

Poznan University of Technology
Faculty of Computing and Telecommunications
Institute of Multimedia Telecommunications

PhD dissertation

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

Analizy, rozszerzenia i optymalizacje wysokowydajnych technik kompresji sekwencji wizyjnych

Jakub Stankowski

Supervisor

dr hab. inż. Damian Karwowski

Auxiliary supervisor

dr inż. Tomasz Grajek

Abstract

This PhD dissertation presents a comprehensive investigation into the analysis, optimization, and extension of hybrid video coding (HVC) techniques, primarily (but not exclusively) focused on the High Efficiency Video Coding (HEVC) standard. As video content accounts for approximately 80% of global network traffic, the development of efficient representation and protection methods is critical for modern telecommunication infrastructure.

In Chapter 1, the analysis of video coding efficiency and statistical properties of data in the encoder has been discussed, along with practical applications of the obtained statistical knowledge.

In Chapter 2, the video encoder optimizations have been discussed. Especially, the improved Rate-Distortion Optimized Quantization (RDOQ) algorithm and optimized motion estimation method have been proposed.

Chapter 3 describes an extension of the HEVC standard to support 3D video and multiview content, particularly for Free-Viewpoint Television (FTV) systems with arbitrary camera configurations.

Chapter 4 summarizes the proposal of fast, energy-efficient watermarking algorithms that operate directly in the compressed bitstream domain for both AVC and HEVC standards.

The findings demonstrate that a deep understanding of bitstream statistics enables the creation of faster, more robust encoders and more versatile video systems. Although the experimental results mainly focus on HEVC, the methodologies and conclusions are universal to hybrid coding schemes and have been actively applied to the development of next-generation technologies like the Enhanced Compression Model (ECM) – the successor of Versatile Video Coding (VVC).

The dissertation takes the form of a collection of published, thematically related scientific articles, in accordance with Article 187, paragraph 3, of the Law on Higher Education and Science.

Streszczenie

Niniejsza rozprawa doktorska stanowi obszerne studium w zakresie analizy, optymalizacji oraz rozszerzenia technik hybrydowego kodowania wizji (HVC), skoncentrowane przede wszystkim (choć nie wyłącznie) na technice High Efficiency Video Coding (HEVC). Ponieważ treści cyfrowego obrazu odpowiadają za około 80% globalnego ruchu w sieciach teleinformatycznych, opracowanie wydajnych metod reprezentacji i ochrony tych danych ma kluczowe znaczenie dla nowoczesnej infrastruktury telekomunikacyjnej.

W rozdziale 1 omówiono analizę efektywności kodowania wizji oraz właściwości statystyczne danych w koderze obrazu, a także praktyczne zastosowania uzyskanej wiedzy statystycznej.

W rozdziale 2 przedstawiono optymalizacje kodera wizyjnego, w szczególności usprawnioną technikę Rate-Distortion Optimized Quantization (RDOQ) oraz zoptymalizowaną estymację ruchu.

Rozdział 3 opisuje rozszerzenie techniki HEVC w celu obsługi wydajnej reprezentacji treści 3D i treści wielowidokowych (Multiview), szczególnie dla systemów telewizji swobodnego wyboru punktu widzenia (FTV) o dowolnym rozmieszczeniu kamer.

Rozdział 4 zawiera propozycję szybkich i energooszczędnych algorytmów znakowania wodnego danych wizyjnych, które operują bezpośrednio w dziedzinie skompresowanego strumienia techniki AVC oraz techniki HEVC.

Wyniki badań wykazują, że dogłębne zrozumienie statystyk danych zakodowanego strumienia bitowego umożliwia tworzenie szybszych oraz bardziej wszechstronnych systemów wizyjnych. Choć wyniki eksperymentalne koncentrują się na technice HEVC, zastosowane metodyki badań i wnioski są uniwersalne dla współczesnego kodowania hybrydowego i zostały aktywnie wykorzystane przy opracowywaniu technologii nowej generacji, takiej jak Enhanced Compression Model (ECM) – potencjalny następca techniki Versatile Video Coding (VVC).

Rozprawa ma formę zbioru opublikowanych i powiązanych tematycznie artykułów naukowych, zgodnie z art. 187 ust. 3 Ustawy o szkolnictwie wyższym i nauce.

Table of Contents

Abstract I
Streszczenie II
Table of Contents III
0 Introduction 1
0.1 Scope of the dissertation, theses, and form 2
0.1.1 PhD theses and form 3
0.2 Own publications constituting the basis for the PhD dissertation 4
0.2.1 Analysis of video coding efficiency and statistical properties of data in the encoder 4
0.2.2 Video encoder optimizations 5
0.2.3 Extensions for encoding of multiview and FTV content 6
0.2.4 Compressed bitstream watermarking 7
0.3 Hybrid video coding 7
0.4 The most commonly used video compression techniques 7
0.5 Video compression concepts 9
0.5.1 Standard vs implementations 10
0.5.2 Frame types and groups of frames 11
0.5.3 Division into blocks 13
0.5.4 Prediction tools 16
0.5.5 Entropy coding and bitstream 17
0.5.6 Hybrid video coding challenges 19
1 Analysis of video coding efficiency and statistical properties of data in the encoder 21
1.1 Methodology of research 21
1.2 Analysis of frame partitioning into blocks in HEVC 22
1.3 Contribution of individual frame types into total bitrate of encoded data 23
1.4 Bitrate distribution of syntax elements in the HEVC 23
1.5 Efficiency of estimating the statistics of syntax elements in CABAC 24
1.6 Homogeneous transcoding and multi-generation encoding 25
1.7 Practical use of gathered statistical properties 27
1.7.1 Video encoder with mode decision assisted by pre-computed statistics of data 28
1.7.2 Encoder with mode decision assisted by noise reduction 29
1.7.3 Research for VVC video compression technology and its improvements 29
1.8 Chapter summary and original achievements of the research 30
2 Video encoder optimizations 32
2.1 Improved Rate-Distortion Optimized Quantization 32
2.2 Optimized motion estimation in contemporary hybrid video encoding 34
2.3 Chapter summary and original achievements of the research 37

3 Extensions for encoding of multiview and FTV content 39
3.1 Development of a multiview video codec based on HEVC technology 40
3.2 Motion and disparity vectors prediction algorithm for multiview codecs 41
3.3 Contribution to the development of a 3D video codec based on HEVC technology 42
3.3.1 Development of a distortion estimation algorithm for depth map coding 43
3.4 Development of FTV video codec with arbitrary camera location 44
3.5 Chapter summary and original achievements of the research 46
4 Compressed bitstream watermarking 48
4.1 Watermarking suite for AVC-compressed bitstreams 49
4.2 Watermark embedding algorithm for HEVC 52
4.3 Chapter summary and original achievements of the research 53
5 Summary and conclusions 55
5.1 Recapitulation 55
5.2 Original achievements of the research 55
5.3 Conclusions 57
6 References 58
6.1 Video compression standards, techniques, and implementations 58
6.2 Patents 59
6.3 Other references 60
Appendix 1 ..... A
Appendix 2 ..... B

0 Introduction

A video is a very common form of processed and transmitted data in the era of digital communication. The obvious example is digital television and its modern successors, such as IPTV and video streaming services. Moreover, a large amount of video data is generated and processed by surveillance systems, as well as by private users themselves. The research from [Cisc20] [Eric24] shows clearly that as much as 80% of all data that is transmitted through the global telecommunication network is related to video.

There is a noticeable problem here arising from the nature of video data. Its original (uncompressed) form requires an enormous number of data bits. For example, the storage requirements for one second of a video are ~300 MB for FullHD video (1920x1080p50/8bps) and ~1,500 MB for 4K video (3840x2160p50/10bps). Storage and transmission of such a huge amount of data is impractical and very expensive in terms of both financial and energy costs. For most applications, operating on such spacious sets of data poses serious technical difficulties.

The obvious conclusion is that efficient representation of a video is necessary. Before transmitting the video or saving it on the device's memory, the video data must be compressed in order to significantly reduce the size of the data describing it. Video compression is, therefore, an important element of contemporary digital communication systems. Moreover, efficient processing of compressed video (i.e., watermarking of compressed video data) is also considered an important scientific and practical problem. The PhD candidate's research described in this thesis was devoted to these problems. The scope of this research was efficient algorithms of video compression and efficient watermarking of compressed video datastreams.

Many different video compression techniques and algorithms have been developed over the years [Karw16][Isna07][Rich24]. However, from the point of view of wide practical application, the typical approach is often based on exploiting similar phenomena, such as exploiting local correlation of pixel values in a natural image (single frame of a video), exploiting temporal redundancy of successive images in a video, and the possibility of removing some information without the introduction of noticeable distortion (realization of lossy coding strategies). Over the years, many different compression techniques have been developed. Some of them used DPCM with Huffman codes [Huffyuv], some were based on simple concatenation of intra-coded pictures (Motion JPEG [RFC2435], Motion JPEG 2000 [MJP2K] or Animated PNG [APNG][PNG]), while others used wavelet-based transforms [MJP2K] or less typical approaches like 3D-DCT [Ichi99] and fractal compression [Ali92]. However, several years of development of video compression techniques resulted in the most successful approach called hybrid video coding (HVC).

For several decades now, hybrid video coding has been the basis for a significant part of global research in the field of efficient image representation. This research is conducted by teams of scientists, but also

by expert groups from international organizations. Subsequent generations of hybrid technology have been included in numerous international video coding standards.

The history of standardized hybrid video compression techniques dates back to H.261 [H261] developed by the International Telecommunication Union (ITU-T) in 1988 and MPEG-1 Part 2 [MPEG1] developed by International Organization for Standardization (ISO/IEC) in 1991. Both techniques share some common ideas, like the usage of discrete cosine transform (DCT), motion-compensated prediction, and entropy coding. Those two early standards set the direction for further development of video compression techniques.

In the following years, both standardization committees (ISO/IEC and ITU-T) have repeatedly established cooperation in order to work together on new techniques. Examples of such a common work are MPEG-2 Part 2 [MPEG2] / H.262 [H262], MPEG-4 Part 10 [AVC] / H.264 [H264], MPEG-H Part 2 [HEVC] / H.265 [H265] or MPEG-I Part 3 [VVC] / H.266 [H266].

Each new technology was a significant breakthrough in video compression and introduced new and improved coding tools. However, their general operating principles have not changed and are shared among all the above-mentioned techniques.

This doctoral thesis is part of this line of research, focusing on improvements in the field of hybrid video coding.

0.1 Scope of the dissertation, theses, and form

The scope of this dissertation is strongly related to hybrid video coding and constitutes a summary of the author`s research and development activities over several years. The dissertation describes a logical research sequence: from fundamental analysis of contemporary video compression technology, through its optimization, then its extension to new applications (multi-view video), to issues related to practical usability of the compression-technique-based systems (watermarking).

The High Efficiency Video Coding (HEVC) standard was finalized and published by ITU-T and ISO/IEC on 7 June 2013 and 25 November 2013, respectively [H264][HEVC]. During the development stage and a few years after the final standard release, HEVC was a new technology with unknown properties and characteristics. Only some properties were measured during development, i.e., encoder/decoder computational complexity, entropy encoding efficiency. The most significant shortage was a lack of information about the statistical properties of the HEVC bitstream and frame (image) partitioning into blocks, which the video encoder is doing for subsequent images of the video.

Most research results described in this dissertation are related to HEVC, due to the time correlation with HEVC development and early adoption phases.

While VVC has been standardized recently, HEVC remains the dominant standard in commercial deployment as of 2025 [Bitm25], making these optimizations critically relevant for a significant share of current infrastructure.

0.1.1 PhD theses and form

This doctoral dissertation is a collection of published, thematically related, and peer-reviewed scientific articles, in accordance with the provisions of the Polish Law on Higher Education and Science (*"Prawo o szkolnictwie wyższym i nauce" w Polsce*), Article 187, paragraph 3.

The subject of this doctoral dissertation is the author's research, which addresses important scientific issues of hybrid video compression. To organize the list of the author's achievements presented in this dissertation, the collection of scientific articles cited in the dissertation has been supplemented with a summary of the most important achievements and a summary of key conclusions from the conducted research.

The detailed scope and objectives of the dissertation stem directly from the three adopted research theses:

PhD Thesis 1:

Statistical analysis of data contained in the encoded video bitstream, together with information about the encoder's decisions regarding the compression modes used, allows for the development of video encoder control algorithms that significantly reduce its computational complexity while still maintaining high compression efficiency of a video.

The original solution to a scientific problem and the main contributions of the dissertation, related to Thesis 1, are contained in Chapters 1 and 2.

PhD Thesis 2:

Using the knowledge of the geometry of the coded scene, it is possible to achieve such an improvement in the efficiency of coding motion information in a 2D video sequence and information about the disparity of subsequent points of two indicated views of the 3D video, which will result in a reasonable improvement in the compression efficiency of the multi-view video.

The original solution to a scientific problem and the main contributions of the dissertation, related to Thesis 2, are contained in Chapter 3.

PhD Thesis 3:

When operating in the compressed video bitstream, it is possible to apply such a modification of the scaling matrices values or a modification of the transform coefficients data of the image sample prediction error that enables efficient and robust watermarking of the compressed video, with computational efforts many times lower than those of methods operating in the image pixels domain.

The original solution to a scientific problem and the main contributions of the dissertation, related to Thesis 3, are contained in Chapter 4.

0.2 Own publications constituting the basis for the PhD dissertation

The list of the author`s own publications includes the following parameters:

- P_p — Ministry of Science and Higher Education points (as of the date of publication) related to the maximum possible number of Ministry points at that time. For pre-2019 publications, the maximum score was 50; since 2019, the maximum score is 200.
- P_n — Ministry of Science and Higher Education points (as of February 2026) related to the maximum possible number of Ministry points at February 2026.
- IF — impact factor (as of February 2026)
- C — number of citations (as of February 2026, according to Google Scholar)

0.2.1 Analysis of video coding efficiency and statistical properties of data in the encoder

A1. **J. Stankowski**, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, and M. Domański. "Analysis of Frame Partitioning in HEVC." In: Computer Vision and Graphics. Ed. by L. J. Chmielewski, R. Kozera, B.-S. Shin, and K. Wojciechowski. Cham: Springer International Publishing, pp. 602–609, 2014.
[Pp= 15/50, Pn=20/200, C=5]
Own contribution 80% – significant role in development of the method, research conceptualization, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

A2. **J. Stankowski**, D. Karwowski, T. Grajek, K. Wegner, J. Siast, K. Klimaszewski, O. Stankiewicz, and M. Domański. "Analysis of Compressed Data Stream Content in HEVC Video Encoder." In: International Journal of Electronics and Telecommunications 61.2, pp. 121–127, 2015.
[Pp=15/50 (list B), Pn=70/200, C=7]
Own contribution 80% – significant role in development of the method, research conceptualization, research methodology development, algorithm development, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

A3. **J. Stankowski**, D. Karwowski, T. Grajek, K. Wegner, J. Siast, K. Klimaszewski, O. Stankiewicz, and M. Domański. "Bitrate distribution of syntax elements in the HEVC encoded video." 2014 International Conference on Signals and Electronic Systems (ICSES), pp. 1–4, Poznań, Poland, 2014.
[Pp=15/50, C=15]
Own contribution 80% – significant role in development of the method, research conceptualization, research methodology development, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

A4. **J. Stankowski**, D. Karwowski, K. Wegner, O. Stankiewicz, K. Klimaszewski, and T. Grajek. "Analysis of CABAC context models efficiency in the HEVC." In: 2015 57th International Symposium ELMAR (ELMAR), pp. 217–220, Zadar, Croatia, 2015.
[Pp=15/50]
Own contribution 80% – significant role in development of the method, research conceptualization, research methodology development, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

A5. **J. Stankowski**, T. Grajek, K. Wegner, M. Domański, "Video Quality in Multiple HEVC Encoding-Decoding Cycles", 20th International Conference on Systems, Signals and Image Processing, IWSSIP 2013, pp. 75-78, Bucharest, Romania, 2013.
[P_p= 15/50, C=11, Publication granted a Kalman Fazekas special award]
Own contribution 60% – research conceptualization, research methodology development, creating the software, conducting experiments, collecting and processing results, and working on the text of the paper.

A6. T. Grajek, **J. Stankowski**, D. Karwowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, "Analysis of video quality losses in the homogeneous HEVC video transcoding", IEEE Access, Vol. 7, pp. 96764 – 96774, 2019.
[Pp=Pn=100/200, IF=3.6, C=4]
Own contribution 20% – co-author of the research idea on homogeneous transcoding, implementation of research scenarios, conducting experimental verification, and working on the text of the paper.

A7. K. Wegner, D. Karwowski, **J. Stankowski**, T. Grajek, K. Klimaszewski, and O. Stankiewicz. "Fast mode selection in the high-efficiency video coding intravideo encoder based on statistics of modes." In: Journal of Electronic Imaging 27.4, pp. 1–13, 2018.
[Pp=20/50 (list A), Pn=40/200, IF=1.0]]
Own contribution 35% – research conceptualization, research methodology development, implementation, and testing of the encoder, involvement in collecting and analyzing results, and proofreading of the paper.

0.2.2 Video encoder optimizations

A8. **J. Stankowski**, C. Korzeniewski, M. Domański, T. Grajek, "Rate-distortion optimized quantization in HEVC: performance limitations", 31st Picture Coding Symposium, PCS 2015, Cairns, Australia, pp. 85-89, 2015.
[Pp=20/50, Pn=20/200, C=40]
Own contribution 30% – research and algorithm conceptualization, supervision of algorithm implementation and experiment execution, conducting experimental verification, editing, and proofreading of the paper.

A9. **J. Stankowski**, D. Karwowski, K. Klimaszewski, K. Wegner, O. Stankiewicz, T. Grajek, "Analysis of the complexity of the HEVC motion estimation", 23rd International Conference on Systems, Signals and Image Processing, IWSSIP 2016, pp. 1-4, Bratislava, Slovakia, 2016.
[Pp=20/50, C=8]
Own contribution 85% – significant role in development of the method, research conceptualization, research methodology development, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

A10. **J. Stankowski**, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, "Computational complexity tradeoffs in HEVC motion estimation", International Conference on Signals and Electronic Systems, ICSES 2016, pp. 124-128, Kraków, Poland, 2016.
[Pp=15/50]
Own contribution 85% – significant role in development of the method, conceptualization, research methodology development, creating the software, conducting experiments, collecting and processing results, analyzing the results, and working on the text of the paper.

0.2.3 Extensions for encoding of multiview and FTV content

A11. **J. Stankowski**, Ł. Kowalski, J. Samelak, M. Domański, T. Grajek, K. Wegner, "3D-HEVC Extension for Circular Camera Arrangements", 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 3DTV-Con 2015, Lisbon, Portugal, 2015.
[Pp=15/50, C=23]
Own contribution 25% – contribution to the development of concepts and algorithms, software co-development, contribution to experiment planning, and results analysis.

A12. **J. Stankowski**, M. Domański, O. Stankiewicz, J. Konieczny, J. Siast, K. Wegner, "Extensions of the HEVC Technology for Efficient Multiview Video Coding", 19th IEEE International Conference on Image Processing (ICIP 2012), pp. 1317-1320, Orlando, Florida, U.S.A., 2012.
[Pp=15/50, Pn=70/200, C=16]
Own contribution 80% – significant role in development of the method, research conceptualization, research methodology development, algorithms development, creating software, conducting experiments, collecting and processing results, analyzing results, and working on the text of the paper.

A13. M. Domański, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, **J. Stankowski**, R. Ratajczak, T. Grajek, "High Efficiency 3D Video Coding Using New Tools Based on View Synthesis", IEEE Transactions on Image Processing, Special Issue on 3D Video Representation, Compression, & Rendering, Vol. 22, No. 9, pp. 3517-3527, 2013.
[Pp = 45/50 (list A), Pn=200/200, IF=13.7, C=57]
Own contribution 20% – contribution to algorithm development, contribution to software development, and experimental evaluation.

0.2.4 Compressed bitstream watermarking

A14. **J. Stankowski**, M. Domański, T. Grajek, "Fast and Energy-Efficient Watermarking of HEVC-Compressed Video Bitstreams", 30th European Signal Processing Conference EUSIPCO, Belgrade, Serbia, 2022.
[Pp=Pn=70/200, C=1]
Own contribution 90% – significant role in development of the method, research conceptualization, research methodology development, algorithms development, creating software, conducting experiments, collecting and processing results, analyzing results, and working on the text of the paper.

A15. **J. Stankowski**, T. Grajek, M. Domański, "Fast watermarking of MPEG-4 AVC/H.264 encoded HDTV video bitstreams", 29th Picture Coding Symposium, PCS 2012, pp. 265-268, Kraków, Poland, 2012.
[Pp=20/50, Pn=20/200, C=8]
Own contribution 70% – significant role in development of the method, development of watermark embedding and detection algorithm, cooperation with the development of software, collecting and processing results, and working on the text of the paper.

A16. J. Siast, **J. Stankowski**, T. Grajek, M. Domański, "Digital Watermarking with Local Strength Adjustment for AVC-Compressed HDTV Bitstreams", 30th Picture Coding Symposium, PCS 2013, pp. 53-56, San Jose, CA, USA, 2013.
[Pp=20/50, Pn=20/200, C=4]
Own contribution 40% – cooperation on the watermark embedding algorithm, development of the watermark detection algorithm, and working on the text of the paper.

0.3 Hybrid video coding

Hybrid video codecs can be characterized as an amalgam of several coding tools. A typical hybrid video codec includes coding tools like: predictive coding, transform coding, quantization (as the sole source of loss in lossy compression), entropy coding, etc. Over the years of development, some of those tools have been improved, and some new tools have been introduced (i.e., loop filters). However, the general coding paradigm is unchanged.

One can also notice that hybrid video encoders share a common scheme of incorporating the inner decoding loop. This means that each encoder contains virtually the entire decoder (with the exception of the entropy decoder). The example of such an encoder is shown in Figure 1.

0.4 The most commonly used video compression techniques

While early techniques such as MPEG-1 Part 2 (Video) [MPEG1] had some limited usage (i.e., Video-CD), the first video compression standard commonly adopted worldwide was MPEG-2 Video. The MPEG-2 Video [MPEG2] / H.262 [H262] was first in a long line of video compression technologies developed cooperatively by the International Organization for Standardization (ISO/IEC) and the International Telecommunication Union (ITU). It was finalized in 1994 and was used by the first wave of digital television (DVB) and was used by Digital Video Disk (DVD) format, which led to intensive development of decoders and encoders (both as software and hardware implementations). Nowadays,

the MPEG-1 and MPEG-2 / H.262 can be considered as legacy standards and have almost disappeared from the market.

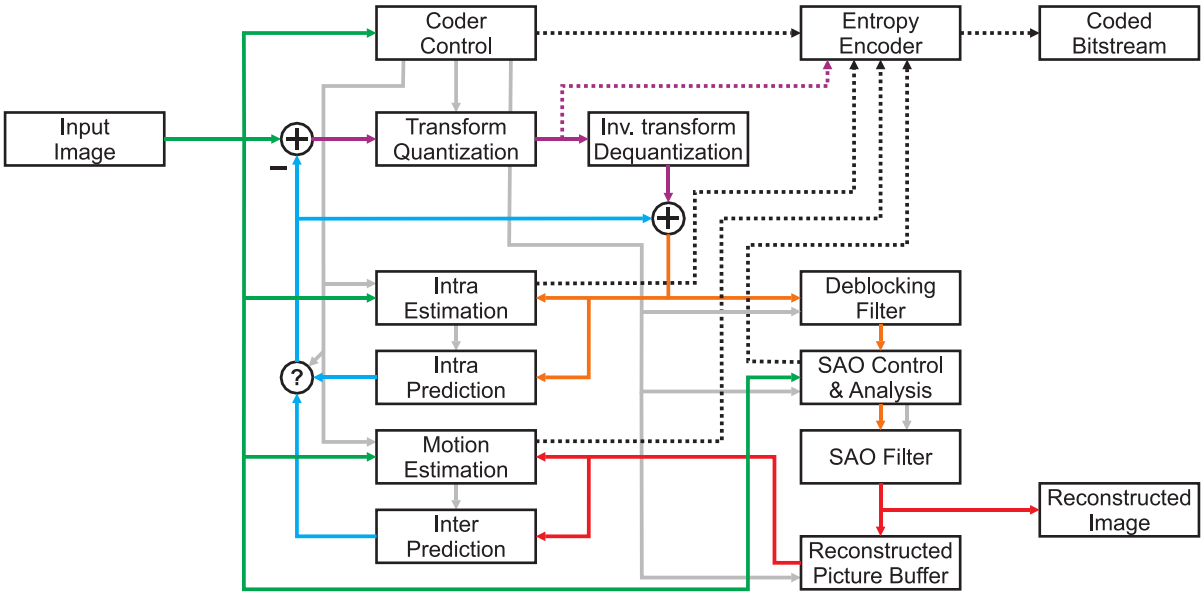


Figure 1. HEVC encoder block diagram.

The following standard – Advanced Video Coding (AVC) [Wieg03] – MPEG-4 Part 10 [AVC] / H.264 [H264] was formally created by the Joint Video Team (JVT), a common body established by the Video Coding Experts Group (VCEG – ITU-T SG16 Q.6) and the Moving Picture Experts Group (MPEG – ISO/IEC JTC 1/SC 29/WG 11) and finalized in 2004. The notable use cases of AVC were once again digital television (DVB) and optical discs (Blu-ray). However, when compared to its predecessors, the most important aspect is wide usage in IP-based systems like video surveillance, Video-on-Demand (VOD), and online video platforms (i.e., YouTube). The AVC is still actively used and maintains noticeable market share [Bitm24][Bitm25]. Moreover, the AVC is now treated as a technique that is compatible with almost any equipment and software.

The High Efficiency Video Coding (HEVC) [Sull12] is another video compression technology developed cooperatively by the International Organization for Standardization (ISO/IEC) and the International Telecommunication Union (ITU). Both organizations established a joint body called Joint Collaborative Team on Video Coding (JCT-VC) and delegated JCT-VC with the task of developing new technology. Finally, HEVC was standardized in 2013 as MPEG-H Part 2 [HEVC] and H.265 [H265] by ISO/IEC and ITU, respectively. The HEVC market adoption resembles a similar pattern – it is used in digital television, video surveillance, and Video-on-Demand (VOD). The adoption of online video platforms is limited due to patent- and licensing-related issues.

The most recent standardized technique is Versatile Video Coding (VVC), developed by Joint Video Experts Team (JVET) – another common body established by VCEG (ITU-T SG16) and the MPEG (ISO/IEC JTC 1/SC 29). The VVC is described as MPEG-I Part 3 [VVC] and H.266 [H266] and was finalized in 2020. Due to being relatively recent (and, once again, patent- and licensing-related issues), the VVC market adoption is almost unnoticeable [Bitm25].

Nowadays, the Joint Video Experts Team (JVET) is working on the development of the VVC successor. The development is carried out under the name Enhanced Compression Model (ECM) [ECM], however, this name is more related to software than standard.

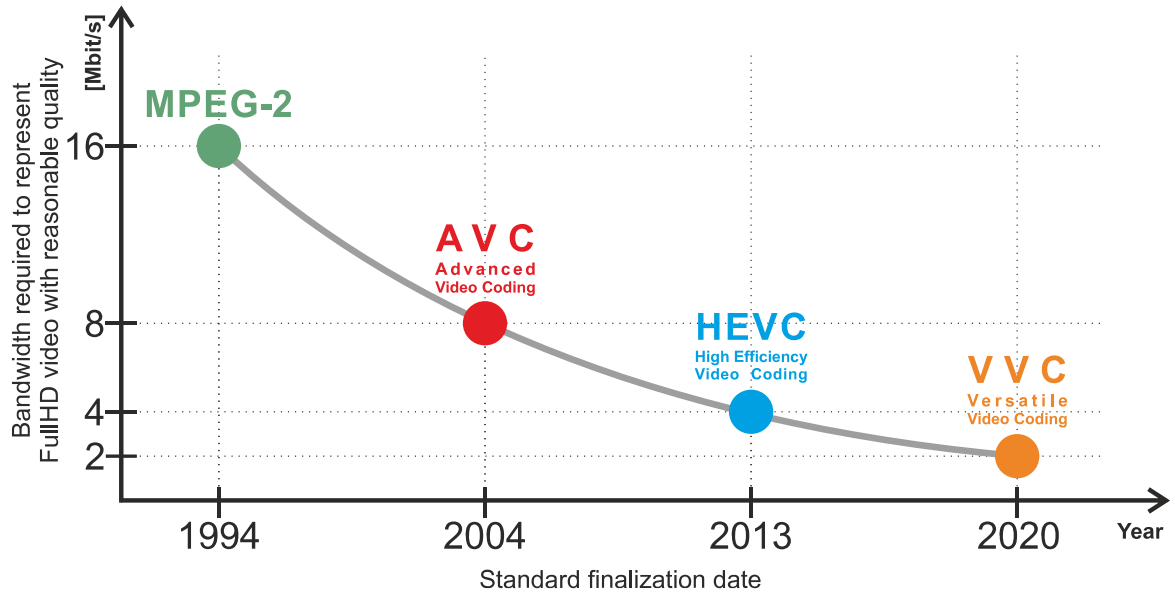


Figure 2. Most important video codecs standardized by ISO/IEC, together with approximated bandwidth required to represent FullHD video with reasonable quality.

Generally, it can be assumed that each compression technique offers twofold improvement in coding efficiency when compared to its predecessor. Simply speaking, it can achieve a similar quality as its predecessor while using half the number of bits to represent video data (Figure 2). The compression efficiency improvement is a result of about 9 years of intensive development in the video (and picture) compression field between each standard [Karw16].

In addition to compression efficiency improvements, some techniques resolve several implementation difficulties caused by previous techniques. The most notable example is HEVC addressing high computational complexity and a lack of a practical approach during the development of AVC (difficult software and hardware implementation) [Vann12].

0.5 Video compression concepts

The list of terms and concepts related to video compression is very long. However, some of them are related to research described in the following chapters; therefore, some selective highlights of those concepts will be described in the following subsections.

0.5.1 Standard vs implementations

Each modern video compression standard reassembles the same pattern. The only parts covered by standardization are:

- bitstream (syntax and semantics),
- decoding process,
- constrains (profiles, levels).

It is important to notice that standards do not cover encoder operation. Therefore, the creator of the encoder has broad freedom in choosing the possible approaches. For example, the author of the encoder can avoid using some coding tools due to complexity reasons. The only important factor is to create a conformant bitstream.

This leads to a distinction between prospective coding performance offered by a standard and coding performance achieved by implementation. For most video compression standards, we can find several encoder implementations differing by many factors. Example implementations include:

- hardware/software/hybrid encoders for online broadcasters,
- software (mostly) implementations for offline mastering,
- hardware implementations for mains-powered consumer devices (i.e., integrated in GPUs),
- hardware implementations for mobile devices (i.e., integrated in smartphone SoCs).

All of them have different operating points dictated by complexity restrictions, available power & thermal environment, real-time processing requirements, etc. Therefore, the codec implementations developed for end users are not suitable for assessing the improvements in coding performance. For example, it is not surprising that an energy-constrained real-time HEVC encoder in a mobile device can achieve lower compression efficiency than a software-based offline encoder using a high-power multi-core server CPU.

Due to the above-mentioned issues, most video compression research is performed using the so-called “reference implementation”. The standardization committees used to maintain both the standard (text) and its reference implementation (software) and keep both of them synchronized.

The reference implementations are available for all techniques since MPEG-1 (i.e., [JM] [HM][VTM][ECM]) and share similar properties. The reference implementations focus on conformance with the standard and achieving the highest possible compression efficiency at the cost of slow operation (sometimes extremely slow).

The experiments conducted during the development of each technique are performed using the reference implementation and according to rules and scenarios described by so called common test conditions. For example, the development of HEVC was based on HM software [HM] and Common test conditions and reference configuration [L1100].

