Pr. Daniel MENARD
INSA Rennes,
IETR Lab
20 avenue des Buttes de Coësmes
CS 70839 - 35 708 RENNES Cedex 7
daniel.menard@insa-rennes.fr

May 23, 2025

Reviewer's opinion
on Ph.D. dissertation authored by
Mateusz Lorkiewicz
entitled:
Block partitioning in video encoding with the use of artificial neural network

1. Problem and its impact

With the emergence of new video services and new video formats, the data traffic due to video has grown exponentially over the past 15 years. To face this high increase, new video standards are proposed to provide coding techniques that are even more efficient than existing ones and limit the bitrate while ensuring a better user quality of experience. This coding efficiency is obtained at the cost of a high increase in the encoding process complexity

This research work aims at proposing new techniques for reducing the complexity and execution time of the video encoding process in the context of the HEVC standards with the All-Intra configuration. The encoding process aims to solve the rate-distortion optimization (RDO) problem, which corresponds to a combinatorial optimization problem. The encoder must select the configuration *i.e.*, coding parameters, leading to the minimal RD cost. The RDO problem comprises various sub-problems, including coding tree partitioning. The search space of the RDO problem is significantly too large to check all possible candidates. This research work focuses on the coding tree partitioning sub-problem, which is known to be the most complex part of the RDO process. Deep learning techniques are considered to predict the most appropriate configurations to test during the determination of the coding tree partitioning. The most important problem considered in this thesis is the acceleration of the coding tree partitioning process.

The definition of complexity reduction techniques requires a scientific approach based on signal processing, artificial intelligence, and computer engineering fields.

The widespread adoption of a new video coding standard requires efficient hardware or software video encoders. Video standard evolution leads to complexity increase for the video encoding process to obtain better video coding efficiency compared to previous standards. Thus, techniques for complexity reduction of video coding are mandatory to obtain efficient encoders. The most promising complexity reduction techniques will be considered by the companies developing video encoders. Also, this kind of research work will improve the global knowledge on video coding and will help in the definition of the new generations of video standards.

2. Contribution

In this research work, new techniques for reducing the complexity and execution time of the video encoding process are proposed in the context of the HEVC standards with the All-Intra configuration. This research work addresses the most complex aspect of the RDO process, specifically the coding tree

partitioning. Two deep learning-based techniques are proposed to predict a tensor that will be used to select the configurations that can be skipped during the determination of the coding tree partitioning. The first technique, named “basic approach”, aims at predicting the size of the coding unit for the coding tree partitioning. The second technique, named “extended approach”, aims at predicting both the size of the coding unit and the prediction unit. In this context, the PU division is treated as an additional level of the quaternary tree for the coding tree partitioning. The strength of the two proposed networks is that they are light with a moderate number of parameters. Two decision algorithms are proposed; the first one is based on hard decision and the second one is based on soft decision. For each algorithm, two variants are proposed, one based on the index of depth level and one based on the probability of depth level.

The author has proposed original techniques to reduce the complexity of the tree partitioning process for video encoding. The contributions exploit advanced techniques based on deep learning. The case of HEVC standards with the All-Intra configuration is considered.

We can regret that the last video coding standard, VVC, was not considered in this research work. For this new standard, using a Quad Tree Multi-Type Tree, the problem of reducing the complexity of the partitioning process is even more acute. Solid arguments were provided in the document to explain this choice. At the beginning of the research work, VVC was not yet standardized. Moreover, the reference software for VVC requires significantly higher processing resources than HEVC. We can also regret that only the All-Intra configuration (intra prediction) has been considered and not the Random Access (RA) configuration (intra and inter prediction). I do not share the argument provided in the document that tree partitioning for Intra mode is more promising than for Inter mode. In RA configuration, most frames utilize inter prediction, and thus significantly impact the global encoding time. However, due to time constraints, attempting to address both intra- and inter-prediction in a PhD research project can be overly ambitious.

This research work has been published in a well-known international journal IEEE Access, with an impact factor of 3.4, which is a high value for a journal in the domain of Information and Communication Technology. Also, four papers have been published in international conferences, including the ICME workshop, a well-known IEEE conference in the field of video processing and coding. This research work has been valorized through a patent. Moreover, it must be noted that the author is very active in the normalization process, with 37 contributions to the ISO/IEC MPEG standards

3. Correctness

In this research work, a rigorous scientific process has been followed. The hypotheses are clearly motivated. The research methodology is well presented, and its description highlights the precautions taken by the author to ensure reliable and reproducible results. Mathematical modeling is used appropriately to present clearly some aspects of the contributions. The experiment results have been analyzed in detail. The precision of the two ANN has been extensively analyzed, allowing to demonstrate that no overfitting occurs. A significant effort has been made to analyze and understand what worked and what didn't, to come up with a high-quality final solution. I have appreciated the symmetry used for the description and the analysis of the two ANN and the two decision algorithms. This facilitates the reading of the manuscript

In this thesis, extensive work has been conducted to explore various solutions and identify which are efficient and which lead to poor results. A complete chapter (Chap. 9) is devoted to the exploration process. For the ANN, the architecture, hyperparameters, and ground truth augmentation have been explored to refine the proposed solutions. Nonetheless, it would have been interesting to test classical techniques to reduce the complexity of the ANN, such as pruning, quantization, and knowledge distillation. Ultimately, this results in a high-quality final solution that achieves a favorable tradeoff

between complexity reduction and quality degradation. This exploration holds significant interest for future research in this domain by identifying promising solutions.

4. Knowledge of the candidate

Chapter 2 describes briefly the HEVC standard and details the tree partitioning and the rate distortion optimization (RDO) processes. This demonstrates the deep knowledge of the HEVC standard. Nonetheless, it would be interesting to present the other coding tools available in the standard to have the big picture of HEVC and to propose a more synthetic view of the tree partitioning and the RDO processes before going into the detail and particular cases.

Chapter 2 also includes the state of the art of existing methods for efficient tree partitioning. ANN and non-ANN-based approaches are distinguished. Moreover, methods exploring the tradeoff between quality and encoding time are presented. The different types of methods are presented without going into detail about each method. This provides a good synthesis of existing works. It would be interesting to have a table summarizing the different existing methods and their main characteristics in order to be able to compare them.

These different elements confirm a general knowledge of the candidate in the discipline of Information and Communication Technology and, more especially, in video coding. More globally, this research work shows the author's expertise in video coding, standardization, machine learning and software development, and optimization. The author has demonstrated his ability to develop into a complex software framework corresponding to the HEVC reference software.

5. 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)[1], 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?

[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 **….**?

[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

Signature

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