'Chapter 1' in the Table of Contents.

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 multimedia 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>

Gregor Pastuszak
Signature

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