For any research related to video compression, the usage of reference software and conditions defined in CTC is an obvious choice, allowing to avoid questions about biased setup and simplifying comparison with the reference.

0.5.2 Frame types and groups of frames

There are three most common frame types used in hybrid video coding:

- I-frame – intra coded only,
- P-frame – allowing for inter prediction (motion compensated prediction) while assuming a single prediction source per encoded unit,
- B-frame – allowing for inter prediction with two prediction sources per encoded unit (bi-predictive).

Since P- and B-frames allow for inter-frame prediction, the encoder and the decoder must agree about prediction sources. Therefore, each P and B frame is encoded using pictures coming from the so-called reference picture list(s). Those lists contain pointers to selected reference pictures and are derived by the decoder or explicitly transmitted in the bitstream.

The I, P, and B frames are organized in Groups of Pictures (GOP). Since the GOP is considered a single Access Unit, it should contain at least one I frame to allow random access decoding of the video sequence (Figure 3).

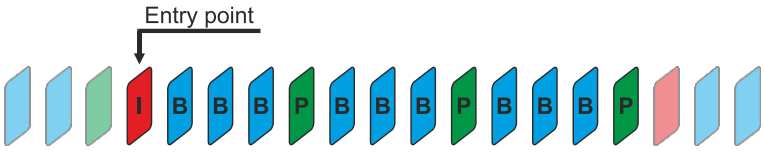


Figure 3. Access unit entry point.

The structure of GOP can be different depending on the use case and encoder decision. However, it is worth noting that there are two common approaches to GOP structure. The first (retrospectively called "classic GOP") assumes that only I and P frames are used as a prediction source for B frames, and all B frames are leaves of coding structure (as shown in Figure 4).

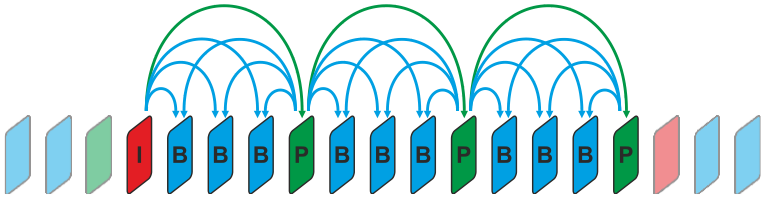


Figure 4. Example of the so-called classic GOP structure.

The classic GOP coding structure is far from optimal; therefore, some AVC encoders (i.e., x264 [x264]) began to use a different structure called hierarchical GOP. In a hierarchical structure, there are no restrictions on prediction sources for B frames, and some B frames can use other B frames as a prediction source. These changes allow for a noticeable increase in compression efficiency and (if GOP is carefully shaped) temporal scalability. Those approaches became more widespread, and hierarchical GOP was used in HEVC Common Test Conditions (CTC) [L1100]. The simplified example of a hierarchical GOP structure is shown in Figure 5.

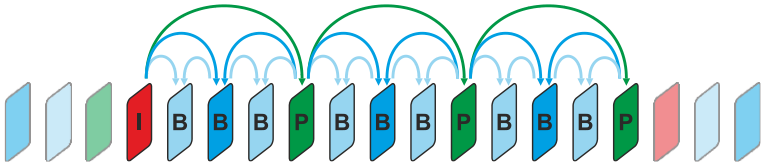


Figure 5. Example of hierarchical GOP structure.

The next step in GOP structure modification is the idea of Generalized B Pictures (GBP). One can notice that P pictures from the example shown in Figure 5 can be replaced by B pictures. Experiments conducted by JCT-VC show that the usage of B frames (even if both reference picture lists contain the same reference pictures) is beneficial from a compression efficiency point of view. Those results in the GOP shape shown in Figure 6.

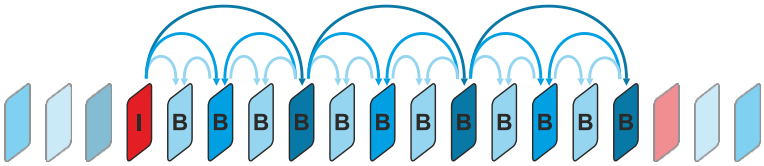


Figure 6. Example of hierarchical GOP structure with GBP.

The combination of hierarchical GOP structure and GBP created a noticeable hierarchy of B frames. In the case of Figure 6, one can distinguish 3 types (or levels, perhaps) of B frames. It can be named B_0, B_1, and B_2 for dark blue, neutral blue, and light blue, respectively.

The inter-frame prediction efficiency is strongly correlated with the temporal distance between the predicted frame and its reference picture. Generally, the closer the distance, the better the prediction signal. When looking at Figure 6, one can notice that GOP`s last B frame closely neighbors with the next group`s I frame, but does not use it as a prediction source. This once again leaves some prospective coding performance on the table.

The GOP structure shown in Figure 6 is called a closed GOP. In opposition to the closed GOP, a new structure called the open GOP was proposed. In this case, the last ("closing") P or B frame is omitted, and the I frame from the successive GOP is used as a prediction source, as shown in Figure 7.

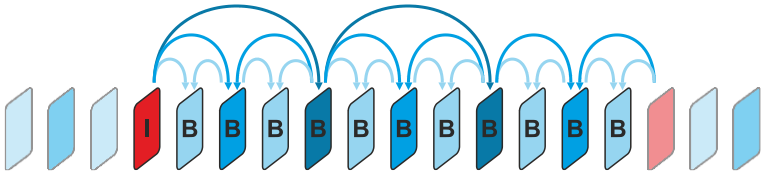


Figure 7. Example of a hierarchical open GOP structure with GBP.

Unfortunately, the idea of open GOP cannot be used in older techniques (i.e., MPEG-2, AVC) due to syntax limitations. For example, in AVC, the frame beginning GOP needs to have some special properties (IDR – Instantaneous Decoder Refresh)[Wieg03] and forces flushing of reference picture lists. Therefore, HEVC and the following standards define several types of decoder refresh methods, i.e., Instantaneous Decoder Refresh (IDC) or Clear Random Access (CRA) [Sull12].

The GOP shape evolution described above led to the GOP used by HEVC Common Test Conditions (CTC) [L1100]. CTC defines the B-frame pyramid of eight picture length and with four B-frame levels (B_0, B_1, B_2, B_3). Obviously, the same structure was used during experiments described in this thesis, especially those described in Chapter 1.

0.5.3 Division into blocks

All hybrid video codecs operate by splitting a frame into rectangular blocks and transmitting the encoded content in a block-by-block manner. Since H.261, the main coding unit was called a macroblock and covered an area of 16x16 pixels. The usage of macroblocks was continued in MPEG-2 Video and AVC. In all mentioned techniques, the decision about coding mode (intra, inter P, inter B) or DCT (discrete cosine transform) size (4x4 or 8x8 in AVC) was possible for the entire macroblock only.

HEVC broke with that approach and introduced a significantly more flexible partitioning scheme [Kim12]. At first, it introduced a new highest-level unit called Coding Tree Unit (CTU) (see Figure 8). The size of CTU is not fixed. Encoder can choose between 16x16, 32x32, and 64x64 CTUs; however, the CTU size has to be constant for the entire sequence. The typical CTU size is 64x64, and this value has been chosen in HEVC Common Test Conditions (CTC).

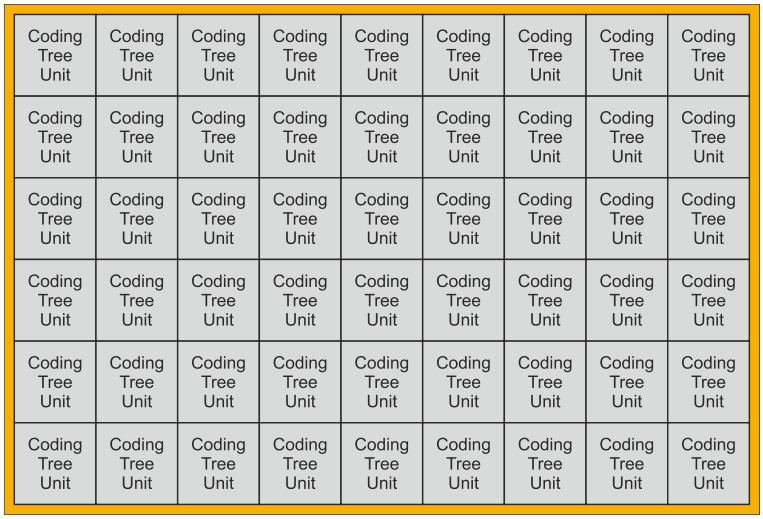


Figure 8. Example of a picture divided into Coding Tree Units.

The CTU contains a quadtree structure of Coding Units (CU). In HEVC, the CU is where the main decision about coding mode (intra, inter P, inter B) is taken. CUs are square units, with four possible sizes (64x64, 32x32, 16x16, and 8x8). Each CTU contains at least one CU, and each CU bigger than 8x8 can be subdivided into four smaller CUs, as illustrated in Figure 9. The example of the CTU division has been presented in Figure 10.

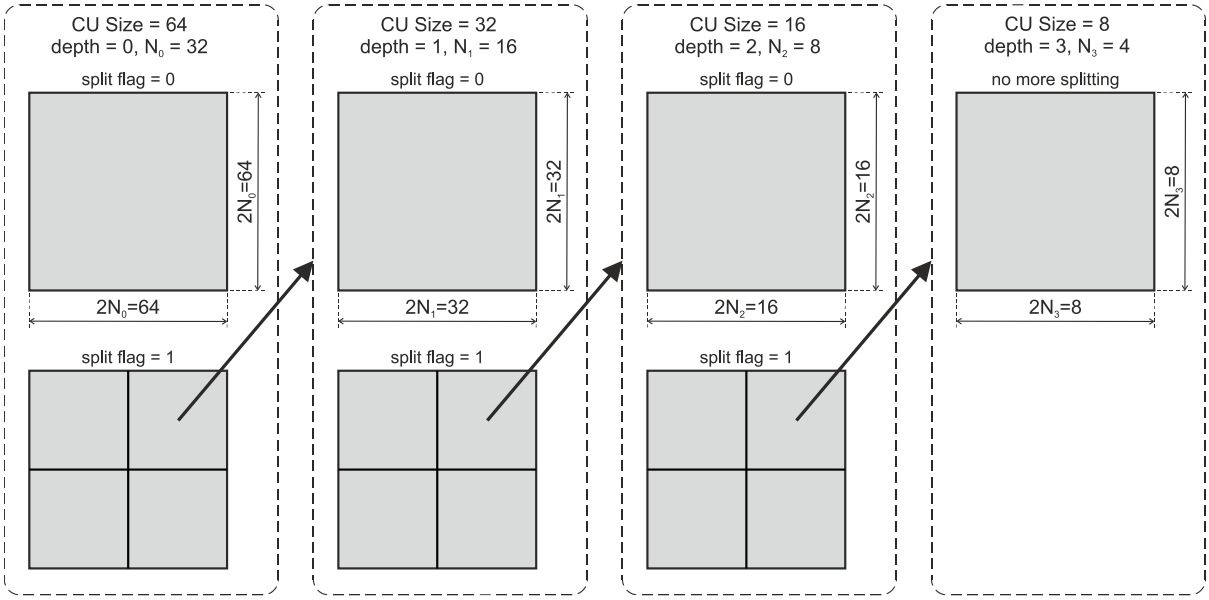


Figure 9. The illustration of possible Coding Unit sizes and divisions.

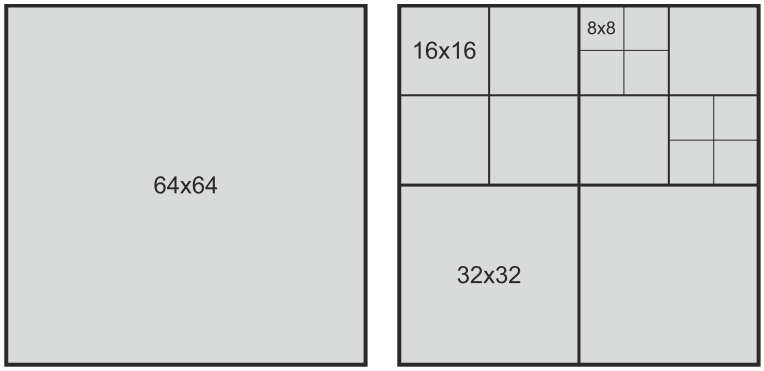


Figure 10. Two examples of possible CTU division into CU(s).

The Coding Unit contains additional structures: one or more Prediction Unit(s) (PU) and (optionally) one or more Transform Unit(s) (TU) organized in Residual Quad Tree (RQT). The PU contains information related to intra or inter prediction (i.e., prediction direction, motion field, etc.). Depending on CU size and prediction type (intra or inter), the CU can contain a different number (and shapes) of prediction units (see Figure 11). In HEVC, the PU is the only unit that can be rectangular.

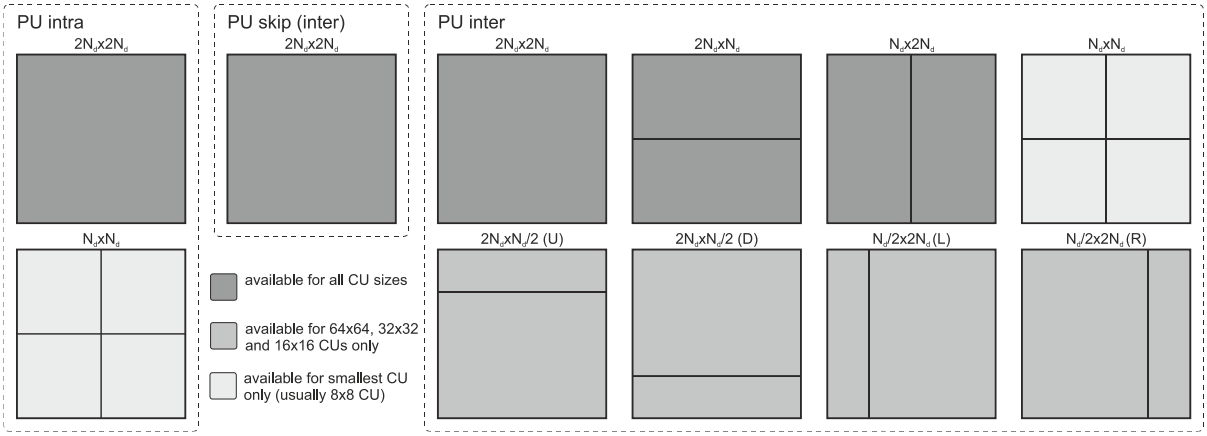


Figure 11. Illustration of available Prediction Units.

Older techniques like H.261, MPEG-1, MPEG-2 used only one DCT transform size – 8x8. AVC initially used 4x4 DCT transform only, but after evidence showed that 4x4 DCT is inefficient for high resolution content, it retreated to 8x8 DCT in High profile. When compared to older techniques, HEVC allows for greater flexibility when it comes to residual coding [Sole12].

The residual data is transmitted within the Residual Quad Tree (RQT) containing the Transform Unit(s). The presence of RQT is mandatory for intra-coded CUs, optional for most inter-coded CUs, and prohibited for special inter mode – the skip mode. The TUs contain quantized transform coefficients and allow for choosing between four sizes (32x32, 16x16, 8x8, and 4x4). Moreover, in some intra modes, the DCT was replaced by DST (discrete sine transform). The Transform Unit subdivision possibilities are illustrated in Figure 12.

The research described in Section 1 contains statistical analysis related to picture partitioning into CUs, PUs, and TUs.

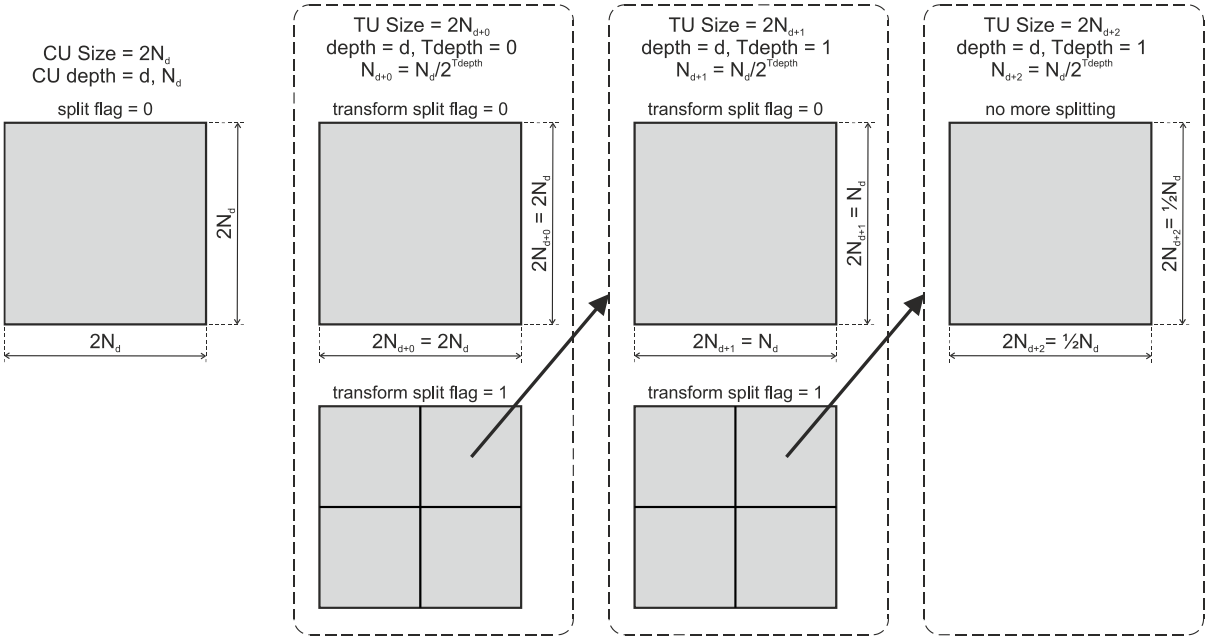


Figure 12. The illustration of possible Transform Unit sizes and divisions.

It is worth mentioning that the HEVC successor, VVC, followed those flexible subdivision schemes, however, with slight differences like the usage of a binary tree and the possibility of rectangular Coding Units [Bros21].

0.5.4 Prediction tools

The prediction tools in hybrid video codecs can be divided into two categories: intra and inter. Intra prediction exploits local correlation or region similarity within a single image. Inter prediction is based similarity between different frames.

In common sense, the inter-frame prediction is associated with motion compensated prediction and exploits similarities between consecutive frames in sequence. In such a case, the encoder maintains a list of reference pictures (used as prediction source) and uses motion vectors[1] to signal location on reference pictures [Sjob12]. The reference pictures are selected from frames that have already been transmitted to the decoder. Since pictures are reordered in time, the encoder can use reference frames from the past and future (see B frames in Figure 7).

Direct transmission of motion vectors would be inefficient; therefore, modern codecs encode motion vectors predictively and contain sophisticated predictors for both motion vector values and reference

[1] It has to be noticed that the name “motion vector” could be a bit misleading. It is definitely true, that the idea of motion compensated prediction is to track object movement and use areas corresponding to the same object as a prediction source. In the real world, the analysis of encoder decisions showed that those vectors are loosely correlated to real motion occurring in sequence. Author of this thesis thinks that the name “vectors pointing to the best prediction source in terms of rate distortion optimization” would be more precise… although “motion vector” is more widely recognized and slightly shorter. Therefore, “motion vector” will be used in the following text.

frame indices. Figure 13 shows possible prediction sources used by the HEVC codec to predict motion vectors.

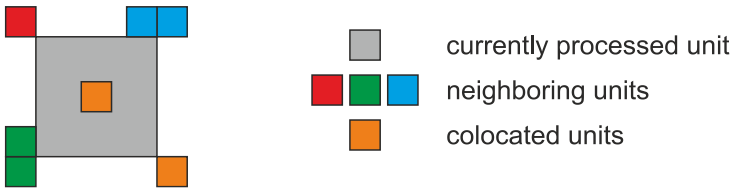


Figure 13. Example of prediction sources used to predictively encode motion vectors in HEVC. The collocated unit corresponds to the same location in 1st reference frame.

The motion vectors are determined by an encoder stage called motion estimation. In most cases, the motion estimator is the most demanding part of the encoder [Boss12] when computational complexity and memory bandwidth are considered. Therefore, optimizations and improvements to motion estimation are critical for the design of efficient encoders (Section 2.2).

The inter prediction can be divided into two variants: uni- and bi-prediction. The uni-prediction (associated with P frames) assumes that only one prediction source is used. In the case of bi-prediction, the encoder chooses two reference blocks, which are combined (averaged) into a single prediction signal. The bi-prediction offers a better prediction signal, but on the other hand, it introduces a significant increase in computational burden and memory bandwidth demand when compared to uni-prediction.

In the multiview codec [MVC][3DHEVC], the inter prediction can also be considered as an interview prediction. In this case, the tool should be called disparity compensated prediction, and the prediction source is pointed by the disparity vector. The study related to improvements in motion- and disparity-vector prediction in the multiview HEVC codec is described in Section 3.

0.5.5 Entropy coding and bitstream

Each video coding technology standard needs to exactly define bitstream syntax and semantics. There are different approaches to design bitstream syntax, with the Network Abstraction Layer (NAL) being a prominent example. The NAL idea was introduced by the AVC standard. NAL defines several syntax units with clear division into units containing video data (Video Coding Layer – VCL) and units containing headers and auxiliary information (non-VCL). The NAL idea was then refined in HEVC and VVC standards.

In the case of HEVC, the main non-VCL units are Video Parameter Set (VPS), Sequence Parameter Set (SPS), and Picture Parameter Set (PPS). There are also several types of units containing video data, divided into slices. Some types of VCL units have special meaning or impose restrictions. For example, IDR and CRA units control the decoder refresh process (end reference picture list flushing) and allow for intra-coded CUs only.

The VCL units contain compressed video data produced by an entropy encoder, which is the mandatory final stage of each hybrid video encoder. Historically, most standardized video encoders (H.261, MPEG-1, MPEG-2) used entropy coding based on predefined tables containing variable length codes (VLC). Those tabularized codes closely reassemble Huffman codes [MPEG2]. AVC was the first to introduce an advanced entropy encoder based on a binary arithmetic encoder. This codec, called CABAC (Context Adaptive Binary Arithmetic Encoder introduced significant improvement in compression efficiency [Marp03][Oste04]. On the other hand, CABAC has significantly higher computational complexity and is difficult to parallelize. Since CABAC is based on a binary arithmetic encoder, it requires an additional stage called binarization in order to transform non-binary symbols into a form acceptable by the binary arithmetic core. The generalized block diagram of CABAC is shown in Figure 14.

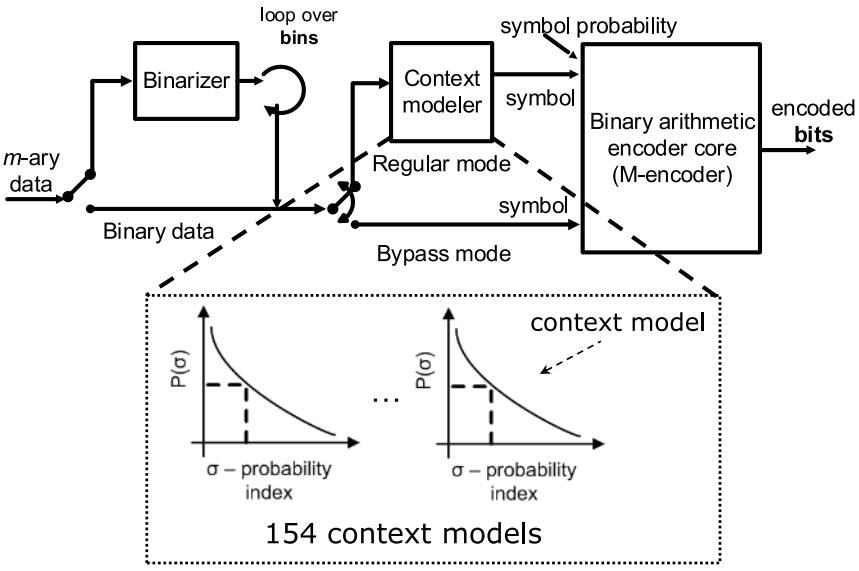


Figure 14. Block diagram of CABAC. Figure from [A4].

Due to computational complexity concerns, the AVC standard defines two entropy coding techniques: a VLC-based technique with adaptive switching of predefined code tables – CAVLC (Context Adaptive Variable Length Coding) and arithmetic coding based CABAC [AVC, Wieg03].

HEVC significantly refined CABAC [Sze12], reducing its computational complexity and facilitating parallel decoding in hardware implementation. The improvements to CABAC throughput were so significant that they resulted in dropping the CAVLC-derived technique (which was developed in parallel with improved CABAC).

CABAC itself consists of two complementary parts – the binary arithmetic encoder core (BAC) and the probability modeling stage (CA). The arithmetic core processes two types of bins:

- regularly coded bins being the subject of probability modeling,

- bypass coded bins with assumed equal probability of both symbols ($P(0) = 50\%$ and $P(1) = 50\%$).

Bypass coded bins are used to unpredictable syntax elements (i.e., transform coefficient sign) or in cases of such types of syntax elements where complex probability modeling is unprofitable (i.e., higher values of transform coefficients of motion vector differences).

The probability modeling stage uses 152 probability estimators. The probability of some syntax elements is modeled by several estimators, and those estimators are selected based on surrounding conditions (context). Probability estimators are a state machine with predefined state changes depending on encoded symbol values (see Figure 14).

The accuracy of the probability modeling stage is critical for the efficiency of the arithmetic encoder. Wrong probability estimation can lead to reduced compression ratio and, in extreme cases, can cause data size expansion. Therefore, Section 1.5 contains excessive analysis of the efficiency of probability modeling of syntax elements.

0.5.6 Hybrid video coding challenges

The flexibility of hybrid video coding offers “infinite” possibilities to introduce improvements to the algorithms and methods used in encoding, as well as possibilities of optimizations of the video encoder. Most of the world’s research that is being conducted in these fields concentrates on two goals:

1. Improving compression efficiency by introducing new coding tools or tuning and optimizing existing ones.

2. Reduction of encoder computational complexity (encoder runtime) while minimizing the loss of compression efficiency.

Since the first point is quite obvious, the reduction of encoder complexity requires a short introduction. Most video coding techniques and standards mentioned in Section 0.4 are designed with the assumption of strong asymmetry in the complexity of the video encoder compared to the complexity of the decoder. To be more precise, the complexity of the encoder is many times greater than the complexity of the decoder. The justification for this state of affairs is the fact that video is encoded only once and decoded by users many times. Video encoding can be performed on powerful hardware, but it is important that decoding, which is performed in users' homes, can be performed on much less advanced, and therefore much cheaper, computing devices. In this way, the very high computational complexity of the encoding process is counterbalanced by the relatively low complexity of the decoding process. The very high complexity of encoding means that the practical implementation of contemporary video compression techniques must be preceded by work and research aimed at optimizing the video encoder. To some extent, this problem was the subject of the doctoral student's research.

Moreover, in parallel to the abovementioned areas, one can point to research related to improving the practical usability of the compression technique-based systems or processing and distribution pipelines. The examples of commonly used "additions" to the compression technique are algorithms related to the processing of the compressed bitstream. Notable examples of such algorithms are:

- Transcoding (including transrating) is required to adapt the encoded bitstream to bandwidth constraints while avoiding costly full-reencoding.
- Digital watermarking of a compressed bitstream is used to protect encoded content.

Both mentioned "additions" were also investigated by the author of this thesis and included in the described research results.

This dissertation describes research related to all the mentioned areas. Chapter 1 describes the statistical analysis of video data properties and the usage of gathered knowledge to design a faster encoder. Chapter 2 contains a description of compression efficiency advances achieved by improving Rate Distortion Optimized Quantization (RDOQ) and motion estimation. Chapter 3 concentrates on the adoption of HEVC to encode multiview and Depth Image Based Rendering (DIBR) [Fehn04] video content. The example of practical usability improvement, as a form of low complexity compressed bitstream watermarking, is described in Chapter 4.

1 Analysis of video coding efficiency and statistical properties of data in the encoder

In the period when the described research was conducted, the High Efficiency Video Coding (HEVC) was a very young technology and required detailed analysis. Some of the computational complexity-related statistics [Boss12] were already published. Other important properties, like partitioning video images into blocks, contribution of individual frame types into the bitstream of encoded data, or distribution of the encoder's syntax elements, were not investigated. Only some partial data from the development stage was available [G569], but gathered for the development version of HEVC, thus for the version that was different from the final standard. The experimental research, conducted by the author of this thesis, significantly complemented the research of other researchers, which was reflected in published articles.

In this part of the dissertation, the author's analysis of important HEVC statistical features is presented. At that time, the research towards discovering HEVC statistical features was important to gain knowledge about new, emerging technology. Nevertheless, the gathered knowledge also became a foundation for subsequent works, realized in relation to later compression techniques (like Versatile Video Coding – VVC or enhanced compression model – ECM), like the implementation of optimized video encoders, the reduction of encoding complexity, etc. This last aspect is indeed worth emphasizing.

The goal of the described research was to gather in-depth knowledge about statistical properties of data in a video encoder (partitioning of video images into blocks, coding modes selected by the encoder, bitrate of encoded data, or syntax elements data distribution in encoded data stream, etc.). The next step of the research was to exploit the gathered knowledge and conclusions and to use them to develop an improved video encoder with reduced computational complexity.

1.1 Methodology of research

The analysis of important properties of data in the video encoder was performed by experimental evaluation. As said before, analyses were performed in relation to HEVC technology. The test sequences were encoded to obtain HEVC bitstreams. The content of the bitstream was then analyzed, and statistics were gathered.

In order to obtain a high number of diverse bitstreams, the wide range of FullHD (1920x1080) test sequences was used: bluesky, pedestrianarea, riverbed, rushhour, station2, sunflower, tractor. Moreover, the sequences were encoded with a wide range of bitrates by using an extensive set of QP (quantization parameter) values. The range of QP parameters was from QP=10 (corresponding to bitstream in the range of hundreds Mb/s and very high quality of decoded images) to QP=51 (corresponding to bitstream in the range of dozens kb/s and terrible image quality). The HEVC Reference Software (HM) [HM] was used as an encoder. Even though the HM encoder is very slow, it

performs an exhaustive search and examines almost all possible compression modes, which makes it suitable for such research. In order to obtain useful data, the research was carried out in full compliance with the recommendations of expert groups ISO/IEC MPEG and ITU-T VCEG as described in Common test conditions and reference configuration [L1100].

The encoded bitstream was later processed by special software - HEVC analyzer developed by the author of the thesis. The analyzer software was written from scratch and consists of an HEVC-compliant entropy decoder and a modified entropy encoder. The modified entropy encoder should be called "entropy counter" since its aim was not to produce a bitstream but to gather all statistics.

The software and methodology developed for this research are used during several research projects related to HEVC technology.

1.2 Analysis of frame partitioning into blocks in HEVC

As described in Section 0.5.3, HEVC introduced a new method of image partitioning into smaller parts with an extensive set of units. This new partitioning scheme also found application in later compression techniques, like VVC. When compared to AVC, HEVC has more unit types (Coding Unit – CU, Prediction Unit – PU, and Transform Unit – TU) and a wide range of unit sizes and shapes. The process of dividing CTBs (Coding Tree Blocks) using PU and TU partitioning results in nearly 3000 combinations of the mentioned partitioning for intra-frames and over 4500 combinations for inter-frames that the encoder needs to analyze for each CTB. In simpler terms, the encoder has a large number of options to consider, and it must assess all of them to select the best one (from the point of view of compression efficiency). However, this approach suffers from a significant drawback: the immense computational complexity it entails. Consequently, during the encoding process, not all combinations of image encoding can be evaluated in practical applications. In practice, it is necessary to adopt solutions (which can be developed based on analyses such as those presented in this research) that will achieve a compromise between coding efficiency and computational effort.

This was the aim of the author's research. The research covers the sizes and shapes of Coding Units, Prediction Units, and Transform Units chosen by the HEVC encoder depending on the bitrate of encoded data and frame type (I, B_0, B_1, B_2, B_3, etc). The research includes an evaluation of the frequency of particular CU, PU, and TU size selection made by the encoder and knowledge of image area covered/encoded with particular CU, PU, and TU sizes, with detailed analysis for individual frame types used by the HEVC reference encoder.

The results described in this section were published in [A1]. These results serve as a foundation for further research focused on developing fast mode decisions in the HEVC encoder (described in Section 1.7.1). The knowledge gained from studying the statistics of frame partitioning in the paper provides

direct insights into which combinations of CU, PU, and TU can be omitted to accelerate the encoder's mode selection process.

1.3 Contribution of individual frame types into total bitrate of encoded data

The HEVC introduces significant coding efficiency improvement when compared to AVC and other predecessors. Development of this technique covered improvements in both intra and inter coding. Moreover, HEVC introduced the idea of generalized B-pictures (also used in newer techniques, like VVC) and made a significant step towards efficient usage of hierarchical groups of pictures (GOP). The changes related to GOP construction include several new slice types allowing for efficient construction of open GOPs (as described in Section 0.5.2). Those changes may lead to different contributions of individual frame types into the bitstream of encoded data (when compared to previous techniques [Kama05]).

The objective of the study was to investigate the relation between the sizes of images of different types within the encoded data stream, specifically I-frames and B-frames. The additional goal was to evaluate the association between the sizes of individual images at various encoding levels in hierarchical coding of images, such as I-frames, B0-frames, B1-frames, B2-frames, and B3-frames, and the utilization of B frames.

The detailed description of the Analysis of Frame Partitioning in HEVC mentioned in this section was published in [A2]. The knowledge about the contribution of individual frame types to the total bitrate of encoded data proved to be useful during the development of compressed bitstream watermarking techniques (described in Section 4). As shown in [A2], in a GOP containing 16 frames, intra frames constitute a significant part of a bitstream (~20%). Consequently, the author ot this thesis decided to avoid processing of I frames to reducte computational complexity of watermark embedding.

1.4 Bitrate distribution of syntax elements in the HEVC

Having a comprehensive understanding of the content within the encoded video data stream is crucial for efficiently controlling the bitrate of encoded data in an encoder. This knowledge enables a more efficient allocation of the bit budget among individual frames in a video sequence. While some previous studies have explored this area, they have mostly provided preliminary findings. Specifically, the literature lacks separate and detailed results specifically focusing on I and B frames.

HEVC syntax contains dozens of individual elements. Analyzing each syntax element separately is difficult and may be non-conclusive. Therefore, HEVC syntax elements were analyzed in groups regarding the type of transmitted information, i.e., related to luma or chroma residual data.

The development of HEVC prediction tools (especially a severe improvement in intra [Sze14] and inter [Rich24] prediction) together with effective partitioning schemes leads to a significant reduction of bits spent on encoding prediction modes. Notable examples of such improvements are intra-predictive

coding of intra prediction modes or merge mode for iter coded units. This leads to an obvious conclusion that this should result in a change in the overall syntax element distribution by making the residual the most significant contributor to the entire bitstream. However, these superficial findings proved to be not true since HEVC significantly improved the efficiency of residual signal coding.

The most important conclusion from that research is the conclusion that the introduction of new, more efficient prediction tools was offset by more efficient residual coding and did not change the overall syntax element distribution when compared to MPEG-4 AVC.

The results described in this section were published in [A3].

1.5 Efficiency of estimating the statistics of syntax elements in CABAC

Since CABAC is a binary arithmetic coder, it can only encode two binary symbols: 0 and 1. Consequently, syntax elements that are not inherently binary, such as motion vectors (to be precise, motion vector differences), must be converted into strings of binary symbols. Each individual binary symbol in this string is referred to as a bin.

A crucial step in CABAC encoding is estimating the probabilities of the encoded data symbols (bins). Therefore, this aspect of encoding was the subject of analysis and research in the doctoral thesis. In CABAC, the estimation of the probability of each bin is accomplished using an advanced data statistics estimation mechanism. The probability of encountering 0 or 1 varies and generally depends on factors such as the specific bin number being encoded, the type of syntax element being encoded, and the values of previously encoded bins within the same syntax element.

To estimate the probability of a bin in a given image block, the context information is also taken into account. This context information contains details about the values of the bins in neighboring blocks of the image (relative to the block currently being encoded). By considering the potential values of the context information for specific bin types and the various syntax elements defined in HEVC, the stream of bins is divided into 154 sub-streams. This division allows for independent tracking of data statistics within each sub-stream. Hence, it can be said that CABAC employs 154 distinct probability estimators (the so-called context modelers) to handle these sub-streams of bins.

The objective of this study was to conduct a thorough analysis of each individual bin in sub-streams of bins encoded by CABAC and determine the proportion of bits produced by CABAC in response to each bin symbol. This fraction of bits produced in response to a bin can be considered as the compression performance of encoding that specific bin. Such studies are of great importance from the perspective of assessing the coding efficiency of individual data sets in CABAC, as well as from the perspective of work on further improvements of the entropy compression technique.

The research aims to answer several questions: How many bins are effectively compressed? How many bins result in data expansion due to ineffective compression? What percentage of bins are compressed with a specific efficiency level, such as generating 0.2 or 0.3 bits at the output of the arithmetic encoder core in response to the appearance of a bin at the input?

Since each encoded bin was analyzed separately, the experiments generated a substantial amount of experimental data. Consequently, the partial results were organized into 154 CABAC sub-streams and further categorized based on the type of syntax element that produced each bin. Three groups of syntax elements were utilized: symbols related to the prediction error of image samples, symbols related to motion data, and symbols related to control data.

The findings indicate that the CABAC entropy encoder achieves efficient compression for the majority of bins. The specific results depend on the type of encoded data and the value of QP, but on average, approximately half of the bins achieve a compression ratio ranging from 0 to 0.8 bits/bin, while most of the other bins achieve a compression ratio between 0.8 and 1 bits/bin. Interestingly, some bins result in data expansion. In extreme cases, these expanding bins accounted for around 5% of all the bins, suggesting that certain CABAC sub-streams could be used more efficiently. In the course of the research, it was indicated which sub-streams are affected by the inefficiency of probability modelling.

The results described above suggest that probability modeling in CABAC should be improved to avoid incorrect probability estimates, which leads to significant data expansion in the entropy encoder stage. The results obtained from this research serve as a foundation for future work on enhancing CABAC and developing fast mode decision algorithms in video encoders. An example of such improvements was presented in Section 1.7.3. The detailed results described in this section were published in [A4].

1.6 Homogeneous transcoding and multi-generation encoding

In some cases, it's necessary to re-encode the video (e.g., because different encoding parameters are required) or to change the format of the encoded data stream. In both cases, we are talking about transcoding the encoded video data stream [Vetr03].

It is commonly known that transcoding introduces inevitable quality loss. Obviously, there are exceptions, but those are special cases with very narrow application. One of those exceptions is described in Section 4.1 – as our watermark embedding software, which incidentally allowed us to convert a CABAC encoded bitstream into a CAVLC encoded one without changing its content.

Unfortunately, in real-world applications, video transcoding is practically unavoidable. Most cameras (including professional devices) compress recorded data. The size of high-resolution content also forces users to compress the immediate content during video processing.

In this section, the study related to quality losses in homogenous HEVC transcoding is presented. The homogenous HEVC transcoding is a transcoding case where both the source and the destination video use the same compression technique, i.e., HEVC.

The first part of the research contains an analysis of multiple encoding-decoding cycles with the same encoder settings and with the same QP value. In this case, 80 cycles were performed, and after each cycle, the bitrate and quality (PSNR) were gathered. Results were compared against the originally encoded sequence. The total measured quality loss is a range of 0.68 – 2.04 dB (1.4 dB on average) and does not depend on the QP value tested. Interestingly, about 41% of the total quality loss is introduced by the first transcoding operation. One can also notice that quality loss after all (80) cycles can be compensated by increasing the initial quality. The research showed that, in general, the decrease of the initial QP value by 2 is sufficient to compensate the total quality loss. The example of a Rate-Distortion plot for a single case of 80 encoding-decoding cycles is presented in Figure 15. It clearly shows the diminishing quality degradation caused by each transcoding cycle.

The publication containing results mentioned in this section [A5] was rewarded with Kalman Fazekas special award during the 20th International Conference on Systems, Signals and Image Processing (IWSSIP 2013).

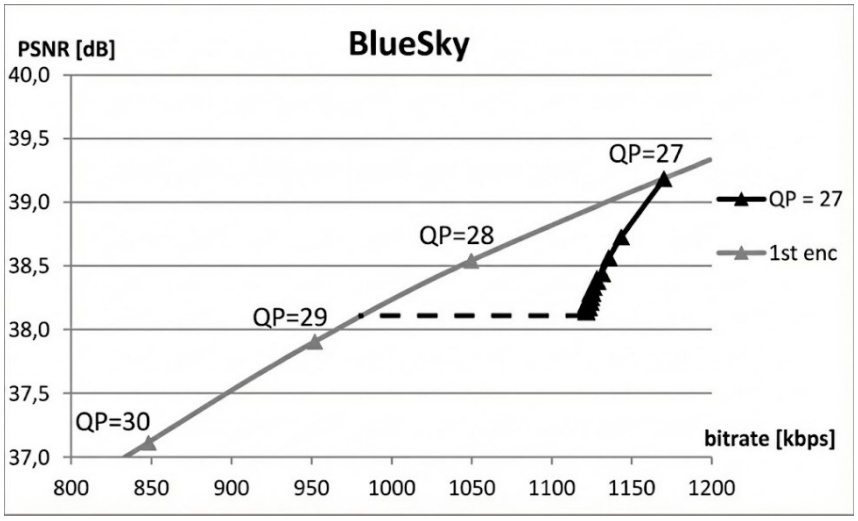


Figure 15. Comparison of the video quality after 80 encoding-decoding cycles with QP=27 and video quality after the first encoding with various QPs for the BlueSky sequence. The grey line represents quality after the first encoding. The black line (and triangles) shows the results of consecutive reencoding. Figure from [A5].

Further studies concentrated on homogeneous HEVC transcoding, with different encoder settings. In this case, the quantization parameter (QP) was altered to simulate transcoding performed in order to reduce the bitrate. In this case, all test sequences were encoded with QP from the set {22, 28, 32, 38} and then transcoded with QP from the entire possible range (1 – 51).

This research and the following analysis clearly show that transcoding-related quality loss depends on the ratio between target bitrate and source bitrate. There is a range of this ratio (between 60% and 90%)

where the quality loss is disproportionate to bitrate reduction (see Figure 16). This leads to an interesting conclusion related to the design of systems that deliver multiple bitstreams with different target bitrates. Investigation suggests that medium quality bitstreams (between 60% and 90% bitrate of base quality bitstream) should be encoded independently, while low quality bitstreams (below 60% of top quality stream) could be transcoded from top quality bitstream. Similarly, near-base bitrate bitstreams (in the range 90-100% of base bitrate) can be effectively produced by a transcoding process.

The results described in this section and the detailed analysis of homogenous transcoding-related problems were published in [A6].

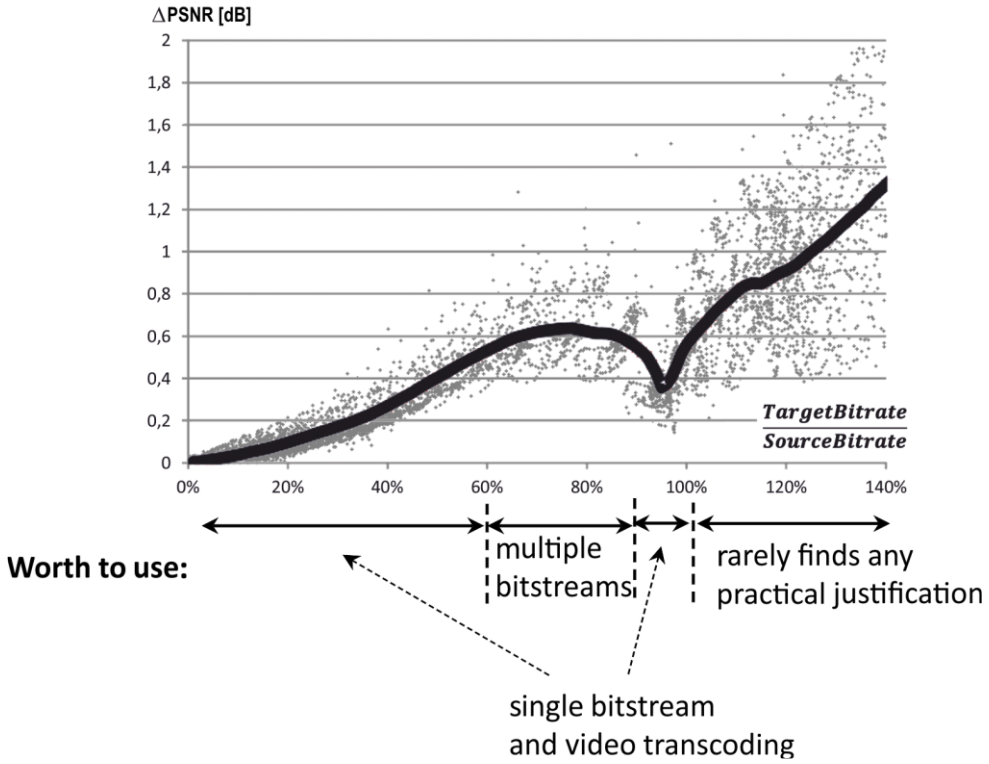


Figure 16. Ranges of bitrate reduction ratios when it is worth using video transcoding instead of the solution with multiple bitstreams. Figure from [A6].

1.7 Practical use of gathered statistical properties

Optimization of the encoding process is a critical area of research in video compression, particularly with the advent of sophisticated but computationally very complex compression technologies such as High-Efficiency Video Coding (HEVC) and its successors. Research described in previous sections of this paragraph allowed for gathering a wide knowledge about the behavior and statistical properties of syntax elements data that are produced by new-generation video encoders. These studies were conducted with reference to a well-known representative of this class of video encoders – the HEVC encoder. The conclusions from that research were later used to develop optimized versions of the video encoder that implement an improved video encoder control mechanism. These new video encoder implementations and the related research are presented later in this thesis.

1.7.1 Video encoder with mode decision assisted by pre-computed statistics of data

The first approach introduces a new method to significantly speed up the encoding process in the HEVC encoder that uses intra compression modes. The core idea is to cut off testing encoding modes (at the stage of selecting compression modes) that are rarely chosen by the HEVC encoder, based on very detailed results of statistical analyses of encoder data. The mentioned cut-off of modes is regulated by a defined value of the percentage threshold th - encoding modes whose usage statistics by the encoder are lower than the th threshold are completely omitted in the encoding mode selection procedure. By using conditional statistics of encoding modes, the proposed method achieves a balance between the scale of reduction of computational complexity of the encoder and efficiency of video compression. By using the th threshold, this balance can be properly adjusted in the encoder. It operates at the Rate-Distortion optimization stage by quickly determining the optimal size for coding tree units, prediction units, and transform units. Additionally, it speeds up the selection of the most appropriate intra-predictor for each PU block.

The proposed method drastically reduces the computational complexity of the encoder, according to research conducted in accordance with the recommendations of the ISO/IEC MPEG and ITU-T VCEG expert groups [L1100]. Simulations show noticeable results: for 1920x1080 video sequences (class B), encoding time can be reduced by three times or two times, with only a slight increase in BD-rate (4% and 1.5% respectively) compared to the HEVC test model version 16.9. This approach is also designed to work seamlessly with other fast mode selection techniques that can be used in HEVC video encoders, which constitutes the additional value of the developed method. Detailed experimental results obtained are presented in Table 1.

Table 1. Summary of performance results for HEVC encoder with mode decision assisted by precomputed statistics.

Sequence class	Operation point 1			Operation point 2		
	Threshold [%]	BD-Rate [%]	Encoding time reduction [%]	Threshold [%]	BD-Rate [%]	Encoding time reduction [%]
A	th = 1	1.09	30.57	th = 1.5	2.28	50.10
B	th = 1.5	1.83	33.32	th = 2.5	4.03	49.82
C	th = 2.5	1.14	34.16	th = 4	8.41	59.41
D	th = 2.5	0.29	18.63	th = 4	5.89	47.07
E	th = 1.5	4.04	30.98	th = 2	8.92	54.67
F	th = 1.5	9.40	23.46	th = 2	25.19	39.87

The detailed description of a fast HEVC encoder with mode decision assisted by precomputed statistics of encoding modes is described in [A7].

1.7.2 Encoder with mode decision assisted by noise reduction

An alternative approach to encoder control is related to observations about the influence of noise in the image on the Rate-Distortion optimization (RDopt) algorithm decisions. A careful analysis of encoder decisions related to unit size and prediction mode is strongly influenced by noise presence. Even a single pixel value change by +/-1 can lead to a completely different chain of decisions that are undertaken in subsequent blocks of the image.

The reason behind this phenomenon is the way almost all video encoders work. Since encoding (and especially the mode decision stage) is a very computationally complex process, encoders use a greedy approach. They try to make optimal decisions for the currently encoded block, without considering further consequences of the current decision. Therefore, the encoder can select a mode that is optimal at the block level but suboptimal on the picture level and never reconsider this decision.

The observations made by the author of this thesis served as an inspiration for a new research idea of improved RDopt. This leads to the original proposal of a mode selection mechanism operating on a noise-reduced version of the input sequence, while the core video information, such as transform coefficients representing residual data, is encoded from the unaltered input. This approach necessitates encoding two versions of the input sequence – a noise-reduced version and the original version (containing information related to noise). The full mode search is performed on noise-reduced video. Further, decisions made by the first encoder are applied by the second encoding pass operating on the original sequence.

The proposed approach leads to a mode decision closer to optimal, since the presence of noise does not disturb the Rate-Distortion optimization stage. The resulting computational complexity is negligibly higher than that of a single encoding since the second pass does not require expensive encoding mode search.

The proposed method has been implemented in the HEVC reference software [HM] and was experimentally evaluated according to CTC [L1100]. Results indicate that this proposal allows for a 1.5% reduction of bitrate for similar compressed video quality, which is a significant achievement in the case of new-generation video compression techniques. The detailed description of HEVC encoding assisted with noise reduction can be found in [Stan18a].

1.7.3 Research for VVC video compression technology and its improvements

The knowledge gathered by analyzing the HEVC properties has also proved to be useful for the newer generations of video encoders. It refers to the Versatile Video Coding (VVC) technique and its improvements. The author of this thesis was involved in work (contracted by a certain company) towards improvements in video compression related to post-VVC development. The development of

the VVC successor is performed by the Joint Video Experts Team (JVET) and uses the ECM software model [ECM].

The author of this thesis (as a member of a team from the Institute of Multimedia Telecommunications) authored three contributions related to entropy encoder improvements and general coding scheme improvements.

The entropy coding related proposal describes optimization of probability estimation in CABAC [AL0198]. The contribution proposes new adaptation strategies for the parameters of the data statistics model, using the gradient-descent-based optimization algorithm (Cauchy optimization) for more accurate probability tracking. This allows for improving compression efficiency up to 0.05% for selected scenarios.

The second topic is related to general coding scheme improvements. As mentioned in Section 1.7.2, encoders use a greedy algorithm and perform optimization on the block level. This can lead to choosing a suboptimal path and, as a result, reducing compression efficiency for the entire picture.

Experiments described in [AK0171] show that changing the starting point (picture corner) and encoding direction can push the encoder into a better path, resulting in improvements in coding efficiency. The solution proposed in [AK0173] and evaluated in [AL0199] uses a reduced set of starting points and coding orders:

1. Starting point – top-left corner, coding order – left-to-right, top-to-bottom (default).
2. Starting point – bottom-left corner, coding order – left-to-right, bottom-to-top ("VERMIR").

In addition, a very simple decision algorithm based on the Sobel filter was developed. All proposed algorithms were implemented in ECM software [ECM] and evaluated according to ECM Common Test Conditions [AI2017]. The proposal resulted in a 0.23% improvement in compression efficiency.

1.8 Chapter summary and original achievements of the research

The original research results described in this chapter refer directly to the analysis of several critical properties of data in contemporary hybrid video encoders. The obtained achievements include: analysis of video frame partitioning into blocks [A1], analysis of individual frame types contribution into total bitrate of encoded data [A2], data on distribution of the encoded bitstream among the individual syntax elements [A3], detailed results on efficiency of probability modeling in CABAC obtained for individual substreams of binary symbols [A4] and results for video transcoding and multi-generation video coding [A5][A6]. The reference point for the mentioned studies was the HEVC video compression technique - the most advanced hybrid video compression technique available at the time of the author's experiments. However, as it has been stated, a large part of the conclusions obtained from the research can also be the basis for improvements to the next generations of video encoders after HEVC.

Knowledge acquired as a result of the above-mentioned analysis was used to develop two optimized HEVC encoder implementations. The first, an encoder with mode decision assisted by precomputed statistics of encoding mode usage [A7], allowed for a significant reduction of computational complexity without significant degradation of compression efficiency. The second, an encoder with mode decision assisted by noise reduction in video frames [Stan18a], allowed for improving compression efficiency by reducing the negative consequences of the greedy approach in encoder mode search.

Furthermore, the knowledge gathered by the author of this thesis proved to be also useful in research aimed at developing improvements to the new technique developed as a successor to VVC. Subsequent research carried out by the author resulted in two original proposals of improvements: 1) the CABAC entropy codec with improved symbols' probability tracking [AL0198], and 2) adaptive changing of the starting point and direction of the encoding process (VERMIR) [AK0173]. Both proposals resulted in noticeable compression efficiency improvements, which is a significant achievement at the current state of compression technology development.

2 Video encoder optimizations

As mentioned in 0.5.1, the video compression standard defines a set of coding tools that can be used to efficiently encode a video sequence. Each new generation of hybrid video encoding techniques allows for achieving better compression ratios and higher efficiency, but with inevitable cost – higher computational complexity.

However, the final compression efficiency depends not only on the selected technique (e.g., HEVC or VVC technique) but mostly on decisions made by the video data encoding mode selection mechanism, which is embedded in the video encoder. The encoder control mechanism is a non-normative part, and its operation is determined by the way it is implemented by the programmer (designer) developing the video encoder software/hardware. If computational power (or energy) is a critical resource, one can create a simpler encoder (not exploiting all decision modes). On the other hand, there is always room for further improvements. Obviously, an encoder designer cannot change any coding tools to preserve conformance with the standard. On the other hand, encoder can use a wide range of techniques, mostly related to mode decision, which can improve the compression ratio while spending more power on computations

Examples of this so-called non-normative techniques are Rate Distortion Optimized Quantization (RDOQ) and Motion Estimation (ME). Both of them are encoder-only algorithms; therefore, they are not covered by a standard, and both of them can have a significant influence on compression efficiency (and encoder computational effort).

2.1 Improved Rate-Distortion Optimized Quantization

The Rate Distortion Optimized Quantization (RDOQ) can be considered as a result of years of systematic improvements in the calculation of quantized transform coefficients. RDOQ estimates the Rate-Distortion cost of encoding a transform block and optimizes quantized coefficient values in order to minimize that cost [AH21]. The RDOQ shares a significant number of components with other parts of Rate-Distortion Optimization (RDopt), being a necessary part of any modern encoder.

The fundamental objective of the presented method is to determine an optimal set of quantized transform coefficients that minimizes the Rate-Distortion (RD) cost. Theoretically, achieving this optimization necessitates an exhaustive search across all possible combinations of quantized coefficients. However, such an exhaustive approach is computationally infeasible due to the extreme complexity it imposes on the encoder.

In order to estimate RD cost, RDOQ calculates the image distortion introduced by the quantization of transform coefficients within an encoded block, alongside the number of bits required to encode the corresponding quantized transform coefficient (or block of coefficients). By evaluating these two metrics, the encoder can determine the optimal coefficient value based on its calculated RD cost.

The RDOQ is not included in any compression standard, since calculation of quantized transform coefficients is non-normative and can be used in any encoder. The RDOQ can be used with any transform-based compression technique. It is even possible to successfully retrofit RDOQ into such an old technique as JPEG [Graj22], which is confirmed by the research of the author of the dissertation.

Unfortunately, the RDOQ introduces significant computational complexity. The encoder performs analysis for each transform coefficient, and for each modification, it has to calculate both distortion and the bit cost. This complexity significantly increases encoder runtime; therefore, most encoders use simplified estimation of distortion and cost. Unfortunately, this simplified approach leads to inaccuracies (in distortion and the bit cost) and results in suboptimal decisions.

Table 2. Exact RDOQ experimental results for "All-Intra" scenario.

Sequence class	HM RDOQ performance (BD-rate) [%]			Exact cost RDOQ performance (BD-rate) [%]			HM RDOQ vs Exact cost RDOQ (BD-rate) [%]		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
A	-4.1	1.0	-0.7	-5.1	0.7	-1.2	-1.0	0.0	-0.4
B	-5.4	0.1	-1.4	-6.5	-0.9	-1.1	-1.2	0.0	-0.2
C	-3.5	1.0	-0.3	-4.4	1.1	-0.2	-1.0	0.4	0.3
D	-3.3	1.2	0.0	-4.2	1.3	0.2	-1.0	0.3	0.4
E	-3.7	-0.9	-2.9	-4.5	-1.0	-3.2	-0.9	0.1	-0.2
F	-2.6	0.0	-0.9	-3.8	-0.3	-1.3	-1.2	-0.1	-0.2
All	**-3.8**	**0.4**	**-1.0**	**-4.8**	**0.1**	**-1.0**	**-1.1**	**0.1**	**0.0**

Table 3. Exact RDOQ experimental results for "Random Access" scenario.

Sequence class	HM RDOQ performance (BD-rate) [%]			Exact cost RDOQ performance (BD-rate) [%]			HM RDOQ vs Exact cost RDOQ (BD-rate) [%]		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
A	-3.3	-4.9	-6.4	-4.0	-3.0	-4.7	-0.9	-0.2	-0.3
B	-5.1	-5.9	-6.3	-5.6	-6.7	-6.4	-0.9	0.0	-0.4
C	-4.0	-5.9	-6.4	-4.8	-5.2	-5.6	-0.9	0.3	0.0
D	-3.9	-5.0	-6.1	-4.8	-4.9	-5.6	-0.9	0.1	0.4
E	-3.2	-2.6	-4.3	-4.0	-2.8	-4.9	-0.9	-0.3	-0.7
F	-3.7	-4.6	-5.1	-4.9	-4.4	-5.0	-1.3	-0.2	-0.2
All	**-4.0**	**-5.0**	**-5.8**	**-4.7**	**-4.5**	**-5.4**	**-1.0**	**-0.1**	**-0.1**

Obviously, the usage of an optimization algorithm that takes suboptimal decisions left some coding performance on the table. It is possible to design an RDOQ algorithm that uses exact calculation of

distortion and bit cost and improves optimized decisions. In order to evaluate this concept, the author of this thesis conceptualized an idea of research and supervised the development of software. As a result, the approach with “exact” calculations has been implemented and tested. The obtained results are summarized in Table 2 and Table 3. It clearly shows that improvements to the RDOQ stage can result in ~1% improvement in coding efficiency of the HEVC encoder. Since RDOQ is an encoder-only tool, the described improvements do not break the conformance with the HEVC standard.

Detailed results for an improved Rate-Distortion Optimized Quantization algorithm for the HEVC encoder are presented in [A8].

2.2 Optimized motion estimation in contemporary hybrid video encoding

Motion-compensated inter-frame prediction is a necessary part of every modern video codec. It is the most efficient tool, exploiting inter-frame similarity and extremely increasing compression efficiency (when compared to intra-only codecs)[Shi08]. However, motion compensated inter-frame prediction requires accurate knowledge about the object's motion in the encoded sequence (assuming here a specific motion model for the objects present in the scene registered by the camera); thus, this information has to be provided by the motion estimation module of the video encoder.

The motion estimation process involves high computational complexity, constituting a significant part of the entire encoder computing burden [Boss12]. Modern video codecs allow for lists containing several reference pictures for inter-frame prediction and the usage of bi-directional prediction (bi-prediction). Those factors make the encoder complexity significantly higher than, for example, a fixed scheme, which assumes a frame that precedes in time the currently encoded frame as a reference frame for inter prediction.

On the other hand, the accuracy of the frame content prediction algorithm is very important, since better prediction results in less residual data to be transmitted in the encoded datastream.

The essential and difficult scientific problem is to find a good tradeoff between the computational complexity of the motion estimation procedure and the accuracy of the prediction signal. This problem has multiple aspects. There are several motion estimation algorithms with TZ-search [Xiu10] (and its derivatives) being recently the most popular [HM][VTM][ECM]. Other aspects include the selection of the starting point and search range used by the motion estimation algorithm, or the adopted metric of similarity of the two image blocks (currently encoded block and the predicted version of the block). These aspects were the subject of the author's research.

The motion estimation needs to select a proper starting point to avoid searching useless areas of the reference picture. In modern video compression techniques, like HEVC, most PUs (for which inter-frame prediction is realized) are surrounded by other, already encoded, units. This allows us to take the neighboring unit motion field as a starting point candidate. Moreover, in most cases, the encoder tries

to use different block (CU) sizes, often beginning from bigger units and investigating smaller splits in later steps. This allows the reuse of motion vectors obtained for a bigger unit within the currently encoded sub-unit. Starting point selection also consumes computational resources; however, as shown in Figure 17, selection of an appropriate starting point has a significant influence on prediction signal quality (especially for smaller blocks).

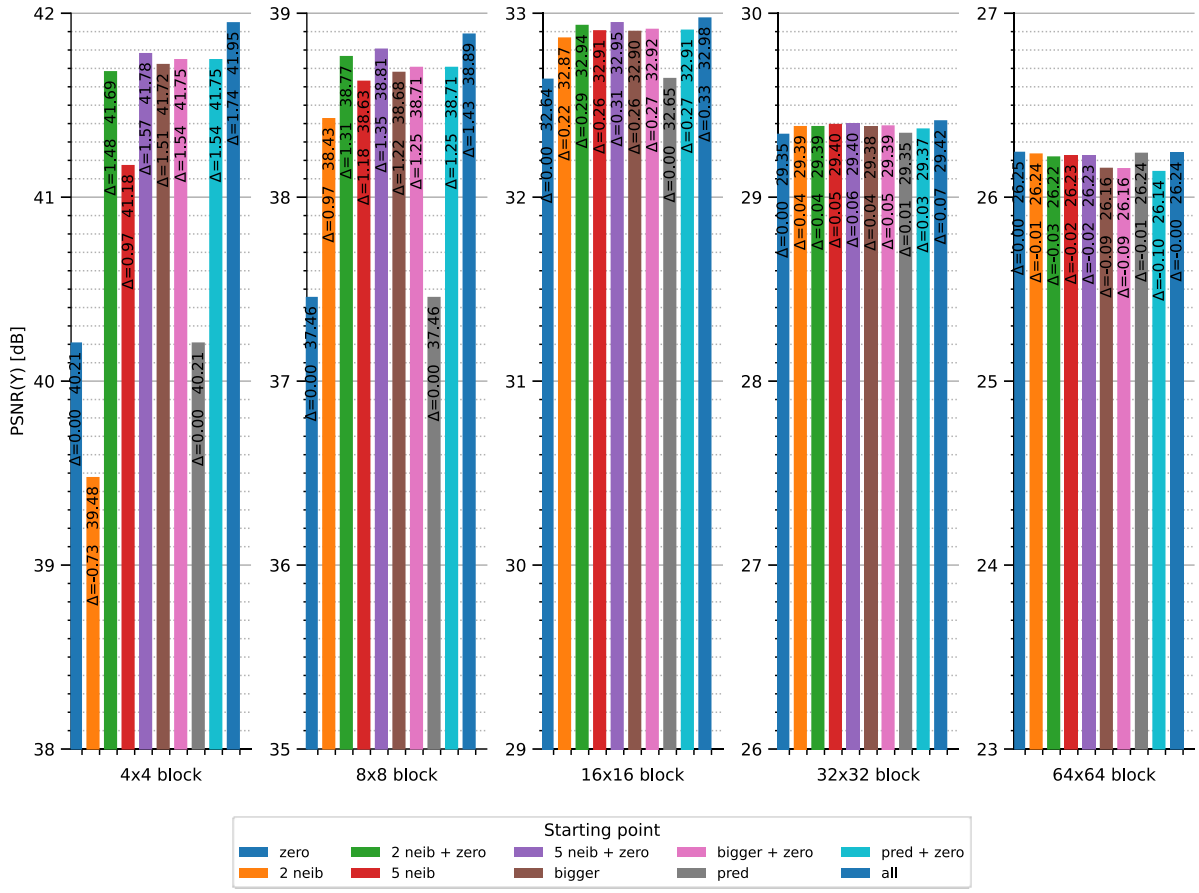


Figure 17. Dependency of the prediction signal accuracy on the starting point selection. The accuracy is represented as PSNR calculated between the original picture and the picture assembled from individually predicted blocks. The predicted blocks were calculated by using the motion vector found by the motion estimation algorithm, with the use of different starting points for estimation. Figure from [A10].

There are also several parameters of the motion estimation procedure that can be altered based on local conditions. One of the most important is the search range, which determines the area being searched around the starting point. A wider area increases the probability of finding a good prediction source, but on the other hand, it increases computational and memory bandwidth requirements.

One of the possibilities to overcome this problem is to adaptively choose the search range. Assuming that the optimal starting point was selected, one can try to reduce the search range for smaller prediction units. Figure 18 shows the author's experimental results obtained by reducing the search based on the size of the prediction unit (PU). Since PUs can be rectangular, the A1 sets the search range to equal the larger dimension of the PU (width or height). Similarly, A2 uses a smaller dimension of PU as the

search range. As a result, the encoding time can be significantly shorter while preserving moderate compression efficiency loss.

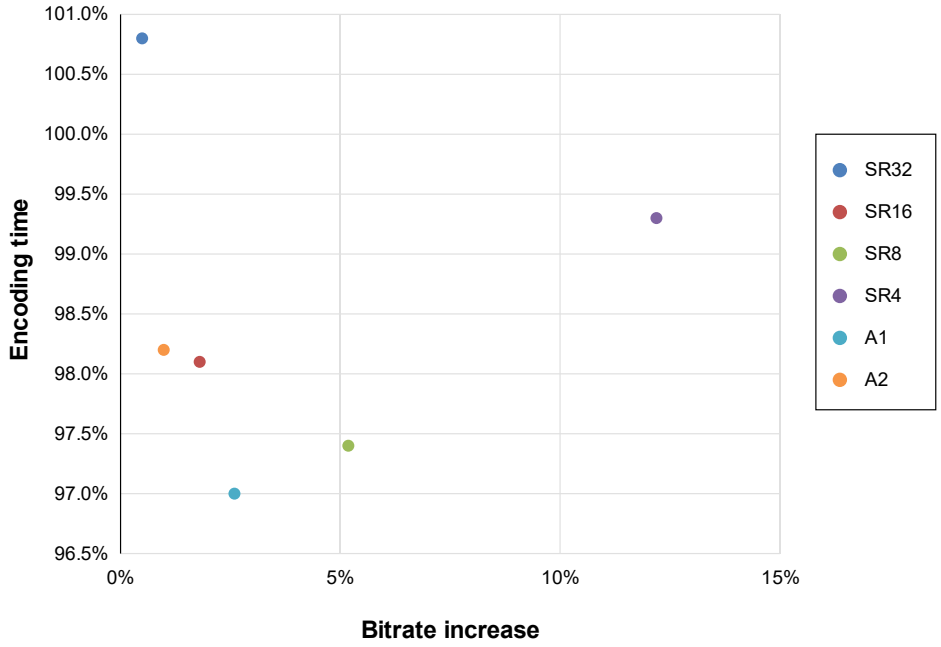


Figure 18. Encoding time (relative to the default HEVC value of ±64 x ±64) for smaller search ranges (SR32, SR16, SR8, SR4) and proposed by the author adaptive approaches (A1, A2). Figure from [A10].

One can also consider a similarity metric used during estimation. The common approach is to use Sum of Absolute Differences (SAD) and Sum of Absolute Transformed Differences (SATD). The common approach is to use SAD for general motion estimation, with per-picture element (per-pel) precision. Then, SATD is used for sub-pel (i.e., half-pel, quarter-pel) refining of previously found full-pel motion vector [HM][VTM].

One can also consider using the Sum of Squared Differences (SSD), and by similarity, the Sum of Squared Transformed Differences (SSTD) can be defined. It should be noted that "Transformed" variants use the Hadamard transform to retain reasonable computational complexity [Oste04]. The influence of the choice of similarity metric into complexity of motion estimation was evaluated by the author of this thesis.

In most publications [Vann12][Setk11], selection of SAD (SATD) is advocated by its lower complexity and avoidance of costly multiplication (required to calculate square difference). However, those naive calculations do not take into account real-world implementations and instructions that are available in modern CPUs. For example, the SIMD extension x86 instruction set has a family of instructions (pmaddwd / vpmaddwd) which are defined as follows: "Multiply packed signed 16-bit integers in SRC_A and SRC_B, producing intermediate signed 32-bit integers. Horizontally add adjacent pairs of intermediate 32-bit integers, and pack the results in DST". This allows for very fast and efficient

calculation of SSD (and SSTD), which completely changes the tradeoff between metric quality and complexity.

Results obtained for plain C++ implementation and vectorized (using SIMD instructions from the SSE instruction set) are presented in Figure 19. It is clearly noticeable that the computational complexity of the SSD metric can be even slightly lower than SAD.

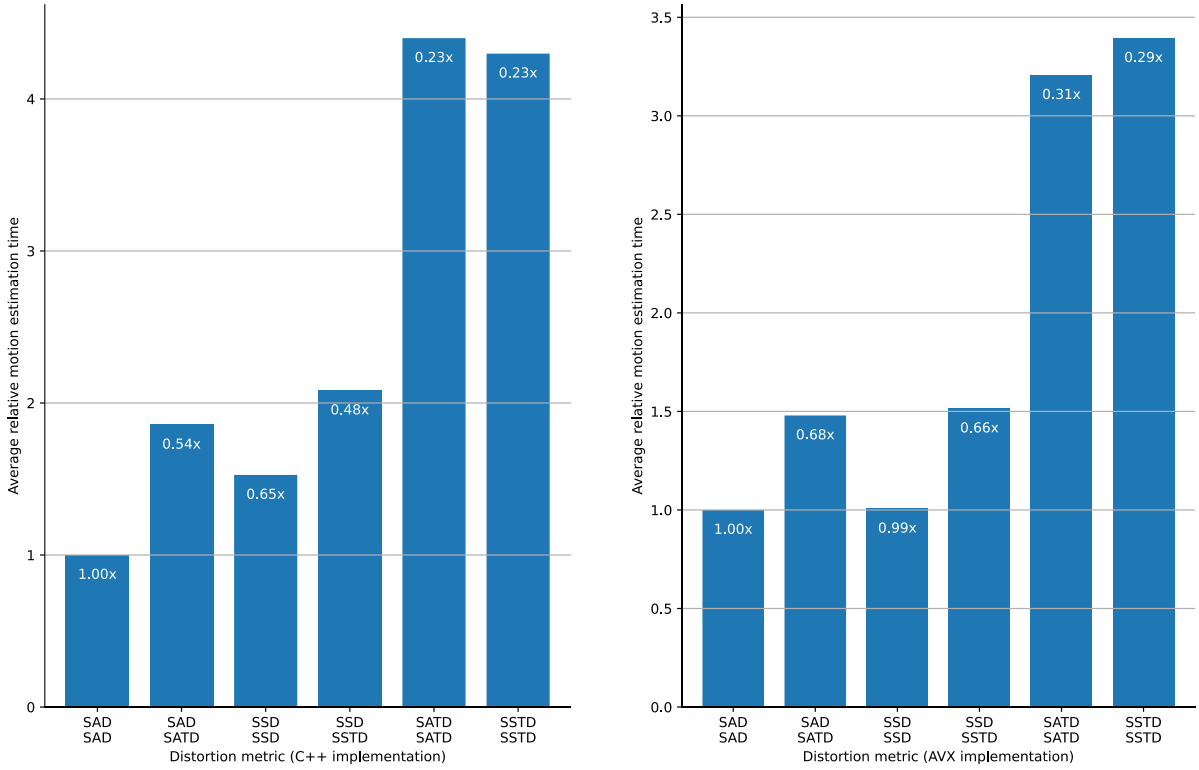


Figure 19. Comparison of different distortion metrics for full-pel / half- and quarter-pel block position. Speed up calculated with SAD/SAD as reference. All results gathered for the TZ search algorithm and 6 working threads. Figure from [A9].

Detailed analysis of the complexity of the HEVC motion estimation and study of Computational complexity tradeoffs in HEVC motion estimation were described in [A9] and [A10], respectively.

2.3 Chapter summary and original achievements of the research

The original research results described in this chapter consist of two non-normative encoder-side improvements. The first is improved rate-distortion optimized quantization (RDOQ) with the author's method of exact cost calculation [A8] that allows for noticeable compression efficiency improvement when compared to the state-of-the-art reference implementation. The second is the optimized motion estimation stage tested on the HEVC encoder [A9][A10]. The author's results show that the accuracy of the prediction signal derived by the motion estimation stage is strongly dependent not only on the search range but also on the selection of the starting point for motion estimation. Moreover, the algorithm for search range adaptation was also proposed, allowing for a noticeable reduction of

encoding time while still maintaining high compression efficiency of video encoding. Lastly, the computational complexity of different distortion metrics used for the motion estimation procedure was evaluated - the obtained research results are of great practical importance because they can be used in the implementation of high-performance video compression techniques of new generations.

3 Extensions for encoding of multiview and FTV content

The most widespread use case of video compression techniques is related to encoding plain 2D video content. There are, however, use cases that require compression of more complex video data. The most prominent examples of such use cases are scalable video encoding and multiview video coding. The scalable codecs (i.e., SVC [SVC] are designed to effectively represent multi-bitrate content. The multiview codecs (i.e., [MVC]) aim to effectively represent a 3D video containing several views, i.e., captured by different cameras in a multicamera system [Stank18]. The encoding of more complex video data requires additional, dedicated coding tools that can exploit data similarities in those complex representations.

The motivation for the author's research, as described in the following paragraphs, was participation in development conducted within an international group of experts.

In March 2011, MPEG issued a "Call for Proposals on 3D Video Coding Technology" [N12036] with the intention to create a codec for 3D video. When compared to MVC [MVC], the 3D video was defined as a set of views (sometimes called "texture") with corresponding depth maps. Such a format, related to Depth Image Based Rendering (DIBR) [Fehn04], allows for synthesizing virtual views and allows the user to smoothly change the viewpoint. The new codec was supposed to be based on HEVC and standardized as an HEVC extension.

The team from the Chair of Multimedia Communication and Microelectronics (KTMiM) entered the contest and developed their own proposal. The author of this thesis was a member of this team and was involved in research, software development, and experimental evaluation of a new proposal. The main achievements of the thesis author in this area are:

- Development of a multiview video codec based on HEVC technology and HEVC reference software [HM] (Section 3.1).
- Research towards improved prediction and predictive coding of motion and disparity vectors (Section 3.2).
- Contribution to teamwork towards creation of KTMiM response to MPEG`s "Call for Proposals on 3D Video Coding Technology", in addition to the points mentioned above:
    - creation of an experimental evaluation framework and software,
    - development of a distortion estimation algorithm for depth map coding.

The research was conducted in response to "Call for Proposals on 3D Video Coding Technology"; therefore, all proposed solutions were developed as extensions of the HEVC codec. However, most of

the techniques and algorithms presented in the following sections can be adapted to any “base” 2D video codec.

3.1 Development of a multiview video codec based on HEVC technology

The first step in the development of a 3D video codec was the creation of a multiview codec based on HEVC technology. A multiview codec should allow for the combined and efficient encoding of multiple views when exploiting the strong similarity of neighboring views. The exemplary coding scheme for the multiview codec was presented in Figure 20.

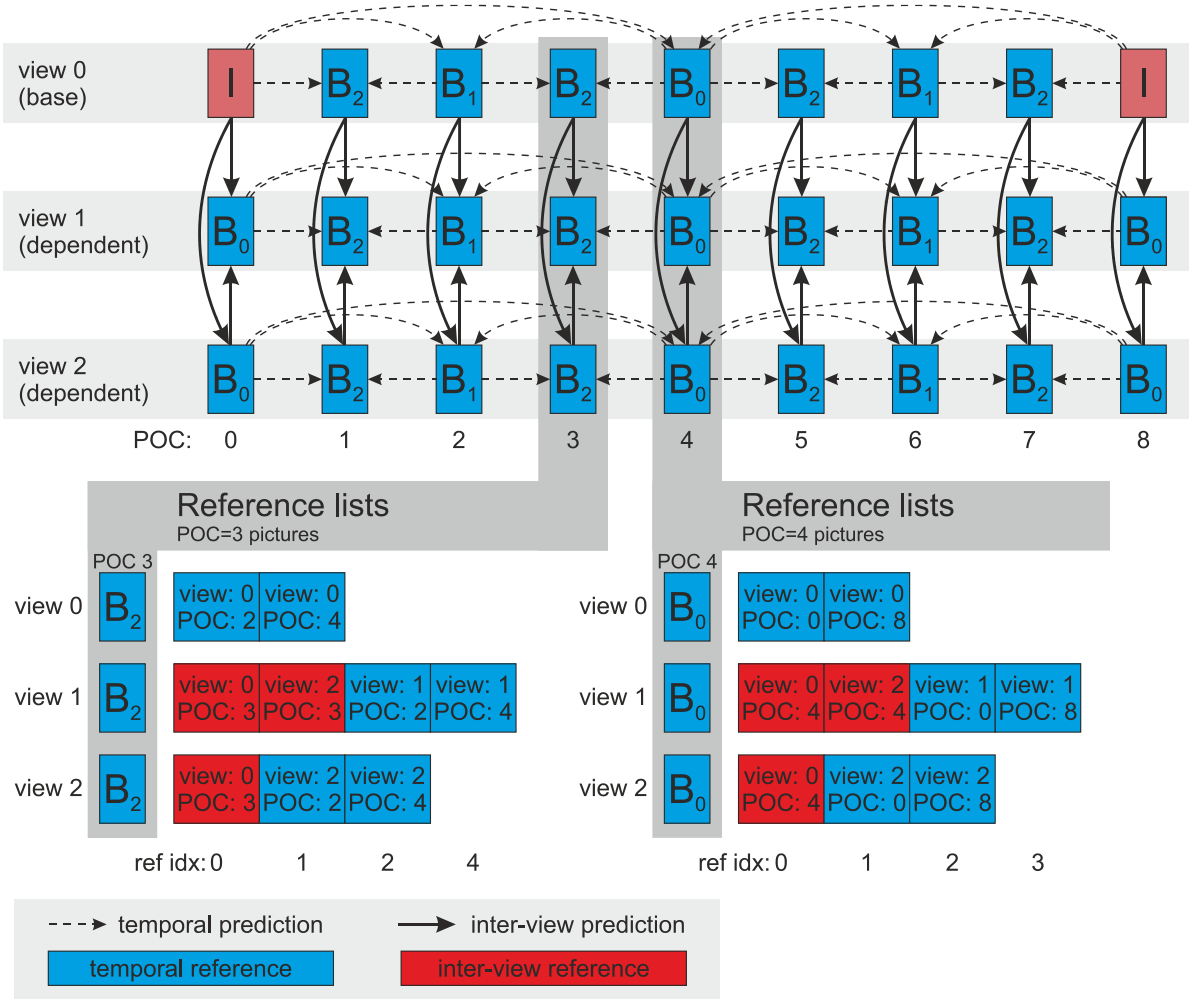


Figure 20. Prediction scheme in exemplary multiview codec and reference lists construction (example for the case of one base view and two dependent views). Figure from [G582].

In a regular video codec, the inter-frame prediction uses reference frames originating from different time points. In the case of a multiview codec, there are two distinct sources of inter-frame prediction – temporal (due to the presence of frames of a given view at different moments in time) and spatial (due to the presence of separate, adjacent views). Both of them share some features, like the general idea of using similar blocks from a reference picture as a prediction source. There are also some differences; motion vectors (temporal) and disparity vectors (spatial) cannot be mixed together since both of them

have different physical meanings. The example of the construction of a reference picture list containing both temporal and spatial sources is presented in Figure 20.

The work towards the creation of a multiview codec was mostly related to software design and development. It was very complex since the HM 3.0 [HM] software (which was the basis for modifications) consisted of about 100000 lines of code.

The HEVC-based multiview codec achieved very promising compression efficiency improvements. When compared to simulcast HEVC, it allowed for a reduction of total bitrate by ~22% and ~30% for the 2-view and 3-view cases, respectively. Detailed description of this codec along with an exhaustive set of performance numbers can be found in [G582].

3.2 Motion and disparity vectors prediction algorithm for multiview codecs

Motion information constitutes a significant part of all data encoded in a classic video encoder. Similarly, in a multiview video encoder, the information describing the disparity between two specified views is also relevant. Consequently, the method of encoding this data has a significant impact on the efficiency of video data compression. Therefore, encoding methods for this data have been the subject of numerous studies and improvements (e.g., [Vetr11] [Zhan13]), including the research of the author of this dissertation.

In HEVC, motion vectors are coded predictively as a difference between the predicted and the used vector. The encoder can select between two prediction candidates and signals its decision. Contemporary codecs (like HEVC) use a very sophisticated algorithm to derive those prediction candidates [Hell12]. The source vectors are sought in already encoded, neighboring blocks (i.e., neighboring PUs). There is also a contradistinction between “direct” prediction sources (from a block using the same reference picture) and “scaled” prediction sources (from a block using a different reference picture). The prediction vector can also be derived from another frame by analyzing the same coordinate (so-called collocated position). All those prediction candidates from the above mentioned list of sources are then processed, and finally, two motion vector predictors are determined. A simplified illustration of motion prediction candidate derivation can be found in [A12].

As mentioned in section 3.1, motion vectors and disparity vectors need to be treated separately. Therefore, disparity vectors cannot be used as a prediction source for motion vectors (and vice versa). The scaling of disparity vectors should also be adapted to consider distances between different views.

Moreover, the two-dimensional grid of coding structure allows for some additional prediction sources. For example, “Nested Prediction” (NP), proposed by the author of this thesis [A12], allows avoiding situations when there is no possibility to derive candidates from the neighboring unit. Neighboring Prediction (NP) finds new motion predictors by looking at a neighboring unit. First, it follows the neighbor's disparity vector to the unit it points to in the reference frame. Then, if that pointed unit has a

useful motion vector, this vector is scaled to become a prediction candidate for the original unit. Figure 21 shows how this process can indirectly add a new, useful candidate to the prediction pool.

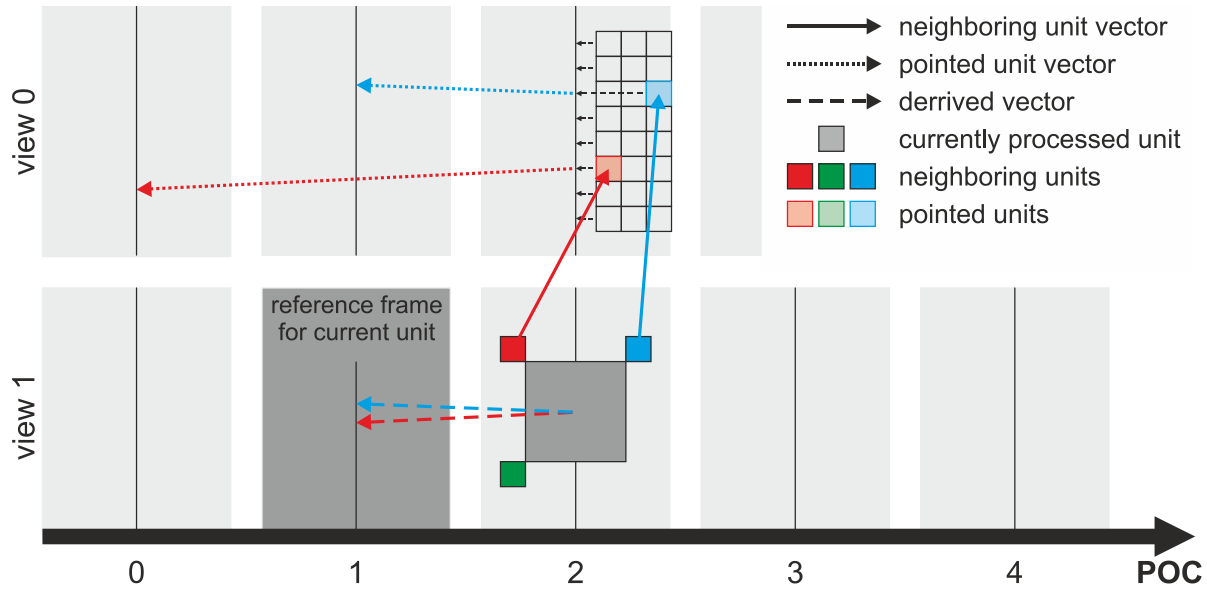


Figure 21. Example of nested prediction (NP). Figure from [A12].

The author's improvements in motion and disparity vector prediction resulted in ~2% improvement in coding efficiency when compared to the multiview codec described in section 3.1. Detailed description of all motion and disparity vector prediction improvements, along with a wide set of performance numbers, is provided in [A12].

3.3 Contribution to the development of a 3D video codec based on HEVC technology

As mentioned at the beginning of paragraph 3, the development of the 3D video codec was teamwork, while the author of this thesis was responsible for selected parts of the new technology development. As a result, a very performant codec was created. The post-CFP evaluation performed in accordance with demanding experimental methodology defined by international groups of ISO/IEC MPEG and ITU-T VCEG experts showed that the proposal submitted by the KTMiM team was one of the two best contributions that took part in the previously mentioned initiative [N12347].

The developed codec was based on an improved multiview HEVC codec, described in section 3.2, with the addition of many new and groundbreaking coding tools [A13]. The introduced tools, among other things, include: Normalized Depth Representation, Disoccluded Region Coding, Joint High-Frequency Layer Representation, and Depth-Based Motion Prediction (DBMP) A13].

It is worth noting that the proposed codec allowed to improve compression efficiency by ~58% and ~70% for 2-view and 3-view case, respectively (see Table 5).

Table 4. 3D codec Bjontegaard metric results for the base view: 2-view and 3-view case.

Sequence	2-view case		3-view case	
	ΔPSNR [dB]	**ΔBitrate [%]**	**ΔPSNR [dB]**	**ΔBitrate [%]**
Poznan_Hall2	1.8	-45.3	2.5	-59.3
Poznan_Street	2.7	-54.6	3.9	-69.6
Undo_Dancer	1.9	-51.0	2.7	-62.6
GT_Fly	2.2	-55.5	3.1	-67.0
Kendo	4.2	-58.8	6.2	-73.3
Balloons	5.2	-64.6	6.7	-74.3
Lovebird1	4.2	-66.7	5.8	-76.4
Newspaper	5.1	-66.5	6.3	-76.1
Average	**3.4**	**-57.9**	**4.7**	**-69.8**

The technical description of those proposals can be found in [M22697]. The detailed study of all algorithms and tools can be found in [A13]. Some selected details of the proposed solution can also be found in the following publications [Doma12a][Doma12b][Doma12c].

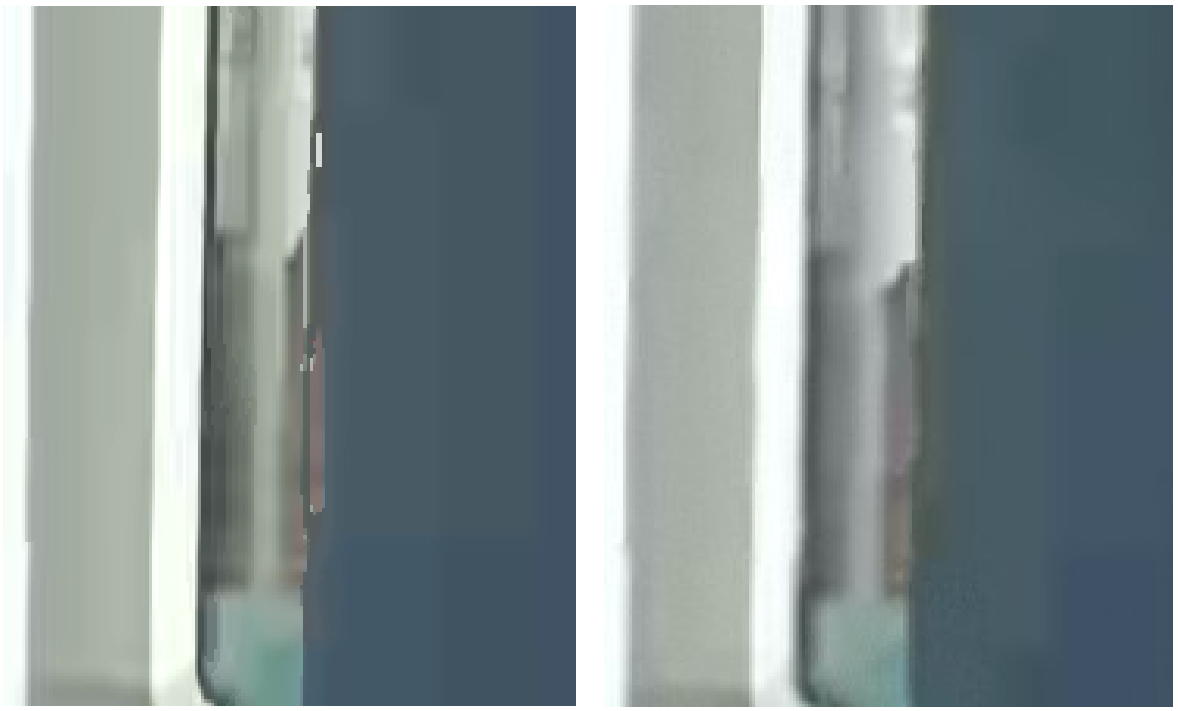
Selected technologies and solutions invented during codec development are covered by patents granted by the United States Patent and Trademark Office [US8761527][US9582859][US9646385] [US9848205][US9992514] or European Patent Office [EP3110149][EP3110155][EP3110156].

3.3.1 Development of a distortion estimation algorithm for depth map coding

In addition to the techniques described in Sections 3.1 and 3.2, the developed 3D codec contains an improved Rate-Distortion optimization module (RDopt) developed by the author of that thesis.

The typical metric of the degree of similarity of two image blocks that is used during Rate-Distortion optimization is the Sum of Squared Differences (SSD). This is a reasonable approach since the encoder tries to achieve the best quality (defined as PSNR) with the lowest possible bitrate. SSD works reasonably well when coding natural images (however, advocates of Structural Similarity [Wang04] related metrics will not agree). Unfortunately, depth maps have different characteristics from natural images.

Moreover, there are some types of distortion that are unnoticeable in the decoded image, but when they appear in the depth map, they can degrade the quality of the synthesized image. One of those types of distortion is a slight shift in edge location. As mentioned in [Dzie22], a small change in object location is unnoticeable for the viewer. However, a slight distortion of the depth map edge can lead to tearing apart of the object edge [Dzie22].



HEVC RD-opt
Visible artifacts in synthetized view

Improved RD-opt
Artifacts reduced

Figure 22. Example of distortion appearing in the synthesized view for original and improved RDopt.

To reduce this effect, an improved Rate-Distortion Optimization (RDopt) mechanism was created for encoding depth maps. The method modifies the distortion metric by switching to the Sum of Squared Transformed Differences (SSTD) metric and adding a new component: a measure of the edge energy in the residual signal. This author's change boosts depth map coding efficiency and reduces the most noticeable visual artifacts (Figure 22).

3.4 Development of FTV video codec with arbitrary camera location

The finalization of HEVC technology was followed by the development of several specialized extensions. Similarly to AVC, the standardization committee worked on "Range extension" (RExt) [Flyn16], scalable extension [Boyc16], and multiview extension [Tech16]. However, the multiview extension of HEVC technology included two separate encoding modes: "regular" multiview (MV-HEVC) and multiview+depth (3D-HEVC) variants.

The MV-HEVC reassembles the idea taken from MVC (the multiview extension of AVC) [MVC]. The MV-HEVC operation is straightforward – it just allows the use of other views as reference pictures. However, the inter-view reference pictures are treated in the same way as long-term reference pictures (LTRP). This means that some sophisticated techniques for motion vector prediction (like motion vector predictor scaling) are disabled.

On the other hand, the 3D-HEVC introduces several new ideas. First of all, it introduces a completely new component type – dept map, with a dedicated coding tool [3DHEVC]. The second novelty was an idea to use scene geometry information (carried by the depth map) to improve inter-view prediction. The prediction signal can be calculated by reprojecting the content of reference view(s) into the currently coded view. The process of calculation of the prediction signal is very similar to operations used during virtual view synthesis [Fach18].

The creators of 3D-HEVC make some assumptions about the encoded content features. The most significant drawbacks of this technology are restrictions related to camera location in multiview systems. The 3D-HEVC assumes that all cameras are arranged along a single line, optical axes perpendicular to this line and parallel to each other (see Figure 23). This type of multiview system is known as linear [M17050] and was popular during the early stage of development of DIBR[Fehn04] systems. This assumption significantly simplifies the reprojection process – in such systems, the reprojection is performed as a simple pixel shift along the horizontal axis.

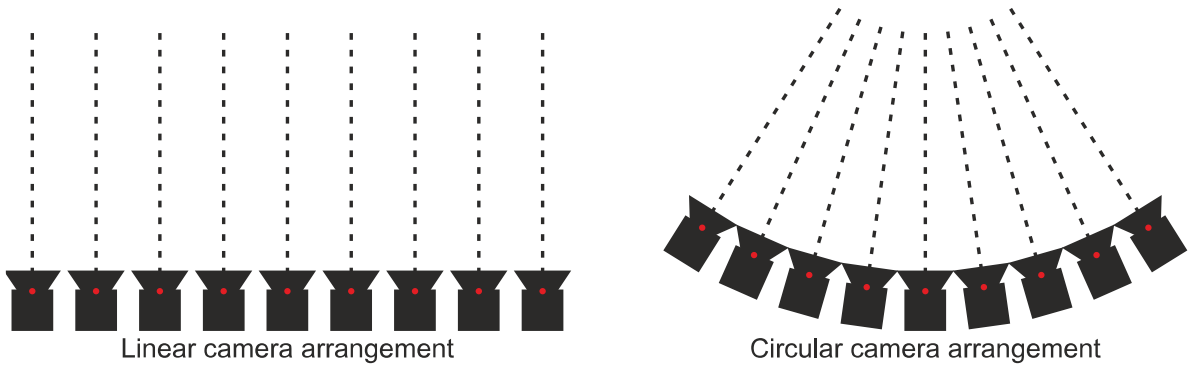


Figure 23. Linear vs circular camera arrangement in multi-view systems.

The following development of multiview systems [M36569] shows that a linear camera arrangement is suboptimal. For example, the systems with circular camera arrangement offered wider freedom of navigation and better user experience (Figure 23). There is also a class of multiview systems where cameras are organized in a 2D grid [M40010], which allows navigation along two axes – horizontal and vertical.

Those shortcomings of 3D-HEVC technology lead to the idea of creating a 3D-HEVC-based codec adapted to arbitrary locations of cameras. The work toward the development of those codecs was performed by the team from the Chair of Multimedia Communication and Microelectronics (KTMiM) during the realization of the OPUS [OPUS12] project funded by the National Science Centre. The author of this dissertation played a significant role in the work of this team.

The work towards generalization of 3D-HEVC coding technology concentrated on circular systems; nevertheless, the created codec is compatible with any arbitrary camera locations, i.e. 2D grid of cameras. The developed codec uses a full set of camera parameters [Cyga09] (intrinsic and extrinsic

camera matrices – representing sensor size/position, lens focal length, camera location in 3D space, rotation, etc.). It also generalizes the processing of any inter-view and depth-related tool, i.e., inter-view prediction.

Table 5 Coding performance results for the codec with an arbitrary camera location

View Arrangement	Sequence	Proposed vs 3D-HEVC	Proposed vs MV-HEVC	3D-HEVC vs MV-HEVC
Linear	Poznan Street	0.05	-7.10	-7.11
	Poznan Hall 2	0.09	-9.14	-9.18
	Dancer	0.02	-14.34	-14.36
	Balloons	0.08	-8.21	-8.20
	Kendo	0.09	-6.05	-6.08
	Newspaper	0.04	-14.26	-14.28
	Average	**0.06**	**-9.85**	**-9.87**
Arc	Poznan Blocks	-6.17	-11.11	-5.28
	Ballet	-5.68	-8.82	-3.35
	Break Dancers	-6.89	-5.94	1.02
	Average	**-6.25**	**-8.62**	**-2.54**

As a result, the proposed coding technology allows for exploiting inter-view dependencies in a system with an arbitrary camera location and significantly improves compression efficiency. The evaluation result was summarized in Table 5. When compared to 3D-HEVC, the proposed codec improves compression efficiency by ~6.25% for sequences with circular camera arrangement.

The detailed description of a 3D-HEVC based codec with arbitrary camera locations can be found in [A11]. Some selected details of the proposed solution can also be found in the following publications [Same16][Doma15][Doma16].

3.5 Chapter summary and original achievements of the research

The original research results described in this chapter consist of several aspects related to the effective coding of multiview and 3D (texture+depth) content. The multiview video codec based on HEVC technology, equipped with the author's advanced motion and disparity vectors prediction algorithm, was developed [A12]. This multiview HEVC-based codec was then the basis for further development of a 3D video codec based on HEVC technology [A13]. Obviously, the 3D video codec was a team effort, but the author of this thesis has a noticeable contribution into development of technology, which was one of the two best solutions submitted in response to the MPEG Call for Proposals. In addition to the aforementioned multiview HEVC, the authors' contribution was also dedicated distortion estimation

algorithm for depth map coding. The follow-up research is the FTV video codec with arbitrary camera location, described in [A11].

4 Compressed bitstream watermarking

Watermarking of digital video content has a number of important applications, i.e., content protection, detection of unauthorized distribution of a video, etc. Therefore, it was the subject of research by the author of this dissertation. In this chapter, the author uses a warermarked video as a communication channel, allowing the transmission of some amount of additional data; however, without inquiring about the content of the transmitted data. In the described case, the watermark should be invisible (or unnoticeable) from the perspective of perceiving the content of subsequent images of a video, but at the same time robust against common attacks (for example, it has to survive after strong image processing, including camcording and recompression).

The most common approach to video watermarking is based on embedding messages directly into uncompressed video, which needs full decoding and then full encoding of the content (Figure 24). It allows for great flexibility, since the watermark embedder can alter any pixel value. Unfortunately, from the perspective of many applications, such a method is too computationally complex, which significantly limits the practical application of this approach. From this perspective, there is a need to develop a method for watermarking video data that operates in the domain of compressed data. The author dealt with this problem in his research.

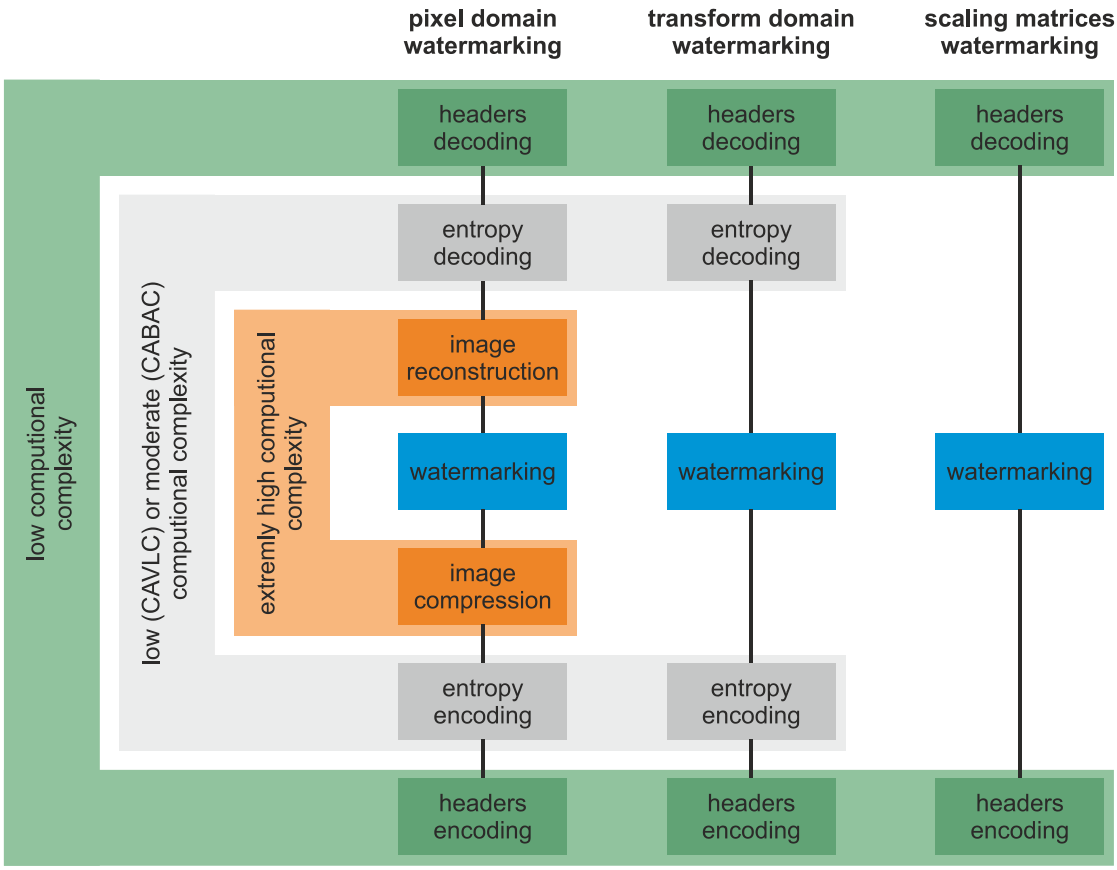


Figure 24. Overview of the most common approaches to compressed bitstream watermarking. Figure from [A16].

Watermarking of compressed bitstreams does not require full decoding and full encoding of a video. Depending on the chosen scenario, the watermark embedding process requires only partial decoding/encoding and does not involve full image reconstruction.

As presented in Figure 24, if full decoding and re-encoding have to be avoided, there are two remaining approaches to compressed domain watermarking: scaling matrices watermarking and transform domain watermarking [Doër03].

The scaling matrices watermarking requires processing only header information (i.e., VPS, SPS, PPS, and Slice Header in HEVC or SPS, PPS, and Slice Header in AVC), which results in negligible computational complexity. However, this method has very limited capacity and control over watermark embedding and introduce significant risk of drift-related distortion.

The transform domain watermarking is slightly more computationally demanding. The embedder needs to perform entropy decoding and entropy encoding of video data. Those gave him access to data of compression modes, transform coefficients, motion vectors, etc. It should be noted that entropy decoding constitutes about 20% of total decoder complexity [Boss12] while on the encoder side, the final entropy encoding is definitely below 1% of total encoder runtime. Those method allows for a higher number of possible embedding opportunities, drift reduction, and local watermark intensity adjustment. The two mentioned types of watermarking have been the subject of analysis in the most advanced studies, including the research of the author of the dissertation.

Both mentioned compressed domain watermarking approaches are very interesting, and both were evaluated during research conducted by the author of this thesis and his coworkers. To be specific, the author of this thesis was responsible for:

- a development af scaling matrices watermarking technique for AVC-compressed bitstreams,
- co-development of transform coefficients watermarking technique for AVC-ompressed bitstreams,
- development of transform coefficients watermarking technique for HEVC-ompressed bitstreams,
- an algorithm for watermark detection, compatible with all the above-mentioned watermark embedding techniques.

4.1 Watermarking suite for AVC-compressed bitstreams

In the context of watermarking video data, the author conducted his research in relation to the AVC technique. The work towards the development of a watermark embedding suite for AVC-compressed bitstreams was based on the idea of using low-frequency AC coefficients for watermark embedding.

This decision was preceded by exhaustive research leading to observations that this type of watermark has the best robustness against attacks.

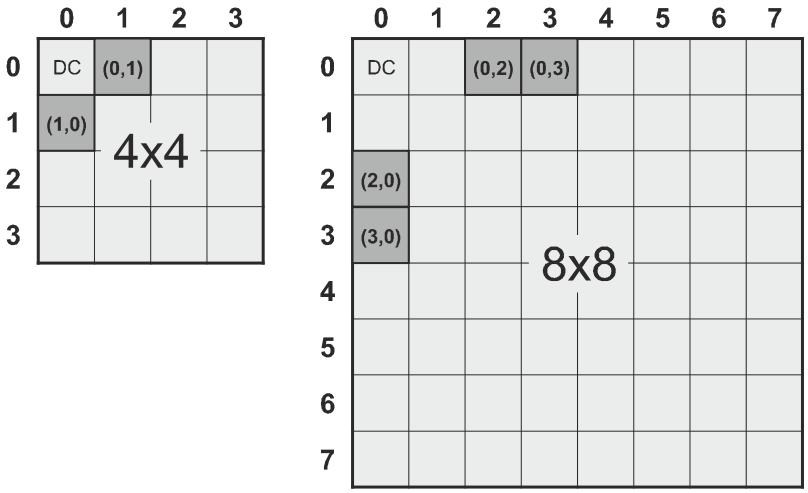


Figure 25. DCT coefficients that are modified during the watermarking process. Figure from [A16].

Two embedding algorithms were developed: one using scaling matrices (created solely by the author of this thesis) and the second using the transform coefficients approach. Both algorithms operated by modifying the low-frequency AC coefficients of the residual signal (as shown in Figure 25). The 0 and 1 bits were signaled by increasing the energy of low-frequency horizontal and vertical components, respectively.

The developed algorithms were then implemented and changed into usable software tools. In the proposed approach, the watermark was embedded in the second half of the GOP, and only P- and B-frames were modified – this allowed for reducing complexity and drift-related distortion. The assumptions mentioned here were an important author's element of the developed watermarking method.

The scaling matrices approach included a set of novel techniques, like: on-the-fly conversion of AVC main profile bitstreams into AVC high profile bitstreams and creation of specialized Picture Parameter Sets containing dedicated Scaling Matrices.

The second proposed method uses a transform coefficients watermarking approach. In that case embedder decodes the transform coefficients using the entropy (CABAC or CAVLC) decoder. Then modifies the selected transform coefficients (according to the scheme described above). In the last phase, the entropy encoding (CABAC or CAVLC) is performed.

In addition, the implemented embedder has a special mode allowing to reduce the complexity of subsequent data decoding at the cost of a slight reduction of compression efficiency – the modified slice of a video frame can be encoded using a faster CAVLC entropy encoder instead of using a slower

CABAC. This last solution allowed the development of a new functionality that could find practical application.

The watermark detection is based on the difference in energy of transform coefficients between the original and the analyzed video sequence. Since the watermarking process disrupts the relation between the average energy of horizontal and vertical frequencies within a frame, the presence of watermark data can be detected by statistical analysis of DCT-transformed frames within a GOP.

Both methods have been evaluated using very long (>=10 minutes, 15000 frames) FullHD (1920x1080) sequences. Those sequences were encoded with the use of classic, closed GOP, with the number of frames in GOP corresponding to half a second of video. The evaluation was performed in the following steps: a watermark was embedded, then attacks against the watermark were carried out, the watermark was detected, and the percentage of correctly detected watermark bits was calculated. The robustness of the algorithm was verified by using three severe attack types resembling real-world scenarios:

- Scenario 1 – Camcording using an HD camera,
- Scenario 2 – Camcording using an HD camera and downscaling to SD resolution,
- Scenario 3 – Camcording using an HD camera and framerate reduction (from 25 to 12,5 fps).

The obtained evaluation results, summarized in Table 6, show that the proposed watermarking can survive even severe attacks like screen camcording while preserving a bit error ratio within a range that can be corrected by properly designed forward error correction codes.

Table 6. Summary of watermark robustness evaluation for both described methods.

Watermarking method	**Properly detected bits [%] in an attack scenario**		
	Camcording	Camcording + downscaling	Camcording + framerate reduction
Scaling matrices	88.0	84.8	87.1
Transform coefficients	86.0	80.5	84.4

The detailed description of the adaptation of the scaling matrices watermarking idea to AVC-compressed bitstreams can be found in [A15], and the transform coefficients watermark embedding and watermark strength adjustment algorithm for AVC can be found in [A16].

It is worth noting that both of the above-mentioned embedding methods used the same detection algorithm described in [A15]. Therefore, they can be used to create a consistent watermarking system with two embedding modes differing in complexity and a single detection method.

4.2 Watermark embedding algorithm for HEVC

The idea of transform domain watermarking was also applied by the author of this thesis to the newer technology of video compression – HEVC. Unfortunately, as it turned out during the research, it was impossible to directly transfer the approach used for AVC technology to HEVC video compression due to significant differences between those two techniques.

Early experiments showed that watermarking based on scaling matrices (done according to [A15]) does not work correctly in HEVC due to the intensive use of intra-coded blocks in some of the B frames. Those lead to strong, uncontrolled drift and annoying artifacts in video images after watermarking.

The transform coefficients watermarking (based on [A16]) was more promising in terms of adaptation to HEVC bitstreams. However, the HEVC coding scheme with multiple unit types (CUs, PUs, and TUs) introduced several difficulties. To efficiently solve the existing limitations, the author developed several additional solutions that were used by the developed watermarking algorithm. These were:

- Distortion-optimized selection of TUs used to embed a watermark.
- Adaptive drift avoidance with detailed analysis of neighboring units and tracking of possible intra prediction sources.
- Complex calculation of watermarked transform coefficients values in watermarked TUs due to interaction with the Sign Data Hiding (SDH) tool, introduced in HEVC [G271].

Even though all the above-mentioned algorithms noticeably increased the complexity of the proposed solution, the overall embedding computational complexity was still orders of magnitude lower than that for pixel domain watermarking.

The watermark embedded in the HEVC bitstream can be detected using exactly the same algorithm as watermarks embedded in AVC bitstreams using methods [A15] and [A16]. This feature creates a unique possibility to use the algorithms proposed by the author of this thesis to build a coherent system using compatible watermarks for both AVC and HEVC bitstreams.

In order to evaluate the proposed method, the author of this thesis implemented the entire watermarking software, including an extremely complex HEVC entropy decoder/encoder. The created software was able to watermark typical HEVC bitstreams.

The evaluation procedure was based on the same set of long sequences as described in Section 4.1. Due to a lack of access to hardware used in earlier experiments, attacks were simulated in order to mimic the distortion caused by the TV-to-camera setup. The robustness was verified by using two attack scenarios:

- Scenario 1 – corresponding to screen-captured content (i.e., VODRip): HEVC reencoding with ½ bitrate and "fast" preset (using x265 encoder [x265]),
- Scenario 2 – corresponding to camera recorded content (i.e., CamRip/Telecine): a strong deblocking filter (in 8x8 regions), slight image rotation, dynamic range clip to 0.1-0.9 of the original range of sample values, dynamic range expansion (contrast enhancement), HEVC encoding with ½ of original bitrate with the usage of "fast" preset (using x265 encoder [x265]).

The evaluation results (Table 7) showed that the embedded watermark is almost invisible, even though the introduced distortion (measured using PSNR) is ~0.3dB and ~1.1dB for weak and strong watermarks, respectively. The weak watermark is robust against screen capture, and the strong watermark is robust against camera recording.

Table 7. Experimental results for transform coefficient watermarking in HEVC.

Parameters	Targeted attack scenario	
	screen capture (VODRip)	camera-recorded (CamRip/Telecine)
Watermark strength	1.8	2.2
Correctly detected bits [%]	93.3	94.3
Bitrate increase [%]	0.313	1.790
Decoded image quality change (delta PSNR) [dB]	-0.332	-1.087

A detailed description of the Fast and Energy-Efficient Watermarking of HEVC-Compressed Video Bitstreams technique, including an extensive set of experimental results, can be found in [A14].

4.3 Chapter summary and original achievements of the research

The original research results described in this chapter are a consistent framework for compressed video bitstream watermarking. The results include the development of the original AVC-compressed bitstream watermarking based on modification of scaling matrices (transmitted in an encoded data stream) [A15], and the co-development of transform coefficients watermark embedding together with a watermark strength adjustment algorithm for AVC [A16]. The sole contribution of the thesis author is the proposal of fast watermarking of HEVC-compressed video bitstreams based on the idea of transform coefficients watermarking [A14]. This HEVC-compressed video watermarking technique includes several additional sophisticated tools allowing for embedding a robust watermark without introducing noticeable quality degradation and unnecessary bitstream expansion. It is worth mentioning that the watermark embedded with any of the proposed embedding algorithms can be detected by the same detector. This feature creates a unique possibility to use the proposed algorithms to build a coherent system using compatible watermarks for both AVC- and HEVC-encoded bitstreams.

5 Summary and conclusions

5.1 Recapitulation

The research presented in this dissertation addresses the critical challenges associated with modern hybrid video coding, focusing mainly on the most widely adopted standardised techniques. The work covers a logical sequence of development, moving from the fundamental statistical analysis of the HEVC technology, through the optimization of encoder algorithms and implementations, to the extension of the technology for multiview applications and practical use cases like compressed bitstream watermarking. The obtained results confirm the validity of the doctoral theses formulated in the introduction and demonstrate significant contributions to the field of video compression.

The research was conducted over a multi-year period in parallel with HEVC development and early adoption. Therefore, a significant part of the thesis contains research results obtained by using HEVC technology. However, drawn conclusions are universal and are applicable to any hybrid video compression technology, especially to HEVC successors. (e.g., ECM).

This dissertation takes the form of a collection of published, peer-reviewed, and thematically related scientific articles, thus meeting the requirements of Polish Law on Higher Education and Science (*"Prawo o szkolnictwie wyższym i nauce"*), article 187, paragraph 3, as formulated for doctoral dissertations. The publication series was supplemented with a self-report presenting the research works performed, research conclusions, and examples of the practical use of the developed solutions.

5.2 Original achievements of the research

The author's research presented in this thesis is part of the current research on efficient representation of a video through the use of modern hybrid video compression techniques and their improvements. The author's achievements and contributions that have been obtained for this area can be arranged into four research directions, described in the following chapters: (1) Analysis of video coding efficiency and statistical properties of data in the encoder, (2) Video encoder optimizations, (3) Extensions for encoding of multiview and FTV content, and (4) Compressed bitstream watermarking.

(1) The analysis of video coding efficiency and statistical properties of data in the encoder, discussed in Chapter 1, includes:

- Analysis of frame partitioning into blocks [A1].
- Analysis of the contribution of individual frame types into total bitrate of encoded data [A2].
- Analysis of bitrate distribution of syntax elements in the HEVC [A3].

- Analysis of the efficiency of estimating the statistics of syntax elements in CABAC [A4].
- Analysis of quality losses in homogenous transcoding and multi-generation encoding [A5][A6].

The obtained knowledge about the statistical properties of the encoded video data has scientific value in itself. However, the author of the thesis conducted additional research aimed at the practical application of the previously obtained data. This led to the development of:

- Video encoder with mode decision assisted by pre-computed statistics of data [A7] – the solutions allowed for a significant reduction of the video encoder runtime while still maintaining high compression efficiency of a video.
- Encoder with mode decision assisted by noise reduction [Stan18a] – the inspiration for the proposal allowed for improving compression efficiency by reducing the negative effect of the greedy coding mode selection approach.
- Improvements that have been applied to video compression technology developed as a successor to VVC [AL0198][AK0173] – both proposed methods delivered tangible improvements in compression efficiency, which is a significant achievement, considering the very high level of advancement and efficiency of modern video compression techniques.

(2) Video encoder optimizations, discussed in Chapter 2, include:

- Improved Rate-Distortion Optimized Quantization (RDOQ) [A8] – a non-normative, encoder-only improvement, allowing for noticeable improvement (~1%) in compression efficiency by adaptively adjusting quantized transform coefficients.
- Optimized motion estimation in contemporary hybrid video encoding [A9][A10] – a motion estimation efficiency and performance improvements (by adaptively selecting: starting point, search range, and distortion metric), allowing for a reduction of encoding time (while still maintaining high compression efficiency).

The doctoral **Thesis 1** formulated at the outset has been validated in **Chapters 1 and 2**.

(3) Extensions for encoding of multiview and FTV content, discussed in Chapter 3, include:

- Development of a multiview video codec based on HEVC technology with motion and disparity vectors prediction algorithm [A12] – resulted in the creation of a multiview codec with an advanced motion and disparity vectors prediction algorithm and paved the way for further development of the HEVC-based 3D codec.

- Contribution to the development of a 3D video codec based on HEVC technology [A13], including a dedicated distortion estimation algorithm for depth map coding – resulted in the development of technology that was one of the two best responses to "Call for Proposals on 3D Video Coding Technology".
- Development of FTV video codec with arbitrary camera location [A11] – allowed to significantly improve 3D-HEVC efficiency for sequences with more complex (e.g., circular) camera arrangements, by using full camera parameters (intrinsic and extrinsic) for calculation of prediction signal.

The doctoral **Thesis 2** formulated at the outset has been validated in **Chapter 3**.

(4) Compressed bitstream watermarking, discussed in Chapter 4, includes:

- Watermark embedding suite for AVC compressed bitstreams [A15] [A16].
- Watermark embedding algorithm for HEVC [A14].

Coherent solution allowing the use of compatible watermarks for both AVC and HEVC bitstreams, including a low complexity technique for robust watermark embedding into AVC/HEVC-compressed bitstreams (by adapting scaling matrices or modifying DCT coefficients), along with a watermark detection algorithm.

The doctoral **Thesis 3** formulated at the outset has been validated in **Chapter 4**.

5.3 Conclusions

This dissertation presents a comprehensive work that advances the field of hybrid video coding. The research navigated from the statistical analysis of a new standard (HEVC) to the practical implementation of optimized encoders, and finally to the expansion of the technology into new domains like 3D video and practical usability of the compression technique-based systems (watermarking).

The practical applicability of the research is evidenced by the implementation of these algorithms in functional software and their potential for integration into commercial systems. Furthermore, the knowledge gathered during this research has been actively applied to the development of the current state-of-the-art technique – the Enhanced Compression Model (ECM) – demonstrating the continued relevance of the findings. The optimizations and extensions proposed herein provide valuable tools for reducing bandwidth requirements and computational costs in an era where video data dominates global telecommunication networks.

6 References

6.1 Video compression standards, techniques, and implementations

[H261] International Telecommunication Union, “H.261: Video codec for audiovisual services at p x 384 kbit/s”, ITU-T Recommendation H.261 (03/93), Geneva, Switzerland, 1993.

[MPEG1] International Organization for Standardization, “Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s — Part 2: Video”, ISO/IEC 11172-2:1993, Geneva, Switzerland, 1993.

[MPEG2] International Organization for Standardization, “Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 2: Video”, ISO/IEC 13818-2:2013, Geneva, Switzerland, 2013.

[H262] International Telecommunication Union, “H.262: Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Video”, ITU-T Recommendation H.262 (02/12), Geneva, Switzerland, 2012.

[AVC] International Organization for Standardization, “Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding”, ISO/IEC 14496-10:2022, Geneva, Switzerland, 2022.

[SVC] International Organization for Standardization, “Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding — Amendment 3: Scalable Video Coding”. ISO/IEC 14496-10:2005/Amd 3:2007, Geneva, Switzerland, 2007.

[MVC] International Organization for Standardization, “Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding — Amendment 1: Multiview Video Coding”, ISO/IEC 14496-10:2008/Amd 1:2009, Geneva, Switzerland, 2009

[H264] International Telecommunication Union, “H.264: Advanced Video Coding for Generic Audiovisual Services”, ITU-T Recommendation H.264 (08/24), Geneva, Switzerland, 2024.

[HEVC] International Organization for Standardization, “Information technology — High efficiency coding and media delivery in heterogeneous environments (MPEG-H) Part 2: High efficiency video coding”, ISO/IEC 23008-2:2023, Geneva, Switzerland, 2023.

[3DHEVC] International Organization for Standardization, “Information Technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High Efficiency Video Coding — Amendment 1: 3D Video Extensions”, ISO/IEC 23008-2:2015/Amd 1:2015, Geneva, Switzerland, 2015.

[H265] International Telecommunication Union, “H.265: High efficiency video coding”, ITU-T Recommendation H.265 (07/24), Geneva, Switzerland, 2024.

[VVC] International Organization for Standardization, “Information Technology — Coded Representation of Immersive Media — Part 3: Versatile Video Coding”, ISO/IEC 23090-3:2024, Geneva, Switzerland, 2024.

[H266] International Telecommunication Union, “H.266: Versatile video coding”, ITU-T Recommendation H.266 (09/23), Geneva, Switzerland, 2023.

[JM] International Organization for Standardization, “Information Technology — Coding of Audio-Visual Objects — Part 5: Reference Software”, ISO/IEC 14496-5:2001. Geneva: International Organization for Standardization, 2001.
Software available at: https://iphome.hhi.de/suehring/tml/download/

[x264] VideoLAN, “x264: A Free H.264/AVC Video Encoder.” Accessed December 17, 2025. Software available at: https://www.videolan.org/developers/x264.html

[HM] International Organization for Standardization, “Information Technology — High Efficiency Coding and Media Delivery in Heterogeneous Environments — Part 5: Reference Software for High Efficiency Video Coding”, ISO/IEC 23008-5:2015, Geneva, 2015.
Software available at: https://vcgit.hhi.fraunhofer.de/jvet/HM

[x265] MulticoreWare, “x265: High Efficiency Video Coding (HEVC) Encoder”, 2024. Software available at: https://www.x265.org/

[VTM] International Organization for Standardization, “Information Technology — Coded Representation of Immersive Media — Part 16: Reference Software for Versatile Video Coding”, ISO/IEC 23090-16:2024. Geneva, Switzerland, 2024.
Software available at: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM

[ECM] M. Coban, R.-L. Liao, K. Naser, J. Ström, L. Zhang, “Algorithm description of Enhanced Compression Model 19 (ECM 19)”, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Doc. JVET-AN2025, Geneva, Switzerland, October 2025.

[RFC2435] L. Berc, W. Fenner, R. Frederick, S. McCanne, P. Stewart, “RTP Payload Format for JPEG-compressed Video”, RFC 2435. Internet Engineering Task Force, October 1998. https://doi.org/10.17487/RFC2435.

[MJP2K] International Organization for Standardization, “Information Technology — JPEG 2000 image coding system — Part 3: Motion JPEG 2000”, ISO/IEC 15444-3:2007, Geneva, 2007.

[PNG] International Organization for Standardization, “Information Technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification”, ISO/IEC 15948:2004. Geneva, 2004.

[APNG] C. Lilley, “Portable Network Graphics (PNG) Specification (Third Edition) Implementation Report”, Editor's draft. World Wide Web Consortium, 2024
Available at: https://w3c.github.io/png/Implementation_Report_3e/

[Huffyuv] B. Rudiak-Gould, “Huffyuv: Lossless Video Codec”, 2002. [Online]
Available at:
https://web.archive.org/web/20051107033649/http://neuron2.net/www.math.berkeley.edu/benrg/huffyuv.html#Tech

6.2 Patents

[EP3110149] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz**,** **J. Stankowski**, K. Wegner, “A system and a method for depth-image-based rendering”, Patent office: EPO, Status: granted, Application number: EP 15461540.5, Filling date: 25.06.2015, Publication number: EP 3110149A1, Publication date: 28.12.2016, Patent number: EP 3110149, Date of Patent: 02.01.2019

[EP3110155] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, “A method for predictive coding of depth maps and a depth map encoder”, Patent office: EPO, Status: granted, Application number: EP 15461541.3, Filling date: 25.06.2015, Publication number: EP 3110155A1, Publication date: 28.12.2016, Patent number: EP 3110155, Date of Patent: 21.06.2017,

[EP3110156] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, "A system and a method for disoccluded region coding in a multiview video data stream", Patent office: EPO, Status: granted, Application number: EP 15173955.4, Filling date: 26.06.2015, Publication number: EP 3110156A1, Publication date: 28.12.2016, Patent number: EP 3110156, Date of Patent: 14.03.2018

[US8761527] M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, "Image coding method", Patent office: USPTO, Status: granted, Application number: US 13/680652, Filling date: 19.11.2012, Publication number: US 2013/0129235A1, Publication date: 23.05.2013, Patent number: US 8761527, Date of Patent: 24.06.2014

[US9582859] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, "A system and a method for depth-image-based rendering", Patent office: USPTO, Status: granted, Application number: US 14/753015, Filling date: 29.06.2015, Publication number: US 2016/0379341, Publication date: 29.12.2016, Patent number: US 9582859, Date of Patent: 28.02.2017

[US9646385] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, "A system and a method for generating a depth map", Patent office: USPTO, Status: granted, Application number: US 14/665004, Filling date: 23.03.2015, Publication number: US 2016/0275691, Publication date: 22.09.2016, Patent number: US 9646385, Date of Patent: 09.05.2017

[US9848205] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, "A method for predictive coding of depth maps and a depth map encoder", Patent office: USPTO, Status: granted, Application number: US 14/753015, Filling date: 29.06.2015, Publication number: U S2016/0381387, Publication date: 29.12.2016, Patent number: US 9848205, Date of Patent: 19.12.2017

[US9992514] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, "A system and a method for disoccluded region coding in a multiview video data stream", Patent office: USPTO, Status: granted, Application number: US 14/753010, Filling date: 29.06.2015, Publication number: US 2016/0381390, Publication date: 29.12.2016, Patent number: US 9992514, Date of Patent: 05.06.2018

6.3 Other references

The list of all remaining references also includes the results of the author's other research that is not directly part of the publication cycle.

[AH21] M. Karczewicz, Y. Ye, I. Chong, "Rate Distortion Optimized Quantization", document ITU-T SG16 Q.6, VCEG-AH21, Antalya, Turkey, 2008

[AI2017] M. Karczewicz, Y. Ye, "Common test conditions and evaluation procedures for enhanced compression tool testing", Joint Video Experts Team of ITU-T and ISO/IEC, in Document JVET-AI2017, Sapporo, Japan, 2024.

[AK0171] D. Mieloch, M. Lorkiewicz, **J. Stankowski**, "Non-EE2: Picture-level mirroring and rotation", Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Doc. JVET-AK0171, Geneva, Switzerland, 2025

[AK0173] D. Mieloch, M. Lorkiewicz, A. Dziembowski, **J. Stankowski**, D. Karwowski, "Non-EE2: Adaptive picture-level vertical mirroring", Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Doc. JVET-AK0173, Geneva, Switzerland, 2025,

[AL0198] D. Karwowski, D. Mieloch, M. Lorkiewicz, **J. Stankowski**, "Non-EE2: Optimization of probability estimation in CABAC", Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Doc. JVET-AL0198, Teleconference, 2025.

[AL0199] D. Mieloch, **J. Stankowski**, M. Lorkiewicz, A. Dziembowski, D. Karwowski, "EE2-6.1: Adaptive picture-level vertical mirroring", Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Doc. JVET-AL0199, Teleconference, 2025.

[Ali92] M. Ali, C. Papadopoulos, T. Clarkson, "The use of fractal theory in a video compression system," Data Compression Conference, 1992, Snowbird, UT, USA, 1992, pp. 259-268, doi: 10.1109/DCC.1992.227455.

[Bitm24] Bitmovin, "The 8th Annual Bitmovin Video Developer Report: Shaping the future of video", Bitmovin Inc., 2024.
Available at: https://bitmovin.com/video-developer-report/

[Bitm25] Bitmovin, "The 9th Annual Bitmovin Video Developer Report: Insights Shaping the Future of Streaming", Bitmovin Inc., 2025.

[Boss12] F. Bossen, B. Bross, K. Suhring, D. Flynn, "HEVC Complexity and Implementation Analysis", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1685-1696, 2012, doi: 10.1109/TCSVT.2012.2221255

[Boyc16] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 20-34, 2016, doi: 10.1109/TCSVT.2015.2461951

[Bros21] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, J.-R. Ohm, "Overview of the Versatile Video Coding (VVC) Standard and its Applications" IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 3736-3764, 2021, doi: 10.1109/TCSVT.2021.3101953

[Cisc20] Cisco Systems Inc, "Cisco Annual Internet Report (2018–2023) White Paper", USA, 2020.
Available at: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[Cyga09] B. Cyganek, J. P. Siebert, "An Introduction to 3D Computer Vision Techniques and Algorithms", John Wiley and Sons, 2009, doi: 10.1002/9780470699720

[Doër03] G. Doërr, J.-L. Dugelay: "A guide tour of video watermarking", Signal Processing: Image Communication, vol. 18, no. 4, pp. 263-282, Elsevier 2003.

[Doma12a] M. Domański, T. Grajek, D. Karwowski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, "Coding of multiple video+depth using HEVC technology and reduced representations of side views and depth maps", 29th Picture Coding Symposium, PCS 2012, pp. 5-8, Kraków, Poland, 2012

[Doma12b] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, "New coding technology for 3D video with depth maps as proposed for standardization within MPEG", 19th International Conference on Systems, Signals and Image Processing, IWSSIP 2012, pp. 401-404, Vienna, Austria, 2012

[Doma12c] [poznancodecicip2] M. Domański, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, "3D video compression by coding of disoccluded regions", 19th IEEE International Conference on Image Processing (ICIP 2012), pp. 1317-1320, Orlando, Florida, U.S.A., 2012

[Doma15] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, Ł. Kowalski, M. Kurc, A. Łuczak, D. Mieloch, R. Ratajczak, J. Samelak, O. Stankiewicz, **J. Stankowski**, K. Wegner, "Methods of High Efficiency Compression for Transmission of Spatial Representation of Motion Scenes", IEEE International Conference on Multimedia and Expo ICME 2015, Torino, Italy, 2015

[Doma16] M. Domański, M. Bartkowiak, A. Dziembowski, T. Grajek, A. Grzelka, A. Łuczak, D. Mieloch, J. Samelak, O. Stankiewicz, **J. Stankowski**, K. Wegner, "New results in free-viewpoint television systems for horizontal virtual navigation", 2016 IEEE International Conference on Multimedia and Expo ICME 2016, Seattle, USA, 2016

[Dzie22] A. Dziembowski, D. Mieloch, **J. Stankowski**, A. Grzelka, "IV-PSNR – the objective quality metric for immersive video applications", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 32, pp. 7575 - 7591, 2022, doi: 10.1109/TCSVT.2022.3179575,

[Eric24] Ericsson, "Ericsson Mobility Report: November 2024", Stockholm, Sweden, 2024. Available at: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2024

[Fach18] S. Fachada, D. Bonatto, A. Schenkel, G. Lafruit, "Depth image based view synthesis with multiple reference views for virtual reality", 3DTV Conference, Stockholm, Sweden, 2018

[Fehn04] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV", SPIE Proceedings, 5291, pp 93-104, 2024, doi: 10.1117/12.524762

[Flyn16] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, "Overview of the Range Extensions for the HEVC Standard: Tools, Profiles, and Performance", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 4-19, 2016, doi: 10.1109/TCSVT.2015.2478707

[G271] G. Clare, Félix Henry, Joël Jung, "Sign Data Hiding", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G271 / MPEG M21833, Geneva, Switzerland, 2011

[G569] J. Lainema, K. Ugur, A. Hallapuro, "Single entropy coder for HEVC with a high throughput binarization mode", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G569, Geneva, Switzerland, 2011

[G582] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, "Multiview HEVC – experimental results", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, MPEG 2011 / M22147, Doc JCTVC-G582, Geneva, Switzerland, 2011

[Graj22] T. Grajek, **J. Stankowski**, "Rate-Distortion Optimized Quantization in Motion JPEG", 30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2022), 2022

[Hell12] P. Helle, S. Oudin, B. Bross, D. Marpe, M. O. Bici, K. Ugur, J. Jung, G. Clare, T. Wiegand, "Block Merging for Quadtree-Based Partitioning in HEVC", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1720-1731, 2012, doi: 10.1109/TCSVT.2012.2223051

[Ichi99] K. Ichino, T. Yoshida, I. Nishihara, Y. Sakai, "2D/3D hybrid video coding based on motion compensation", Proceedings of International Conference on Image Processing (ICIP 1999), vol.2, pp. 644-648, Kobe, Japan, 1999, doi: 10.1109/ICIP.1999.822974

[Isna07] M. A. Isnardi, “Historical Overview of Video Compression in Consumer Electronic Devices”, Digest of Technical Papers International Conference on Consumer Electronics, pp. 1-2, Las Vegas, NV, USA, 2007 doi: 10.1109/ICCE.2007.341449

[Kama05] N. Kamaci, Y. Altunbasak, R. M. Mersereau, <Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 8, pp. 994-1006, 2005, doi: 10.1109/TCSVT.2005.852400.

[Karw16] D. Karwowski, T. Grajek, K. Klimaszewski, O. Stankiewicz, **J. Stankowski**, K. Wegner, “20 Years of Progress in Video Compression - from MPEG-1 to MPEG-H HEVC. General View on the Path of Video Coding Development”, Image Processing and Communications Challenges 8, Advances in Intelligent Systems and Computing: No. 525, Springer International, pp. 3-15, 2016, doi: 10.1007/978-3-319-47274-4_1

[Kim12] I.-K. Kim, J. Min, T. Lee, W.-J. Han and J. Park, “Block Partitioning Structure in the HEVC Standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1697-1706, 2012, doi: 10.1109/TCSVT.2012.2223011

[L1100] F. Bossen, “Common test conditions and software reference configurations”, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCTVC-L1100, Geneva, Switzerland, 2013

[M17050] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, “Poznań Multiview Video Test Sequences and Camera Parameters”, ISO/IEC JTC1/SC29/WG11, MPEG/M17050, Xian, China, 2009

[M22697] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, **J. Stankowski**, K. Wegner, “Technical Description of Poznan University of Technology proposal for Call on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11, MPEG 2011 / M22697, Geneva, Switzerland, 2011

[M36569] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz, K. Wegner, “Video and depth multiview test sequences acquired with circular camera arrangement – Poznan Service and Poznan People”, ISO/IEC JTC1/SC29/WG11 MPEG2015, M36569, Warsaw, Poland, 2015

[M40010] D. Doyen,T. Langlois, B. Vandame, F. Babon, G. Boisson, N. Sabater, R. Gendrot, A. Schubert, “Light field content from 16-camera rig”. ISO/IEC JTC1/SC29/WG11 MPEG, M40010, Geneva, Switzerland, 2017

[Marp03] D. Marpe, H. Schwarz, T. Wiegand, “Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard” IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 620-636, 2003, doi: 10.1109/TCSVT.2003.815173

[N12036] International Organisation for Standardisation, “Call for Proposals on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, Geneva, Switzerland, March 2011

[N12347] K. Müller, A. Vetro, V. Baroncini, “Report of Subjective Test Results from the Call for Proposals on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11, MPEG2011/N12347, Geneva, Switzerland, 2011

[OPUS12] Narodowe Centrum Nauki, 2012/05/B/ST7/01279
Description available at: https://multimedia.edu.pl/?page=projects_OPUS3

[Oste04] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, T. Wedi, “Video coding with H.264/AVC: tools, performance, and complexity”, IEEE Circuits and Systems Magazine, vol. 4, no. 1, pp. 7-28, 2004 doi: 10.1109/MCAS.2004.1286980.

[Rich24] I. E. Richardson, “Coding Video: A Practical Guide to HEVC and Beyond”, Wiley, Chichester, UK, 2024.

[Same16] J. Samelak, **J. Stankowski**, M. Domański, “Adaptation of the 3D-HEVC coding tools to arbitrary locations of cameras”, International Conference on Signals and Electronic Systems, ICSES 2016, Kraków, Poland, 2016,

[Setk11] A. Setkov, “Matching quality investigation of Motion Estimation Algorithms”, Proceedings of 9th Conference of Open Innovations Association FRUCT, pp. 116–122, Petrozavodsk, Russia, 2011

[Shi08] Y. Q. Shi, H. Sun, “Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards (2nd ed.)”, CRC Press, Inc., USA, 2008

[Sjob12] R. Sjoberg, Y. Chen, A. Fujibayashi, H. Akiraksela, M. M. Hannuksela, J. Samuelsson, T. K. Tan, K. Thiow, Y.-K. Wang, S. Wenger, “Overview of HEVC High-Level Syntax and Reference Picture Management”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1858-1870, Dec. 2012, doi: 10.1109/TCSVT.2012.2223052

[Sole12] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, A. Duenas, “Transform Coefficient Coding in HEVC”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1765-1777, Dec. 2012, doi: 10.1109/TCSVT.2012.2223055

[Stan18a] O. Stankiewicz, K. Wegner, D. Karwowski, **J. Stankowski**, K. Klimaszewski, T. Grajek, “HEVC encoding assisted with noise reduction”, International Journal of Electronics and Telecommunications, Vol. 64, No. 3, pp. 285-292, Electronics and Telecommunications Committee of Polish Academy of Sciences, Warsaw, 2018

[Stan18b] O. Stankiewicz, G. Lafruit, M. Domański, “Multiview video: Acquisition, processing, compression and virtual view rendering”, Image and Video Processing and Analysis and Computer Vision, Academic Press Library in Signal Processing: Vol. 6, pp. 3-74, Academic Press, 2018, doi: 10.1016/B978-0-12-811889-4.00001-4,

[Sull12] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, 2012, doi: 10.1109/TCSVT.2012.2221191

[Sze12] V. Sze, M. Budagavi, “High Throughput CABAC Entropy Coding in HEVC”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1778-1791, 2012, doi: 10.1109/TCSVT.2012.2221526

[Sze14] V. Sze, M. Budagavi, G. J. Sullivan, Eds., “High Efficiency Video Coding (HEVC): Algorithms and Architectures”, Springer, Cham, Switzerland, 2014

[Tech16] G. Tech, Y. Chen, K. Müller, J. -R. Ohm, A. Vetro, Y.-K. Wang, “Overview of the Multiview and 3D Extensions of High Efficiency Video Coding”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, 2016, doi: 10.1109/TCSVT.2015.2477935

[Vann12] J. Vanne, M. Viitanen, T. D. Hamalainen, A. Hallapuro, “Comparative Rate-Distortion-Complexity Analysis of HEVC and AVC Video Codecs”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1885-1898, 2012, doi: 10.1109/TCSVT.2012.2223013

[Vetr03] A. Vetro, C. Christopoulos, H. Sun, “Video transcoding architectures and techniques: an overview”, IEEE Signal Processing Magazine, vol. 20, no. 2, pp. 18-29, 2003, doi: 10.1109/MSP.2003.1184336

[Vetr11] A. Vetro, T. Wiegand, G. J. Sullivan, “Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard”, Proceedings of the IEEE, vol. 99, no. 4, pp. 626-642, 2011, doi: 10.1109/JPROC.2010.2098830

[Wang04] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, “Image quality assessment: From error measurement to structural similarity,” IEEE Transactions of Image Processing, vol. 13, pp. 600–613, 2004

[Wieg03] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, “Overview of the H.264/AVC video coding standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, 2003, doi: 0.1109/TCSVT.2003.815165

[Xiu10] T. Xiu-li, H. Univ, D. Sheng-kui, “An analysis of TZSearch algorithm in JMVC”, International Conference on Green Circuits and Systems (ICGCS), pp. 516-520, Shanghai, 2010

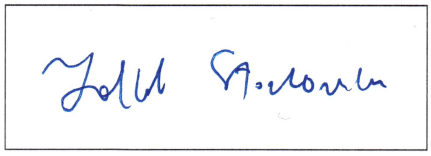
[Zhan13] L. Zhang, J. Kang, X. Zhao, Y. Chen, R. Joshi, “Neighboring block based disparity vector derivation for 3D-AVC”, 2013 Visual Communications and Image Processing (VCIP), pp. 1-6, Kuching, Malaysia, 2013, doi: 10.1109/VCIP.2013.6706401

Appendix 1

Declaration of contribution to the authorship of the publications

Author contribution statement

I hereby declare that the author contributions described in Section 0.2 of the dissertation ("Analysis, extensions, and optimizations of high-efficiency video compression techniques") accurately reflects the facts.



Jakub Stankowski

Appendix 2

Copies of publications mentioned in 0.2

Analysis of frame partitioning in HEVC *

Jakub Stankowski, Tomasz Grajek, Damian Karwowski, Krzysztof Klimaszewski, Olgierd Stankiewicz, Krzysztof Wegner, and Marek Domański

Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology, ul. Polanka 3, 60-965 Poznań, Poland
{jstankowski,tgrajek,dkarwow,kklima,ostank,kwegner}@multimedia.edu.pl

Abstract. The paper presents analysis of frame partitioning in the next generation video coding standard HEVC. Complex study on frequency of the particular coding unit, prediction unit and transform unit sizes selection made by the encoder compliant with HEVC technology is included. General conclusions based on extensive experiments for HD video sequences are presented. Such knowledge may be the first step to the development of more efficient mode selection algorithms.

1 Introduction

In the last years the growing importance of multimedia systems transmitting high quality videos is observed. In this context, studies on efficient representation of high resolution images are crucial. Such works have been conducted in last years, and as a result the new High Efficiency Video Coding (HEVC) technology has been developed. The technology has been standardized as ISO/IEC 23008-2 (MPEG-H part 2) and ITU-T H.265 in 2013 [1].

HEVC allows for encoding of high resolution video at half of bitrate of previous technology namely AVC (MPEG-4 Part 10 and H.264)[2]. Higher coding efficiency came at a price of higher encoder complexity.

General idea of the HEVC technology is very similar to the older generation solutions (e.g. MPEG-2, H.263, AVS, VC-1, AVC) [3] and exploits intra- and inter-frame correlation by using prediction coding of image blocks together with block-based transform coding of residual data. But the HEVC allows for greater flexibility in terms of image partitioning and prediction mode selection leading to a stronger compression of high resolution video [4]. Classical macroblock structure has been replaced by more flexible coding tree block structure (CTB). Each CTB allows to quad tree block partitioning from size of 64x64 pixels downto 8x8 pixels. Each block in a partitioning tree is called a coding unit (CU) and can be encoded in intra-frame or inter-frame mode. Within the CU, prediction is performed in prediction units (PU) of various size (both square and rectangular shapes of PU are available). Each CU can also be further recursively divided into transform units (TU) of various sizes from 32x32 downto 4x4 in

* The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-11331-9_72

which transform-based coding of residual data is performed. Size of PU and TU can not exceed the size of CU.

CTB partitioning with PU and TU partitioning leads to almost 3000 combinations in case of intra-frames and more than 4500 combinations in case of inter-frames to be analysed by encoder for a single CTB. In other words, great number of combinations is available in the encoder and all of them should be checked in order to choose the optimal one. However, huge computational complexity is the main drawback of such approach. Therefore, not all of combinations are evaluated in practice during encoding. Various methods of speeding up encoder decision have been developed so far, for example [5–8]. Basically all of them are based on observation that some partitioning schemes are very rarely used by the encoder, so they can be omitted resulting in negligible coding efficiency loss.

Nevertheless there are no complex studies of frequency of the particular partitioning selection made by the encoder. Such a knowledge can be further used to develop even more efficient encoder mode selection algorithms. Thorough analysis of image partitioning makes the topic of the paper.

2 Research problem and goal of the work

In this paper we have evaluated frequency of particular CU, PU and TU sizes selection made by the encoder. In particular we are interested in the knowledge of image area covered/encoded with particular CU, PU and TU sizes. The goal is to make detailed analysis for individual frames types used in the HEVC encoder.

3 Methodology of the research

In extensive experiments, seven HD 1920x1080 test video sequences (Fig. 1): *Bluesky*, *Pedestrianarea*, *Riverbed*, *Rushhour*, *Station2*, *Sunflower* and *Tractor*, recommended by ISO/IEC MPEG and ITU-T VCEG working groups have been encoded by the HEVC encoder. In particular HM 10.0 reference software [9] working under Common Test Condition (CTC) [10] was used. CTC defines a set of conditions and encoder configurations designed as a common ground for evaluation of HEVC related technology. In experiments "Random Access" scenario has been used to make the research for high efficiency mode (Fig. 2). Each sequence was encoded multiple times, each time with different quantization parameter (QP) value from a range $10 \div 48$ resulting in a wide range of bitrates. It must be emphasized that, the "Random Access" scenario assumes different values of QP for individual types of frames in the hierarchical coding structure.

4 Experimental results

Coding unit analysis. QP value affects the statistics of CU sizes in frames. Experimental results averaged over all sequences and all frames allow to draw the following general conclusions:



Fig. 1. Test video sequences used in experiments in order from top-left: *Bluesky*, *Pedestrianarea*, *Riverbed*, *Rushhour*, *Station2*, *Sunflower* and *Tractor*.

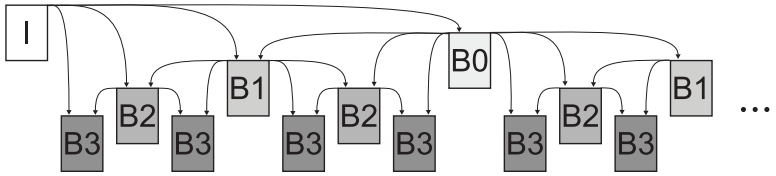


Fig. 2. Hierarchical structure of frame prediction used in experiments (according to CTC). Only part of inter-prediction sources is marked.

- In the case of small QPs, the large CU (i.e. 64x64) is rarely used (below 5%). In such a case the most commonly used sizes of CU are 32x32 and 16x16 (about 75%).
- With the increase of QP the amount of big CUs also increases (up to 90%). In particular, for really high QPs 16x16 CUs and smaller ones are rarely used by encoder (below 10%).

Going into details there is a large convergence of the results obtained for $B0$, $B1$, $B2$, and $B3$ frames, but these results are significantly different from those obtained for I frames. In particular, significantly higher percentage contribution of 8x8 CUs is observed in I frames in the case of small QPs. Why is this happening? In the HEVC encoder the size of CU determines the size of PUs in which image prediction is realized. Selected size of CU is the upper limit for the size of PUs and TUs. In the case of complex image textures, CU is divided into smaller blocks in order to increase encoding efficiency. Thus, the increased use of smaller CUs for small QP values, where details of textures are preserved (for intra- and inter-frames). In addition, the prediction of complex image parts is more demanding in the case of intra-frames, hence more frequent (than in inter-frames) selection of small CUs. Detailed data for these experiments are shown in Fig. 3 for a wide range of QP values. The most likely explanation for this is the following. By strong data quantization (high QP values) the vast majority of transform coefficients take zero value. Therefore, it is more reasonable to carry out encoding in smaller CUs in order to preserve from sending large amount of control data. In this case, large CUs are more often applied.

Prediction unit analysis. There are two categories of PUs: square and rectangular. Overall square PUs are chosen more frequently than rectangular ones (see Fig. 4). With increasing QP, frequency of choosing small and rectangular PUs decreases. Interesting observation is that rectangular PUs of size 64xN and Nx64 are chosen more frequently for average values of QP than for extreme ones. On average PUs of size 4x4 are hardly ever selected for QP greater than 24.

The distribution of PUs for I frames and B frames is quite different. Mainly due to lack of rectangular PUs in I frames (which results from HEVC standard). Moreover small PUs are chosen 5 times more frequently in I frames than in B frames. For I frames it is harder to accurately predict content in larger PUs so encoder chooses smaller units. 4x4 PUs are not used for QPs greater than 18 in B frames and for QP greater than 36 in I frames which is a huge difference. Distribution of frequency of various sizes PUs in $B0$, $B1$, $B2$, and $B3$ frames is quite similar. The only difference is a frequency of choosing 4x4 PUs. Results in $B0$ frames and comparable to frequency in I frames. But having in mind the number of $B0$ frames, it has no significant impact on overall statistics for all B frames.

Transform unit analysis. For each CU, encoder decides how to divide it into TU blocks (i.e. TU block size equal to the CU size or CU divided into smaller TUs). Therefore there is a strong relationship between statistics of the TU and CU size. This is confirmed by the results of experiments (see Fig. 5):

- With the increase of the QP the share of large TUs (i.e 32x32 and 16x16) also increases. In particular, due to strong quantization (high QP values) most of transform coefficients are equal to zero. Therefore, it is better to carry out encoding in larger blocks (less control data which means higher coding efficiency).
- In the case of small values of QP, the smaller sized TUs are more frequently used. It is related to CU statistics discussed earlier.

In the HEVC encoder a large area of the frame is covered by the TU blocks for which prediction error signal is not sent to decoder (all quantized transform coefficients are equal to zero). This is particularly evident for B frames for which higher efficiency of predicting coding may be observed. Although detailed results differ between the $B0$, $B1$, $B2$, and $B3$ frames, in each of these frames increase of QP value increases the number of such blocks (up to 50% of the frame). In contrary, for I frames share of such TUs is very low (below 5% due to smaller efficiency of intra prediction).

5 Conclusions and final remarks

The results documented in the paper give useful information about the statistics of selection of CUs, PUs and TUs of different sizes by the HEVC encoder. In order to draw detailed conclusions an independent analysis was done for individual

frame types (I, $B0$, $B1$, $B2$, and $B3$). Additionally all the results were averaged and presented in the paper to highlight the general remarks. Huge collection of experimental results enables analysis of frame partitioning for a wide range of bitrates (QP values from 10 to 48). Achieved results allow to formulate two general conclusions:

1) The higher QP the major area of the frame is covered by larger CUs, PUs and TUs (up to 80%, 83% and 87% respectively).

2) There is a significant difference between results obtained for I and B frames. For the same QP value for I frames smaller sizes of CUs, PUs and TUs than for B frames are used.

The results are the basis for further study leading to the development of fast mode decisions in the HEVC encoder. The knowledge on statistics of frame partitioning acquired in the paper gives direct information which combinations of CU, PU and TU can be omitted in order to speed up the encoder mode selection process. Appropriate remarks have been presented in experimental section.

Acknowledgement Project was supported by The National Centre for Research and Development, Poland. Grant no. LIDER/023/541/L-4/12/NCBR/2013.

References

1. ISO/IEC 23008-2 (MPEG-H Part 2) / ITU-T Rec. H.265: High Efficiency Video Coding (HEVC) (2013)
2. Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T.-K., Wiegand, T.: Comparison of the Coding Efficiency of Video Coding Standards Including High Efficiency Video Coding (HEVC). IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1669–1684 (2012)
3. Shi, Y.Q., Sun, H.: Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards. Second Edition, ISBN 978-0-8493-7364-0, CRC Press, New York (2008)
4. Sullivan, G.J., Ohm, J.-R., Han, W.-J., Wiegand, T.: Overview of the High Efficiency Video Coding (HEVC) Standard, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668 (2012)
5. Zhao, L., Zhang, L., Ma, S., Zhao, D.: Fast mode decision algorithm for intra prediction in HEVC. IEEE Visual Communications and Image Processing (VCIP), Tainan, Taiwan, pp. 1–4 (2011)
6. Palomino, D., Cavichioli, E., Susin, A., Agostini, L., Shafique, M., Henkel, J.: Fast HEVC Intra Mode Decision Algorithm Based on New Evaluation Order in the Coding Tree Block. Picture Coding Symposium PCS 2013, San Jose, USA, pp. 209–212 (2013)
7. Xiong, J., Li, H., Wu, Q., Meng, F.: A Fast HEVC Inter CU Selection Method Based on Pyramid Motion Divergence. IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 559–564 (2014)
8. Yoo, H.M., Suh, J.W.: Fast coding unit decision algorithm based on inter and intra prediction unit termination for HEVC. 2013 IEEE International Conference on Consumer Electronics (ICCE),Las Vegas, USA, pp. 300–301 (2013)
9. HEVC test model reference software - available online: https://hevc.hhi.fraunhofer.de/svn/

10. Bossen, F.: Common test conditions and software reference configurations. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, doc. JCTVC-J1100, Stockholm, Sweden (2012)

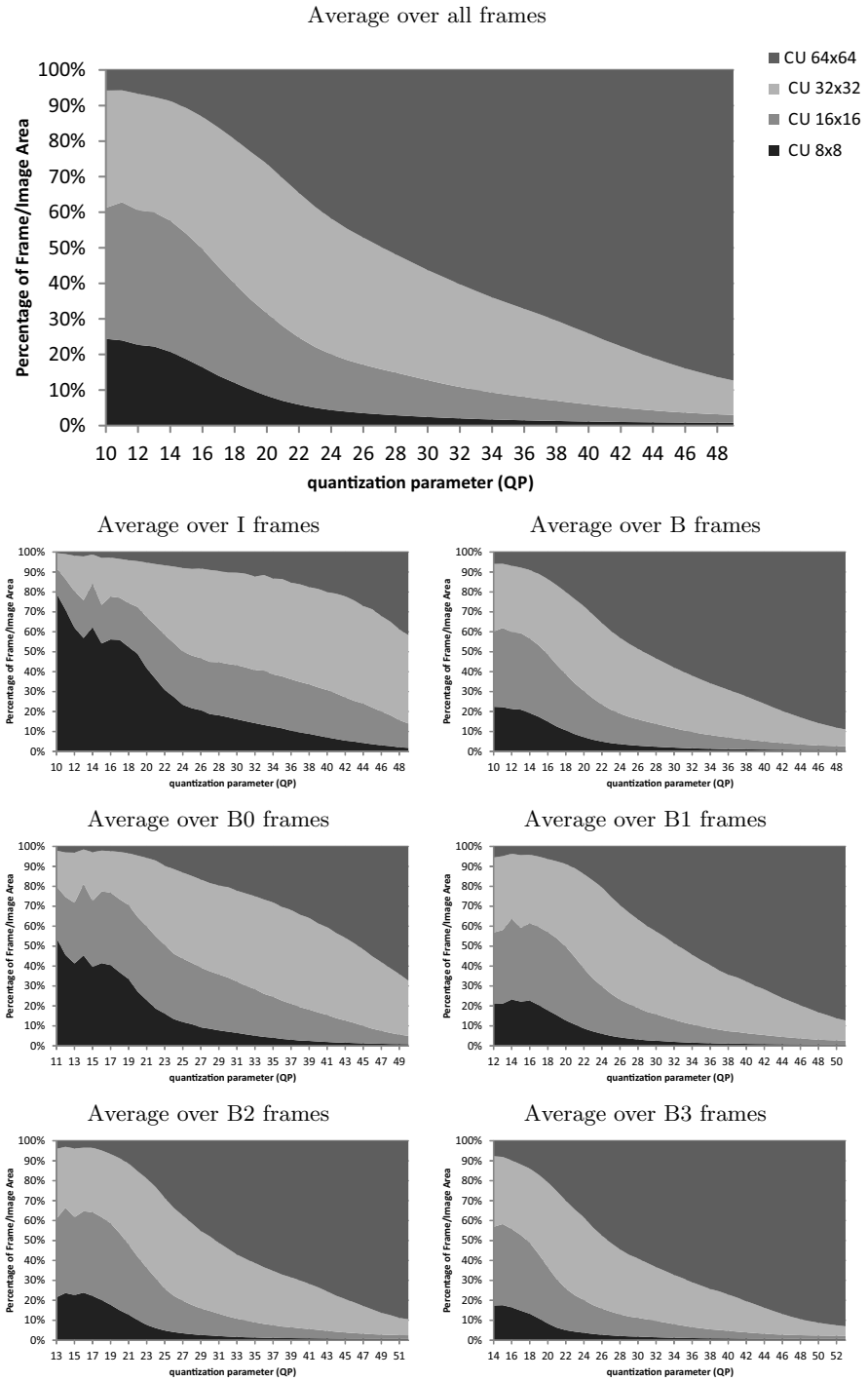


Fig. 3. Percentage of frame/image area covered by different sizes of CU. The QP values for individual types of frames in the hierarchical coding structure are adjusted according to "Random Access" scenario.

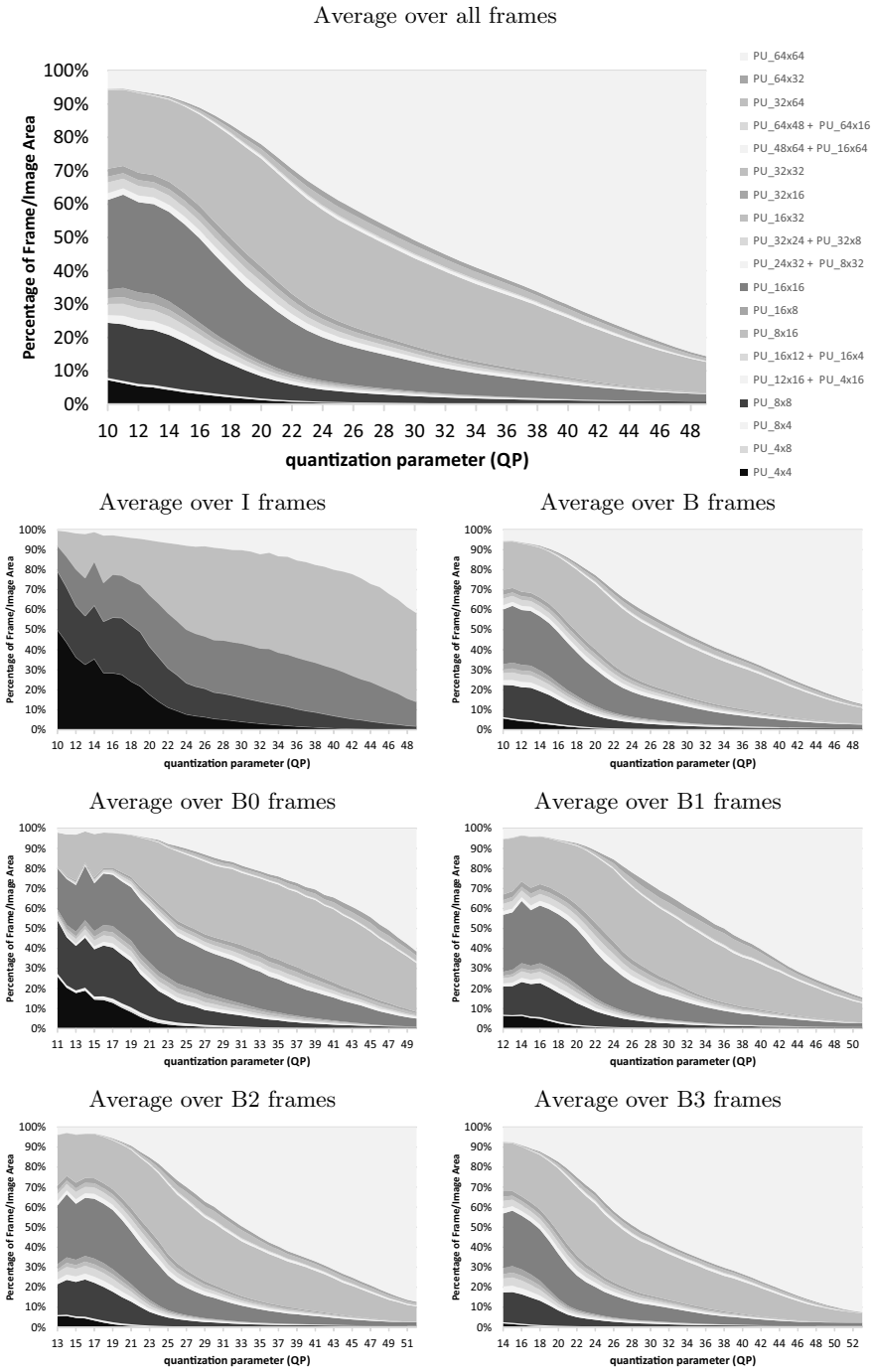


Fig. 4. Percentage of frame/image area covered by different sizes of PU. The QP values for individual types of frames in the hierarchical coding structure are adjusted according to "Random Access" scenario.

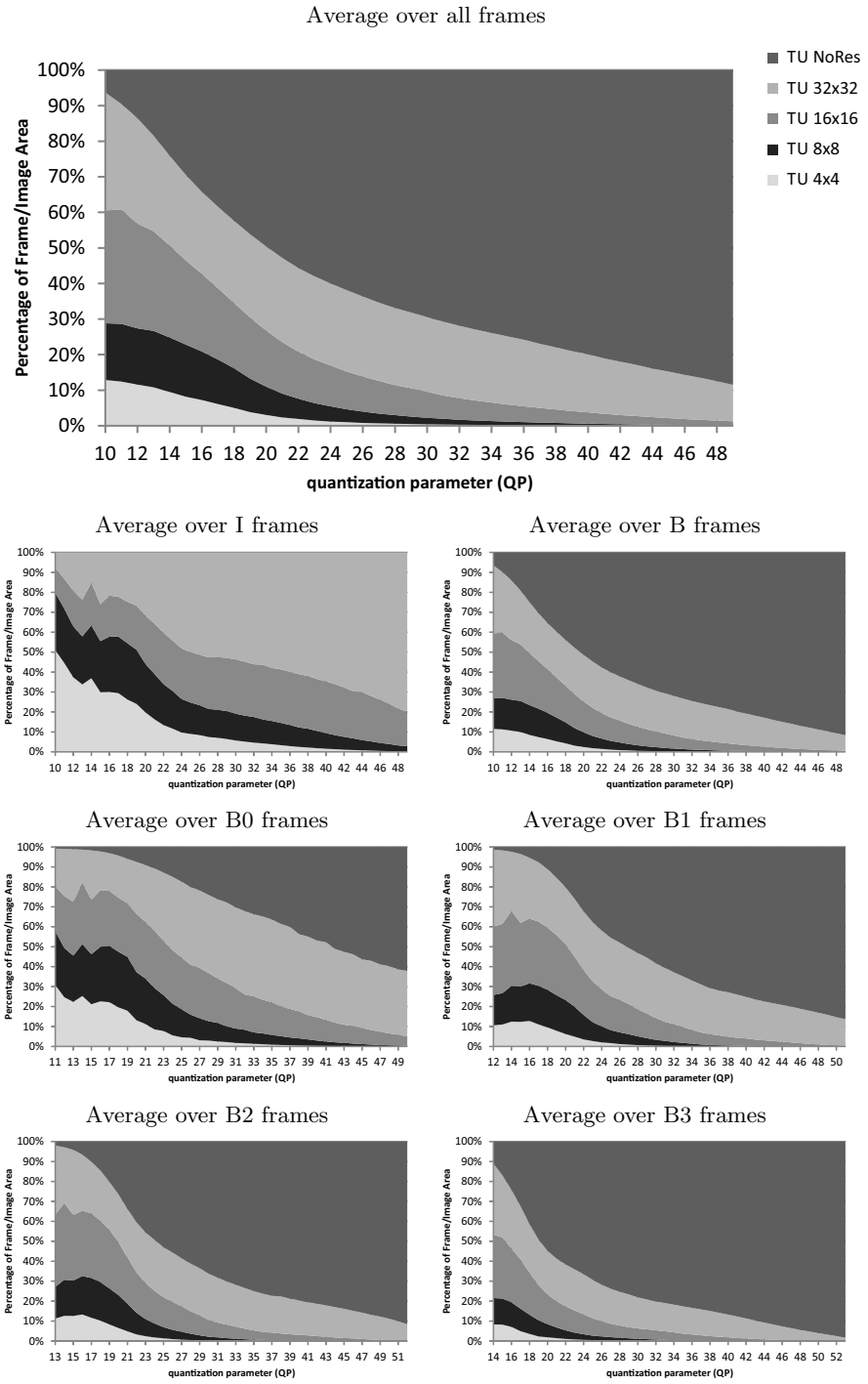


Fig. 5. Percentage of frame/image area covered by different sizes of TU. The QP values for individual types of frames in the hierarchical coding structure are adjusted according to "Random Access" scenario.

Analysis of Compressed Data Stream Content in HEVC Video Encoder

Jakub Stankowski, Damian Karwowski, Tomasz Grajek, Krzysztof Wegner, Jakub Siast, Krzysztof Klimaszewski, Olgierd Stankiewicz, Marek Domański,

***Abstract*—In this paper, a detailed analysis of the content of the bitstream, produced by the HEVC video encoder is presented. With the use of the HM 10.0 reference software the following statistics were investigated: 1) the amount of data in the encoded stream related to individual frame types, 2) the relationship between the value of the QP and the size of the bitstream at the output of the encoder, 3) contribution of individual types of data to I and B frames. The above mentioned aspects have been thoroughly explored for a wide range of target bitrates. The obtained results became the basis for highlighting guidelines that allow for efficient bitrate control in the HEVC encoder.**

***Keywords*—HEVC; video compression; compressed data stream analysis**

I. Introduction

A huge amount of data related to original video signal became an incentive to work out techniques of video compression. Over the years a number of video compression methods has been developed, many of them have been presented in the literature [1], [2]. Among the known solutions, the hybrid coding schemes are of particular importance, due to their high compression performance and consequently, many applications on the market [1], [2]. The state-of-the-art in the field of hybrid video compression is the new High Efficiency Video Coding (HEVC) technique [3], developed jointly by ISO/IEC MPEG and ITU-T VCEG experts, and standardized in 2013 [4]. Thanks to many improvements as well as new coding tools [3], HEVC offers up to 50% bitrate reduction in comparison to the AVC, while maintaining the same quality of encoded video [5]. For this reason, in the coming years, HEVC is envisaged to be the subject of industrial deployments in systems with high definition video signals.

The deployment of the new HEVC technology needs to be preceded by the works on efficient mechanisms of bitrate control e.g. [6]. Some of the known encoder control methods utilize the information about the content of the encoded data stream that is produced by the encoder [7]. The results of the analysis of data stream produced by the encoders of older generations (e.g. MPEG-2, AVC) are available in [8]. In the HEVC, the content of the encoded data stream is different, when compared to MPEG-2 or AVC (due to the new solutions in the structure of the encoder). First of all, HEVC encoder produces different syntax elements. The quantitative features of the stream (percentage distribution of individual types of data in the stream) are also different. As a result, research on the content of the encoded data stream must be repeated in order to work out the new encoder control solutions, dedicated to the HEVC technique. Thorough analysis of the HEVC bitstream content, and its properties, make the topic of the paper. In particular, the aim of this work is the analysis of such a HEVC data stream, which is the result of the encoder mode selection mechanism that ensures high efficiency video coding (see methodology of experiments). In this encoding scenario, the relationship between the bitrate and the quality of decoded pictures has already been the subject of thorough study on HEVC [5]. For this reason, the quality of the decoded pictures was not a subject of research in this paper.

The rest of the paper is organized as follows. In Section II detailed goals of the work have been presented. Section III reveals what experiments have been carried out and what was the methodology of these studies. Section IV presents experimental results together with comments and general conclusions. And the last, Section V summarizes the paper.

Jakub Stankowski, Damian Karwowski, Tomasz Grajek, Krzysztof Wegner, Jakub Siast, Krzysztof Klimaszewski, Olgierd Stankiewicz, Marek Domański are with the Chair of Multimedia Telecommunications and Microelectronics, Poznan University of Technology, Poznan, Poland (e-mail: jstankowski;dkarwow;tgrajek;kwegner;jsiast;kklima;ostank@multimedia.edu.pl)

II. Research problem and goal of the work

Detailed knowledge about the content of the encoded video data stream is the basis for efficient control of bitrate in the encoder. Such knowledge will allow for more efficient distribution of the bit budget between individual frames of a video sequence. Although, some studies have been performed in this area earlier, they present preliminary results only. For example, separate and detailed results for I and B frames have not been presented in the literature so far (only averaged results were presented in [9]). The analysis has been performed for a specific (and narrow) range of quantization parameter (QP) values. Or, the results were averaged without presenting detailed values for individual syntax elements, that have been defined in HEVC [10]. The detailed analysis of the HEVC data stream content, performed for different scenarios of video encoding is the goal of this paper. Preliminary research has already been done by the authors and published in [11]. In this paper the continuation of the cited work is presented together with new results which are supplementary to the data already published in [11]. In particular, the following problems are explored in this paper:

1) What is the portion of the total bitstream used to encode I frames? What is that portion for B frames?
2) What amount of the encoded data stream is related to individual frame types?

3) What is the share of individual categories of syntax elements (residual data, motion data, control data) in encoded data stream? How does the value of bitrate affect the results?
4) What are the statistics of syntax elements in frames of a given type (I and B frames)?
5) What portion of image area can be encoded without the need of transmitting residual data for transform coefficients?

III. Methodology of the Research

In order to perform a thorough analysis of the HEVC encoded data stream and draw reliable conclusions, extensive experiments have been done. In particular, in experiments the commonly available HM 10.0 reference software of the HEVC [12] working under Common Test Condition (CTC) [13] has been used. CTC defines a set of conditions and encoder configurations designed as a common ground for the evaluation of HEVC related technology. Using this software, a series of experiments was done with the following encoding conditions:

- Seven 1920x1080@30Hz test video sequences were used (Fig. 1): *bluesky*, *pedestrianarea*, *riverbed*, *rushhour*, *station2*, *sunflower*, *tractor*. The sequences are recommended by ISO/IEC MPEG as well as ITU-T VCEG working groups as appropriate for the purposes of research works on video compression.
- Random access encoding scenario was used, as defined in CTC [13]. In particular, a video sequence was divided into Groups Of Pictures (GOP), with a strict division of a GOP into I-, B-frame types and hierarchical encoding of B-frame types (the so-called B_0, B_1, B_2 and B_3 frames). The higher value of subscript of B frame the lowest level in the hierarchical structure (see Fig. 2 for details).
- Experiments were carried out for a wide range of target bitrates (controlled by the quantization parameter QP in encoder). Tests were done for QP values from 10 to 40 with step equal to 1.
- Other settings of the encoder were set as suggested in CTC. Essential configuration parameters used for HEVC are presented in Table I.

TABLE I: Essential Configuration Parameters Used for HEVC Encoder

Parameter	Parameter Value
Profile	Main
GOP structure	$IB_3B_2B_3B_1B_3B_2B_3B_0$ ...
GOP size	8
Intra period	32
Hierarchical GOP	Yes
Number of referenced frames	4
Rate-distortion optimization	On
Search range for motion estimation	± 64

As a result, 280 encoded data streams were achieved. Each of compressed video bitstream has been examined by using the Bitstream Analyzer software. The analyzer has been written in order to perform precise analysis of bitstream components. The analyzer entropy decodes the input bitstream and simulates the encoding process in order to calculate the number of bits per syntax element. The bits calculation is performed by the use of the CABAC counter precise estimator [14] instead of CABAC arithmetic encoder core. The CABAC counter allows for precise calculation of bits per syntax element (2^{-15} precision) and allows to avoid the arithmetic encoder flushing issue.



Fig. 1: Test video sequences used in experiments in order from top-left: *Bluesky*, *Pedestrianarea*, *Riverbed*, *Rushhour*, *Station2*, *Sunflower* and *Tractor*.

All results have been calculated for each of QP point with constant QP encoded bitstream and interpolated to common bitrate values.

The list of analyzed syntax elements in each syntax element category includes:

- ControlData:
  - *split_cu_flag,*
  - *cu_skip_flag,*
  - *pred_mode_flag,*
  - *part_mode,*
- IntraPred:
  - *prev_intra_luma_pred_flag,*
  - *mpm_idx,*
  - *rem_intra_luma_pred_mode,*
  - *intra_chroma_pred_mode,*
- MotionData:
  - *merge_idx,*
  - *merge_flag,*
  - *inter_pred_idc,*
  - *ref_idx_l0,*
  - *mvp_l0_flag,*
  - *ref_idx_l1,*
  - *mvp_l1_flag,*
  - *abs_mvd_greater0_flag,*
  - *abs_mvd_greater1_flag,*
  - *abs_mvd_minus2,*
  - *mvd_sign_flag,*

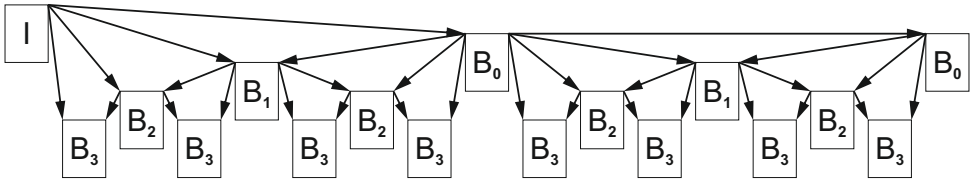


Fig. 2: Hierarchical coding scheme. For the sake of simplicity, only part of prediction directions are marked.

- ResidualData:
  - *rqt_root_cbf,*
  - *split_transform_flag,*
  - *cbf_cb,*
  - *cbf_cr,*
  - *cbf_luma,*
  - *transform_skip_flag,*
  - *last_sig_coeff_x_prefix,*
  - *last_sig_coeff_y_prefix,*
  - *last_sig_coeff_x_suffix,*
  - *last_sig_coeff_y_suffix,*
  - *coded_sub_block_flag,*
  - *sig_coeff_flag,*
  - *coeff_abs_level_greater1_flag,*
  - *coeff_abs_level_greater2_flag,*
  - *coeff_sign_flag,*
  - *coeff_abs_level_remaining,*
- Other:
  - *end_of_slice_segment_flag,*
  - *sao_merge_left_flag,*
  - *sao_merge_up_flag,*
  - *sao_type_idx_luma,*
  - *sao_type_idx_chroma,*
  - *sao_offset_abs,*
  - *sao_offset_sign,*
  - *sao_band_position,*
  - *sao_eo_class_luma,*
  - *sao_eo_class_chroma,*

Further in the paper, the numbers of bits that correspond to the above mentioned syntax element categories in pictures of individual types, i.e. I, B_0, B_1, B_2 and B_3 are studied.

IV. Experimental results

In the encoder, the bitrate control mechanism relies on the adjustment of the quantization step size. In the HEVC encoder the choice is made at frame, slice and Coding Unit (CU) levels [3]. Quantization step size is determined in the HEVC by defining the value of quantization parameter (QP). The current value of QP is sent in the encoded data stream. This parameter influences quantization of transform coefficients and thus has a direct impact on the bitrate of encoded data stream that is produced by the encoder. Therefore, the relationship between the value of QP and the bitrate of the encoded data stream has been studied in the first phase of the experiment.

A. Bitrate vs. Quantization Parameter

The results shown in Fig. 3 present the relationship between the value of QP and the bitrate for a set of test video sequences. As can be seen it depends significantly on the content of a video sequence. Thus, the analysis of the complexity of image content is an essential step of the algorithm that determines the value of QP for a given bitrate.

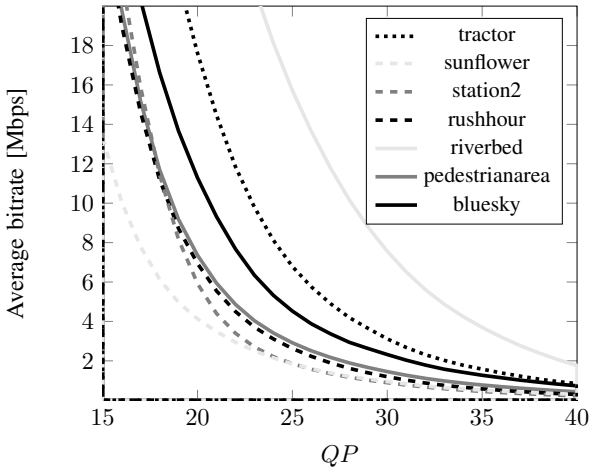


Fig. 3: Experimental curves for all test sequences. Relationship between QP and bitrate.

B. Contribution of individual frame types

In practice, the Group of Pictures (GOP) structure contains both the I and B types of frames. Therefore, there is a need for bitrate allocation inside the GOP between images of different types. During the experiments, the structure of GOP that is recommended by JCT-VC (Joint Collaborative Team on Video Coding of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11) experts group as appropriate for encoding scenarios with high compression performance [13] has been used. The distribution of bits between the images of different types for assumed GOP structure is shown in Fig. 4, 5, 6 and 7.

For the commonly used QP values in the range of 15–40, the results can be summarized as following.

Although there is only one I frame in the GOP (the whole GOP contains 32 frames), on average even 10–30% bits are devoted to encode this frame of type I (see Fig. 4 and 5,). Such a high percentage of bits for the I frames is a result of a much lower efficiency of intra-frame encoding as compared to inter-frame compression in the case of B frames. The averaged results revealed that, the contribution of the bits required by I frames ($bits_I$) to the total number of bits representing entire

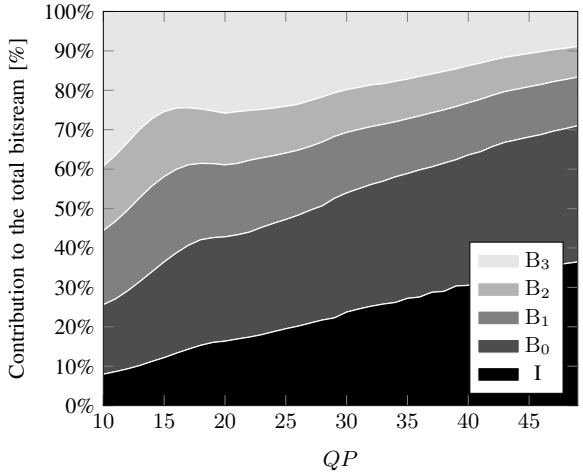


Fig. 4: Contributions of individual frame types to the total bitstream, averaged over 7 test sequences in relation to QP value.

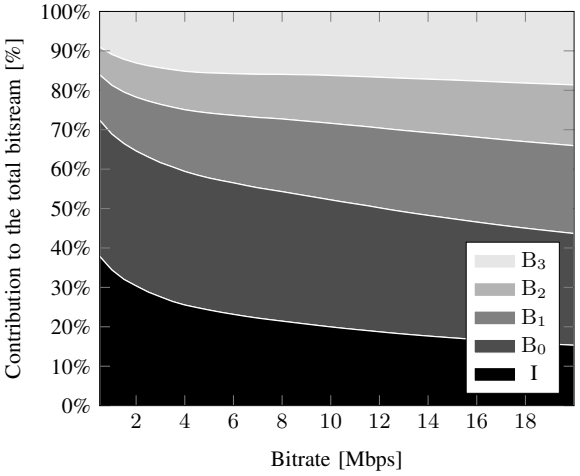


Fig. 6: Contributions of individual frame types to the total bitrate, averaged over 7 test sequences in relation to sequence bitrate.

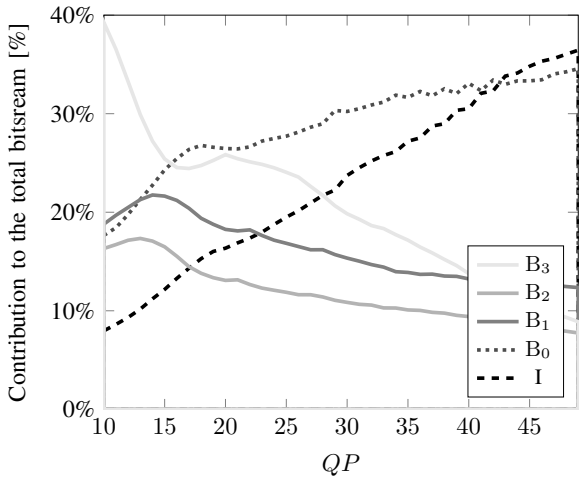


Fig. 5: Contributions of individual frame types to the total bitstream, averaged over 7 test sequences in relation to QP value.

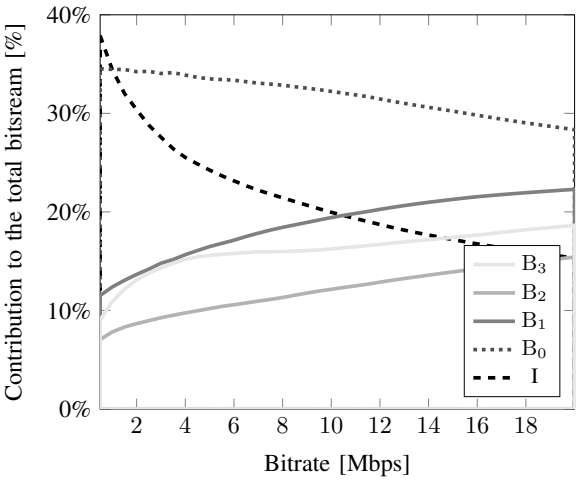


Fig. 7: Contributions of individual frame types to the total bitrate, averaged over 7 test sequences in relation to sequence bitrate.

sequence ($bits_{total}$) can be modeled as:

$$\frac{bits_I}{bits_{total}} \approx 0.0073 \cdot QP + 0.0141 \quad (1)$$

with the coefficient of determination (in statistical programs often marked as R^2) [15] equal to 0.9973. Such high value of coefficient of determination means that this approximation is almost perfect.

For B_0 frame types (the highest level of B frames in the hierarchical structure of GOP), the comparable number of bits as for I frames is used: 15-40% of the total bitstream. In this way, the I and B_0 frame types constitute 60-70% of the encoded data stream (for the lowest tested QP it is about 25%).

The rest of bitstream conveys data for B_1, B_2 and B_3 frames, specifically: 20% for B_1 frames, about 15% for B_2 frames, and depending on the value of QP, between 10-40% for B_3 frames.

The presented quantitative results of the research can be the basis of the proper allocation of the bits budget between the individual images in the GOP. In order to increase the efficiency of the bitrate control algorithm, the nature of image content (image complexity) should additionally be taken into account when distributing bits among images.

From the point of view of bit allocation between images, it is more desirable to know distribution of bits between images, but not in the function of QP, but in the function of bitrate of encoded data instead. The results of the study, which was carried out for a wide range of bitrates (500 kbps to 20 Mbps) are shown in Fig. 6 and 7. As can be seen, reducing the

bitrate results in monotonic increase of the amount of data associated with I and B_0 frames. It is obvious that smaller bitrates correspond to stronger quantization of residual data. Due to the fact, that the prediction of image samples works more efficiently in the case of B_1, B_2 and B_3 frames (when compared to I and B_0 frames), a stronger quantization reduces more data in B_1, B_2 and B_3 frames than in other types of images. It has been experimentally confirmed in our additional study described in Section IV-D.

For applications in high-quality digital television (including Full HD images) and the HEVC technology, the useful range of bitrates is 4–5 Mbps. In this particular application, the I frames constitute about 25% of the total bitstream, B_0 frames 34%, B_1 frames about 16%, B_2 frames 10%, and B_3 frames 15% of the total bitstream.

When presenting the results of Fig. 6 in the form as in Fig. 7 it is possible to propose mathematical models, that allow approximation of the share of individual types of frames in encoded data stream (as a function of bitrate). On average, the number of bits required by individual types of frames ($bits_{type-of-frame}$) to the total number of bits representing entire sequence ($bits_{total}$) in relation to bitrate in Mbps can be approximated as follows:

1) Data for I frames may be approximated by function:

$$\frac{bits_I}{bits_{total}} \approx -0.0438 \cdot log_2(Bitrate) + 0.3437 \quad (2)$$

with the coefficient of determination equal to 0.9988. Such high value of coefficient of determination means that this approximation is almost perfect.

2) In the case of B_0 frames the linear function is sufficient:

$$\frac{bits_{B_0}}{bits_{total}} \approx -0.0033 \cdot Bitrate + 0.0352 \quad (3)$$

with the coefficient of determination equal to 0.9862.

3) For B_1 frames, high accuracy of approximation is provided by power function:

$$\frac{bits_{B_1}}{bits_{total}} \approx 0.1214 \cdot Bitrate^{0.2028} \quad (4)$$

with the coefficient of determination equal to 0.9875.

4) In the case of B_2 frames data may be approximated by the linear function:

$$\frac{bits_{B_2}}{bits_{total}} \approx 0.0037 \cdot Bitrate + 0.0825 \quad (5)$$

with the coefficient of determination equal to 0.994.

5) In the case of B_3 frames data may be approximated by the linear function:

$$\frac{bits_{B_3}}{bits_{total}} \approx 0.0159 \cdot log_2(Bitrate) + 0.1134 \quad (6)$$

with the coefficient of determination equal to 0.97.

Adjustment of the coefficients values of the above mentioned functions is beyond the scope of this work and should be the subject of further research.

C. Contribution of categories of syntax elements

On the basis of detailed results that were collected during the experiments (distribution of individual syntax elements in encoded data stream - see methodology of experiments), the data that present distribution of five syntax element categories were prepared:

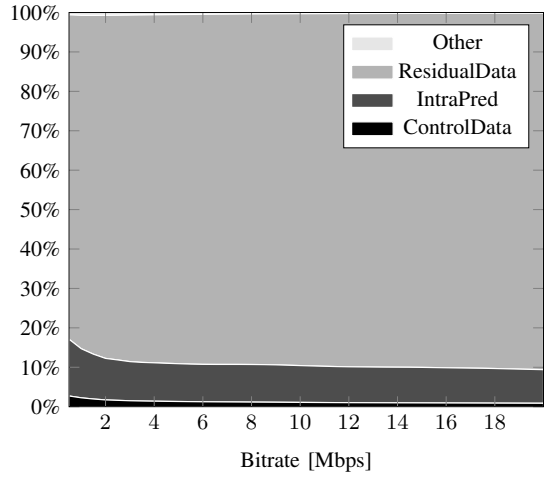
1) residual data of picture samples (ResidualData),
2) motion data (MotionData),
3) control data (ControlData),
4) information about intra prediction mode (IntraPred),
5) other types of data (Other).

Detailed statistics of these categories of syntax elements are different in individual types of frames - Fig. 8. In the case of I frames, there is no motion data (see Fig. 8a). It results in a greater share of data related to residual data of picture samples - which is 85-90% of the total bitstream. The remainder of the stream is intra prediction signalization and control data. There is a small variability of the results in function of bitrate. The value of bitrate affects the bitstream content of B frames to a greater extent (see Fig. 8b). In the case of residual data, they constitute 50-80% of the bitstream, for the low and high bitrates respectively. The motion data takes up 10-30% of bitstream, with larger share for smaller bitrates. The remaining types of data make 5-15% of the bitstream. The detailed results for individual types of B frames are presented in Fig. 8c, 8d, 8e and 8f. As can be seen, the results for B_1 and B_2 types of frames are very similar, but different than those obtained for B_0 and B_3 frames. Smaller contribution of residual data for the lower bitrates, and thereby higher share of other types of data in this scenario is a direct result of a stronger quantization of residual data.

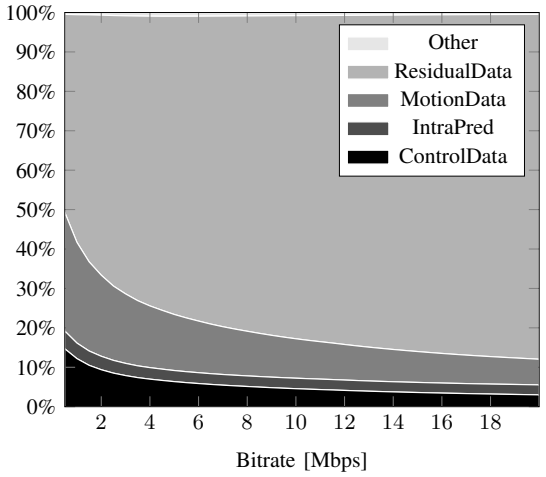
It should be noted that the size of the HEVC data stream is twice lower than that at the output of the AVC encoder. However, due to the use of more efficient techniques of image prediction and motion data coding, the proportions of individual types of data in encoded stream of data have not substantially changed. Residual data of image prediction is still an essential part of the encoded data stream. Therefore, the selection of QP value in the HEVC encoder will be the basic tool of the bitrate control, similarly as in the case of the older video compression techniques e.g. [16], [17] for MPEG-2 and [18], [19] for AVC.

D. Quantized transform coefficients

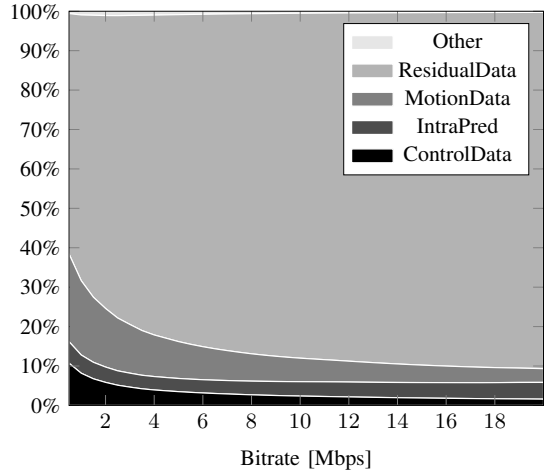
As mentioned before, the prediction of image samples works more efficiently in the case of B_1, B_2 and B_3 frames. Moreover, a stronger quantization reduces more data in B_1, B_2 and B_3 frames than in other types of frames. In order to confirm this, additional study has been performed. The percentage of image blocks with no residual signal of image samples prediction was measured for I, B_0, B_1, B_2 and B_3 frames (see Fig. 9). Due to limited efficiency of intra-frame prediction, for only 3% of image area, there are no non-zero transform coefficients of prediction error samples. Moreover, this percentage is quite independent from the value of bitrate. Much higher efficiency of predictive coding has been observed in B frames. For broadcast applications a useful range of



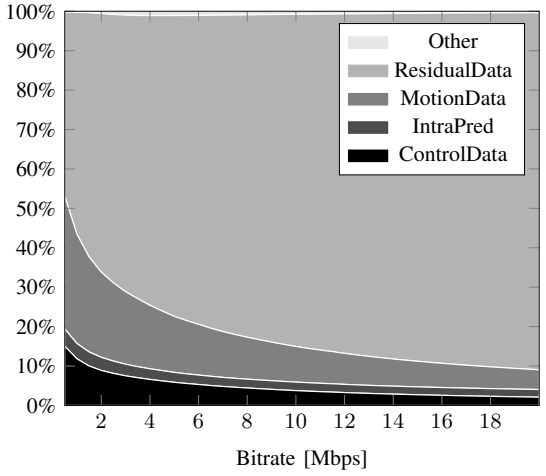
(a) I frames



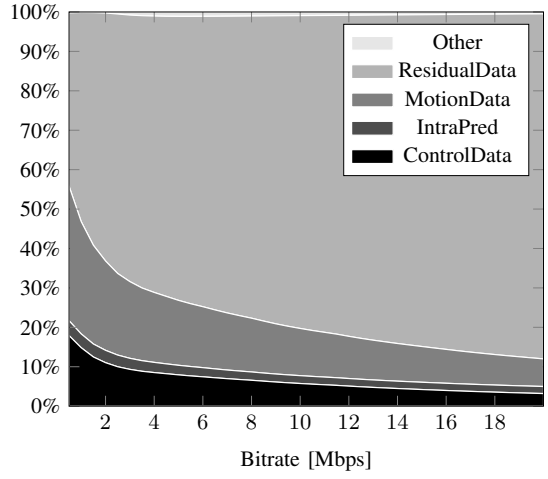
(b) B frames



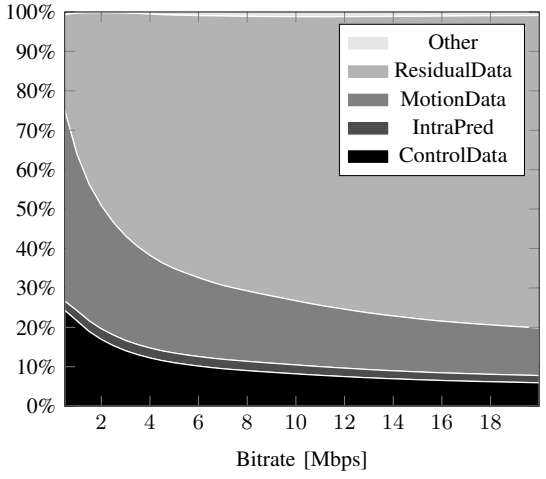
(c) B_0 frames



(d) B_1 frames



(e) B_2 frames



(f) B_3 frames

Fig. 8: Contributions of categories of syntax elements.

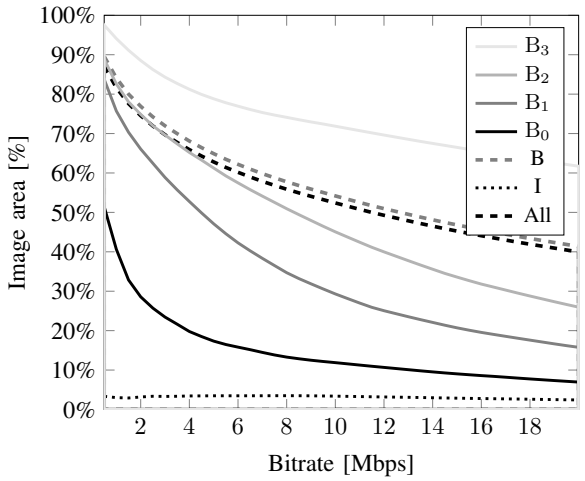


Fig. 9: Percentage of image area where no residual data (no transform coefficients) have been transmitted.

bitrates (4–5 Mbps for HD resolution, 30 fps), for B_0 frames, the percentage of the image area in which no prediction error of image samples is transmitted is about 17%. In the case of B_1 frames it is about 47%. In the case of B_2 and B_3 frames, 61% and 79% respectively. While the bitrate increases, the above mentioned values decrease. The experimental results indicate a very high efficiency of inter-frame prediction in the HEVC.

V. Conclusions and final remarks

The paper documents the results of research that was conducted in order to analyze the compressed data stream from the HEVC video encoder. The analysis was focused on the extraction of the data that will be useful in the development of the mechanisms for HEVC bitrate control. Detailed results were presented and discussed in section IV.

On the basis of the obtained results, the following general conclusions can be highlighted:

1) As expected, the contribution of individual frame types in the stream of the encoded data is not the same. The exact share of data strictly depends on bitrate and the type of frame (refer to Fig. 6 and 7). Depending on the type of frame, data can be modeled by linear, logarithmic or power functions (equations 2–6 in Section IV).
2) The relationship between QP and the target bitrate strongly depends on the type of a video. Although this relation may be modeled by a forth order polynomial function, the analysis of the image content is necessary to calculate the parameters of the function.
3) Although the prediction mechanisms in HEVC are much better than the ones from the previous standards, there is still a large percentage of data related to the prediction error of image samples present in the HEVC encoded bitstream. In the case of I frames it is 85–90% of all data, whilst in B frames it is 50–80%. Therefore, similarly as in older standards (e.g. AVC) the bitrate control can be done on the basis of the amount of data that represents the residual signal.
4) The highly improved compression efficiency of HEVC is related to very efficient prediction of image samples, in particular inter-frame prediction. For broadcast quality video, on average, for about 70% of picture blocks, no prediction error is transmitted.

Presented results make the basis for further works of authors that concern the bitrate control of the HEVC encoder. The works are carried out under the LIDER Grant of the National Center for Research and Development, Poland.

Acknowledgment

Research project was supported by The National Centre for Research and Development, Poland. Grant no. LIDER/023/541/L-4/12/NCBR/2013.

References

[1] W. A. Pearlman, A. Said, *Digital Signal Compression. Principles and Practice*, Cambridge University Press, 2011.
[2] J. W. Woods, *Multidimensional Signal, Image, and Video Processing and Coding*, Second Edition, Elsevier, 2011.
[3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, *Overview of the High Efficiency Video Coding (HEVC) Standard*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668, December 2012.
[4] ISO/IEC and ITU-T, *High Efficiency Video Coding (HEVC)*, ISO/IEC 23008-2 (MPEG-H Part 2) / ITU-T Rec. H.265, April 2013.
[5] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T.-K. Tan; T. Wiegand, *Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC)*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1669–1684, December 2012
[6] B. Li, H. Li, L. Li, J. Zhang, *λ-domain rate control algorithm for high efficiency video coding*, IEEE Transactions on Image Processing, vol. 23, no. 9, pp. 3841–3854, September 2014.
[7] T. Grajek, M. Domański, *New model of MPEG-4 AVC/H.264 video encoders*, 17th IEEE International Conference on Image Processing (ICIP), pp. 961–964, September 2010.
[8] T. Grajek, M. Domański, *Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders*, Lecture Notes in Computer Science - Computer Vision and Graphics, pp. 384–391, Springer Berlin Heidelberg, 2010.
[9] J. Lainema, K. Ugur, A. Hallapuro, *Single entropy coder for HEVC with a high throughput binarization mode*, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G569, Geneva, Switzerland, November 2011.
[10] V. Sze, M. Budagavi, *High Throughput CABAC Entropy Coding in HEVC*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1778–1791, December 2012.
[11] J. Stankowski, D. Karwowski, T. Grajek, K. Wegner, J. Siast, K. Klimaszewski, O. Stankiewicz, M. Domański, *Bitrate Distribution of Syntax Elements in the HEVC Encoded Video*, International Conference on Signals and Electronic Systems (ICSES), Poznan, Poland, 2014.
[12] HEVC test model reference software available online: https://hevc.hhi.fraunhofer.de/svn/
[13] F. Bossen, *Common test conditions and software reference configurations*, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-J1100, Stockholm, Sweden, July 2012.
[14] F. Bossen, *CE1: Table-based bit estimation for CABAC*, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document: JCTVC-G763, Geneva, Switzerland, November 2011.
[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing*, Third Edition, Cambridge University Press, 2007.
[16] T. Chiang, Y.-Q. Zhang, *A new rate control scheme using quadratic rate distortion model*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, pp. 246–250, 1997.
[17] M. Domański, A. Łuczak, *Simple global model of an MPEG-2 bitstream*, IEEE International Conference on Multimedia and Expo, ICME, pp. 645–648, Lausanne, Switzerland, 2002.

[18] S. C. Lim, H. R. Na, Y. L. Lee, *Rate control based on linear regression for H.264/MPEG-4 AVC*, Signal Processing: Image Communication, vol. 22, no. 1, pp. 39–58, Elsevier, 2007.
[19] Y. Zhou, Y. Sun, Z. Feng, S. Sun, *New rate-distortion modeling and efficient rate control for H.264/AVC video coding*, Signal Processing: Image Communication, vol. 24, no. 5, pp. 345–356, Elsevier, 2009.

Bitrate Distribution of Syntax Elements in the HEVC Encoded Video

Jakub Stankowski, Damian Karwowski, Tomasz Grajek, Krzysztof Wegner, Jakub Siast, Krzysztof Klimaszewski, Olgierd Stankiewicz, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology, POLAND
{jstankowski, dkarwow, tgajek, kwegner, jsiast}@multimedia.edu.pl

***Abstract*— The paper focuses on two problems: 1) analysis of bitrate contribution of I and B frames in encoded HEVC data stream, and 2) analysis of distribution of HEVC syntax elements in I and B frames. The relationship between sizes of I and B frames located at a different levels in hierarchical coding scheme was experimentally determined, as well as contribution of HEVC syntax elements in encoded streams of I, B0, B1, B2, and B3 frames. The results are presented in the paper. All presented results concern a wide range of QP values in a video encoder.**

Keywords— HEVC; video compression; video data analysis; video bitstream content

I. INTRODUCTION

Currently, more than half of the total traffic in telecommunication networks is related to video. Without any doubts, this factor will steadily increase in the next years. This raises the need for an efficient compression of a video in order to reduce the data transmission costs. Among the known techniques of a video compression, of particular importance are hybrid coding schemes of a video with intra-frame and inter-frame prediction followed by lossy transform coding of residual image data [1]. The hybrid coding scheme has been included in many international and commercial recommendations (MPEG-2, H.263, AVC, VC-1, AVS, HEVC), and find now widespread applications.

The state-of-the-art in the field of hybrid video coding is the new High Efficiency Video Coding technology defined in international recommendation of ISO/IEC and ITU-T [2]. The HEVC technology is a breakthrough in a video compression and enables significant reduction of a bitstream representing encoded images [3]. Although the main idea of encoding as well as the general structure of the video encoder remained unchanged when compared to previous solutions (e.g. MPEG-2, AVC), there are a lot of improvements and new coding tools that distinguish the HEVC from its predecessors [4]. As a result, the encoded data stream of HEVC is different and not compatible with data streams that are produced by older encoders.

Significant differences in the stream structure pose a problem for efficient HEVC encoder control, since the older solutions, for older encoders (e.g. MPEG-2, H.263, MPEG-4 AVC) do not work as expected. These solutions are extremely specialized, since the base of their action is the knowledge of the structure of encoded data that is produced by encoder e.g. [5],[6]. Compared to the predecessors, different syntax elements are produced in the HEVC encoder. Quantitative features of data that form a bitstream are also different in the case of HEVC. As a result, bitrate control methods that are used in older encoders cannot directly be used in the new HEVC encoder. Experiments on the structure of a video bitstream, already done for older encoders, must be repeated for the HEVC technology. Thus, the topic of our research was to analyze the HEVC bitstream properties and contents.

II. RESEARCH PROBLEM AND GOAL OF THE WORK

This work is a contribution to further study on efficient bitrate control in the HEVC encoder. The goal of the work is to perform a thorough analysis of data stream that is produced by the HEVC encoder. Results obtained in this field up to date and presented in the literature do not give answers for many detailed questions. For example, in [7] results obtained for I and B frames were averaged. There are no separate results for individual frame types. Additionally, experiments were conducted only for two QP values. Moreover above mentioned results are gathered using early version of HM software and do not correspond to final HEVC standard. In [8] authors focused on the average number of context/bypass coded bins without detailed analysis for separate syntax elements.

Our goal is to examine the following issues:

- the relationship between the size of images of a different types in encoded data stream (I- and B-frames),
- the relationship between the size of individual images that belong to different encoding levels in hierarchical coding of images i.e. I-, B0-, B1-, B2-, B3-frames (see explanation in the next section) and use of B frames,
- the distribution of syntax elements in encoded stream of data representing I-, B0-, B1-, B2-, B3-frames,
- influence of the value of quantization parameter (QP) on the abovementioned properties.

III. Methodology of the Research

This work focuses on analysis of HEVC encoded data stream that will probably be a typical stream to be used in the future home theatre systems and broadcasting environments. Therefore, the specific assumptions (presented below) about the methodology of experiments were made.

Research works were done using the commonly available HM 10.0 reference software of the HEVC standard [9]. Using this software, a series of experiments was done with the following encoding conditions.

- Seven 1920x1080 test video sequences were used: *bluesky*, *pedestrianarea*, *riverbed*, *rushhour*, *station2*, *sunflower*, *tractor*. The sequences are recommended by ISO/IEC MPEG as well as ITU-T VCEG working groups as appropriate for the purposes of research works on video compression.
- 'Random access' encoding scenario was used, as defined in 'common test conditions' [10]. In particular, a video sequence was divided into Groups Of Pictures (GOP), with a strict division of a GOP into I-, B-frame types and hierarchical encoding of B-frame types (the so-called B0, B1, B2 and B3 frames) within GOP (see Fig. 1. for details).
- Experiments were done for a wide range of target bitrates by changing the value of quantization parameter (QP) in encoder. Tests were done for QP from 10 to 40 with step equal to 1.
- Other settings of the encoder were set as suggested in 'common test conditions'.

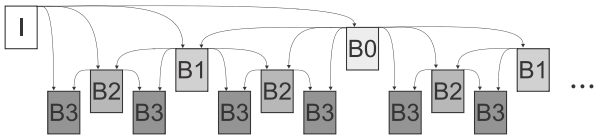


Fig. 1. Hierarchical coding scheme. Only part of prediction sources are marked.

As a result of research, 320 encoded data streams were achieved. Each of data streams was the subject of careful examination, which was conducted for the purposes of:

- Analysis of contribution of individual frames (I, B0, B1, B2, B3) in the encoded data stream, as a function of QP value.
- Detailed analysis of contribution of HEVC syntax elements in individual frames (I, B0, B1, B2, B3), depending on the value of QP.

The partial results were then averaged in order to draw more universal conclusions. The volume of data that represents the partial results does not allow for their inclusion in this paper. Only averaged results were presented in the paper.

The partial results have been published on the companion website to this paper: *www.multimedia.edu.pl/icses2014-hevc.*

IV. Experimental Results

In the first phase of experiments, the general information about distribution of the bits among individual frames of HEVC was collected. The results are presented in Fig. 2 as a function of QP value.

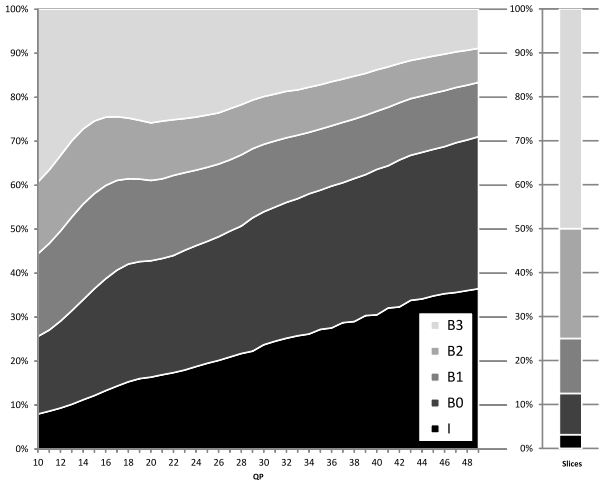


Fig. 2. Contribution of individual frames in encoded data stream, averaged over 7 test sequences results for I, B0, B1, B2, and B3 frames (left). Contribution of frames in total number of slices (right).

In adopted scenario of experiments, the number of I, B0, B1, B2, and B3 frames constituted 3.1%, 9.3%, 12.5%, 25%, 50% of all encoded frames, respectively (it results from the assumed structure of GOP during experiments). Nonetheless, from the viewpoint of the amount of data, the I frames still make a large part of encoded data stream (from 10% to 30% depending on the value of QP). This result is comparable to contribution of B0 frames in a stream – from 15% to 40% depending on the QP value. In this way, I and B0 frames make a significant part of the encoded data stream – from 25% to 70% of the whole data stream depending on the value of QP. The rest of the bitstream is formed by the remaining frames: about 20% for B1 frames, about 15% for B2 frames and 10%-40% for B3 images (depending on QP value).

Presented above knowledge about distribution of data between I and B frames can be only a basis for frame-level mechanism of bitrate control in encoder. Therefore, in the second phase of experiments, the individual frames were additionally thoroughly analyzed in terms of syntax elements distribution in a frame. Detailed statistics obtained for I, B0, B1, B2, and B3 frames were presented in Fig. 3 below for a wide range of QP values.

The main observation is that a type of frame (I or B) as well as the value of QP affect strongly quantitative results.

In the case of I frames, the essential part of the bitstream is formed by three types of data: transform data for luma (70% for lower QPs and 55% for higher QPs), transform data for chroma (20% for lower QPs and 10% for higher QPs), and intra prediction mode (10% -15% depending on value of QP).

In the case of B frames (B0, B1, B2, B3) there is higher number of syntax elements that form a bitstream. Among the

syntax elements present in B frames, the following types of data should be distinguished: transform data for luma (80% for lower QPs and 20% for higher QPs), transform data for chroma (20% for lower QPs and 1% for higher QPs), intra and inter prediction modes (2%-7% and 2%-20% respectively), merge index (1% for lower QPs and 40% for higher QPs) and skip flag (1%-9% depending on value of QP). The other types of data make relatively small part of the image data stream.

Besides the value of QP, detailed statistics are obviously dependent on the hierarchy of a B frame in a GOP. What is worth noting, the amount of merge index and skip flag data in B2 and B3 frames is higher relative to that for B0 and B1 frames. At the same time, the amount of transform data is smaller in B2 and B3 when compared to B0 and B1 frames.

The results obtained for QPs in the neighborhood of 30 are of particular importance for high quality video encoding scenario. Bar charts in Fig. 3 present these results in more detailed way. In this case, the essential part of I frame falls on transform data for luma (about 65%), transform data for chroma (about 16%), and information on intra prediction mode (10%). Distribution of data in B0, B1, B2, and B3 frames is relatively aligned: transform data for luma makes 65%-72% of total bitstream while the inter prediction mode signaling makes 7% to 10% of the bitstream. Notable difference between individual B frame types lie in contribution of transform data for chroma as well as skip flag.

For comparison purposes, the influence of hierarchical QP value in the case of B0, B1, B2, and B3 frames was presented in the second bar chart on Fig. 3. The value of QP affects the amount of transform data in the final stream.

The following conclusions can be drawn from the presented results:

- separate mechanisms of bitrate control should be used for I and B frames due to different statistics of syntax elements,
- the statistics for B0 and B1 frames are similar to each other, so one common mechanism of bitrate control for these frames can be used,
- the results for B2 and B3 correspond well, so one method of bitrate control can be used for both, B2 and B3 frames.

V. Conlusions and Final Remarks

Analysis of the HEVC bitstream content was documented in this paper.

The experimental data give useful information about the size of individual frames in encoded stream, but also information about distribution of syntax elements in frames of a given type. Valuable result of the work is a huge collection of partial results which are accessible via companion website: www.multimedia.edu.pl/icses2014-hevc. They are an important contribution to the knowledge about efficient coder control mechanisms for the new HEVC standard.

On the basis of detailed results, two general conclusions can be highlighted. First, I and B0 frames make a significant part of the encoded data stream: 25% to 70% of the total bitstream (exact value depends on QP). Second, four syntax elements constitute an important part of data stream: transform data for luma and chroma, mode of prediction, merge index and skip flag. Detailed values depend on image type and QP value. When compared to AVC, the residual data is significantly larger part of the bitstream (HEVC: 75% on average, AVC: 65% on average). The contribution of motion data is very similar in both standards (~10%). The results presented in the paper may be a contribution to further study on efficient HEVC bitrate control.

Acknowledgment

Research project was supported by The National Centre for Research and Development, Poland. Grant no. LIDER/023/541/L-4/12/NCBR/2013.

References

[1] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," Selected Areas in Communications, IEEE Journal on , vol.5, no.7, pp. 1140-1154, Aug 1987.

[2] ISO/IEC 23008-2 (MPEG-H Part 2) / ITU-T Rec. H.265: High Efficiency Video Coding (HEVC), Apr. 2013.

[3] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T.-K. Tan; T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp. 1669-1684, Dec. 2012

[4] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp. 1649-1668, Dec. 2012.

[5] T. Grajek, M. Domańnski, "Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders", Lecture Notes in Computer Science - Computer Vision and Graphics, pp. 384-391, Springer Berlin Heidelberg, 2010.

[6] T. Grajek, M. Domanski, "New model of MPEG-4 AVC/H.264 video encoders," 17th IEEE International Conference on Image Processing (ICIP), pp. 961-964, Sept. 2010.

[7] J. Lainema, K. Ugur, A. Hallapuro, "Single entropy coder for HEVC with a high throughput binarization mode", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G569, Geneva, Switzerland, Nov. 2011.

[8] V. Sze, M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp. 1778-1791, Dec. 2012.

[9] HEVC test model reference software – available online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[10] F. Bossen, " Common test conditions and software reference configurations ", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-J1100, Stockholm, Sweeden, Jul. 2012.

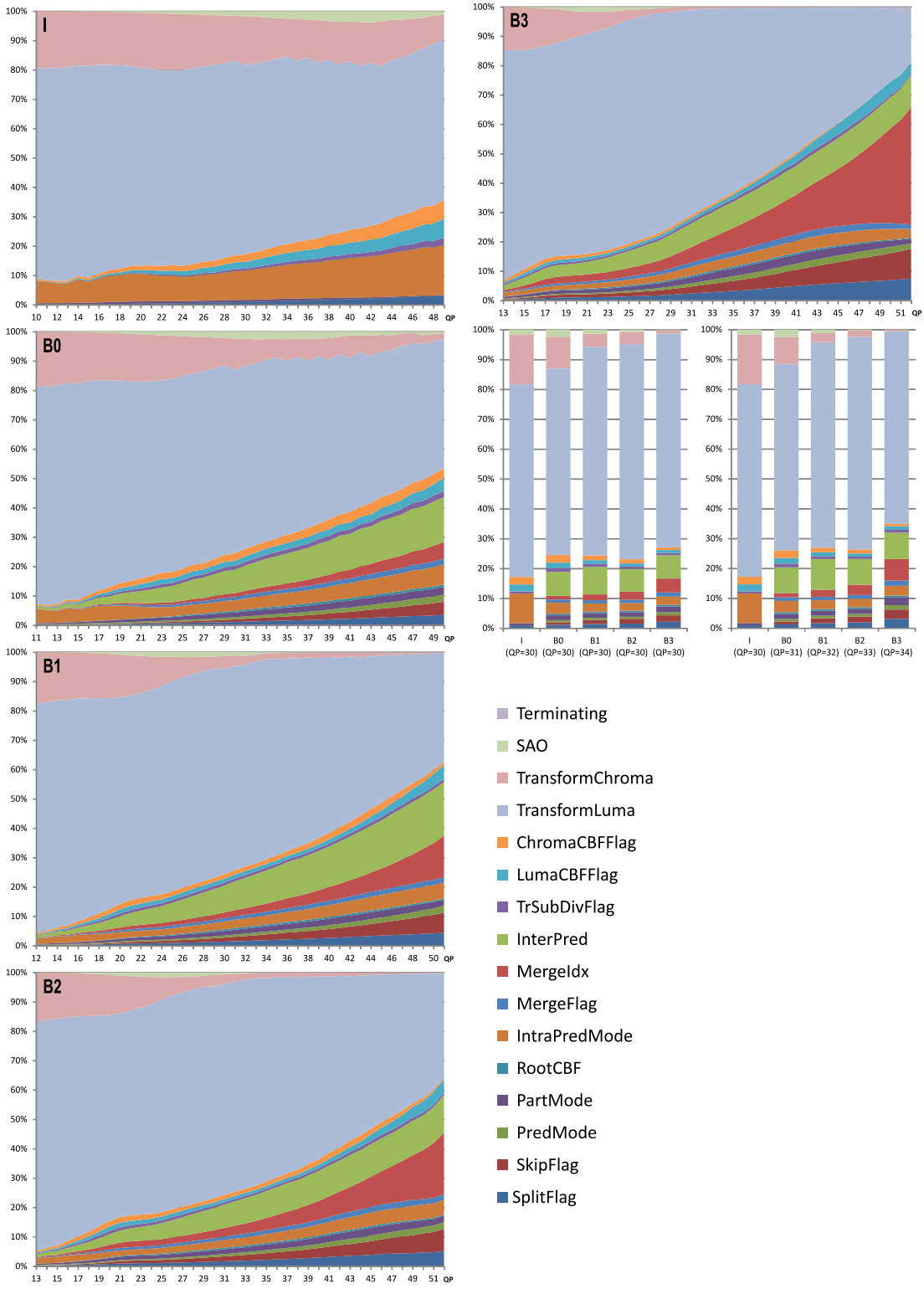


Fig. 3. Contribution of individual HEVC syntax elements in encoded **I**, **B0**, **B1**, **B2**, **B3** frames (averaged results from 7 test sequences).

Analysis of CABAC context models efficiency in the HEVC

J. Stankowski, D. Karwowski, K. Wegner, O. Stankiewicz , K. Klimaszewski, T. Grajek
[1] Poznan University of Technology, Chair of Multimedia Telecommunications and Microelectronics, POLAND
{jstankowski, dkarwow}@multimedia.edu.pl

***Abstract* – This paper presents detailed analysis of the CABAC efficiency when operating in the HEVC in a number of bits per number of bins manner. The paper reveals what fraction of bit is produced by the CABAC for each and every binary symbols encoded in HEVC. The statistics of bins encoded with a given efficiency are presented for the three types of video data: 1) prediction error of image samples, 2) motion data, and 3) control data.**

Keywords – ***CABAC entropy coding, HEVC, Context models***

I. INTRODUCTION

In the final stage of video encoding entropy coding is always used. Entropy coding plays an important role in video compression, because it reduces the statistical redundancy that remains within syntax elements that are produced by video encoder.

In both the AVC and the HEVC video encoders, the syntax elements that are the input for the entropy coding represent various types of video data. In general, these are the following types: 1) prediction error of image samples, 2) motion data, and 3) control data. The three types of data listed above have different statistical properties. In addition, the statistics of the data varies with the change of image content.

The state-of-the-art in the field of entropy coding of video data is Context-based Adaptive Binary Arithmetic Coding (CABAC) [1,2] that has been introduced firstly in Advanced Video Coding AVC MPEG 4 part 10/H.264 standard of ISO/IEC and ITU-T [3], and later in the new High Efficiency Video Coding HEVC MPEG-H part 2 / H.265 international standard of the same organizations [4]. Research on efficiency of CABAC clearly indicate that it ranks among the most efficient entropy encoders used in video compression techniques [1].

As the CABAC is in fact binary arithmetic coder it is able to encode only two binary symbols: 0 and 1. Because of that each syntax element (that are not binary in nature, likes i.e. motion vector differences) needs to be translated into some string of binary symbols. A single binary symbol of this string is called a bin.

In CABAC estimation of probability of each bin is done by the so-called context model. Probability of appearance of 0 and 1 varies and in general depends on actual number of the encoded bin, type of the encoded syntax element and values of the previously encoded bins within a given syntax element.

Therefore, for a given image block, the probability of a bin is estimated with respect to the context information, that holds the information about the values of the given bin in neighboring blocks of the image. Taking into account the possible values of the context information (for a given types of bins), and different types of syntax elements that are defined in HEVC, the whole stream of bins is divided into the total number of 154 sub-streams of bins, in order to track the statistics of data independently in each of the sub-stream [2,4]. Therefore it is also said that CABAC uses 154 independent probability estimators (context models), as illustrated in Fig. 1.

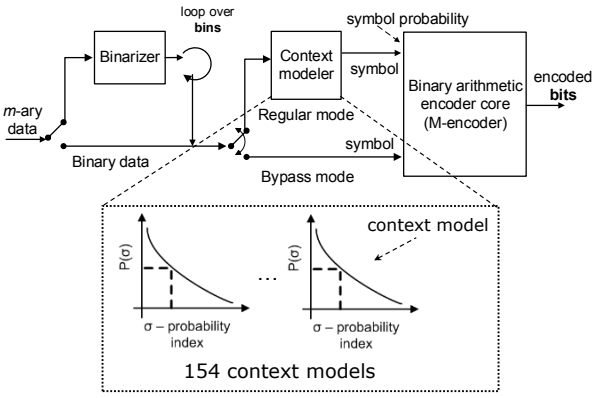


Figure 1. General block diagram of CABAC entropy encoder in HEVC.

All bins are fed into a binary arithmetic encoder core (the so-called M-encoder) which produces bits representing encoded bitstream. It is commonly known that not every bin feed into a binary arithmetic encoder produces a bit at the output. On average CABAC has an ability to represent each bin merely on a fraction of bit.

II. RESEARCH PROBLEM

The goal of this work is to analyze in details for every bin (every 0 and 1 from a stream of bins) what fraction of bit was produced by the CABAC in response to a given bin at the input. This fraction of bit produced in response to a given bin can be understand as compression performance of that particular bin.

The questions we are trying to answer is: How many bins are compressed effectively? How many bins are compressed ineffectively leading to data expansion? What portion of bins is compressed with particular efficiency (i.e. leading to

production of 0.2 or 0.3 bit at the output of arithmetic encoder core).

Because we have analyzed each encoded bin separately, conducted experiments gives a large set of experimental data. For this reason, the partial results have been grouped together, first in 154 CABAC context models, and then according to a type of syntax element that have produced a given bin. Following three groups of syntax elements have been used: symbols related to prediction error of image samples, symbols related to motion data and symbols related to control data.

III. Efficiency of syntax element CABAC compression – methodology

Experiments were done with the HM 10.0 reference software of HEVC [5] configured in the Common Test Conditions (CTC) scenario [6]. In particular, random access test settings were used with a strict division of encoded video into I, and B frame types. In contrast to the CTC guidelines, a wider range of QP values was considered in the research (QP=1-40) when encoding a set of 1920x1080 test video sequences (*bluesky, pedestrianarea, riverbed, rushhour, station2, sunflower, tractor*). The results of the encoding is a comprehensive collection of 280 encoded data streams.

The encoded data streams were then analyzed using the authors' Bitstream Analyzer software. The analyzer performs full CABAC decoding of encoded data streams and calculates the fraction of bit produced by the CABAC in response to each and every bin. The fraction of bit was calculated by the use of the CABAC counter estimator [7]. Gathered fraction of bit produced for every bin was the basis for evaluating performance of CABAC for each of the sub-stream.

Based on the analysis of CABAC efficiency for each of the context models, the bins were divided into separable groups containing bins for which the ratio (defined as: fraction of bit produced by each bin) is contained in a fixed range of values. In particular, the range 0–2 of the ratio was studied, which has been divided into smaller sub-ranges of length 0.2.

IV. Efficiency of CABAC context Models – results

A. *Results for all the CABAC bins*

Among all the coding bins, the largest group makes these symbols, for which the ratio is in the range 0.8–1.0. This group represents 30%–55% of all data symbols. A much smaller proportion of symbols was observed in the other groups. And so, for the 15%–20% of symbols, there was obtained the ratio from range 0.6–0.8, 8%–13% of symbols is encoded with the ratio 0.4–0.6, the ratio in the range 0.2–0.4 was obtained for 5%–12% of the bins, and even lower ratio (0.0–0.2) was obtained for 3%–20% of the bins. In the case of less than 5% of the bins encoder led to the expansion of data, which means that those particular bins was encoded on more than one bit i.e 1.5 bit. Research has shown that the exact proportions of the individual groups of bins in the total data stream depends on the value of QP. With the increase of QP the share of bins for which greater coding efficiency was achieved (i.e. the ratio from range 0–0.8) also increases, and thus the share of bins coded with lower efficiency (the ratio from range 0.8–1) decreases. Detailed results are shown in Fig. 2.

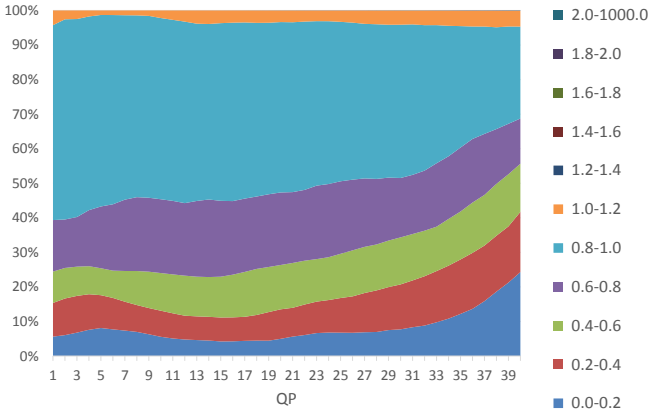


Figure 2. The relative proportion (in %) of the number of bins that have been encoded with the ratio belonging to a given range (0-0.2, 0.2-0.4, etc.). All of the CABAC bins have been considered in statistics.

In order to answer the question why the value of QP affects the results, a detailed analysis of results obtained for individual groups of data is needed. Next subsections present analysis for three main groups of data: 1) data that represent prediction error of image samples, 2) motion data, and 3) control data.

B. *Results for control and motion data*

In the case of both the motion and the control data a strong correlation of proportions of the respective groups of bins with the value of QP was observed. And so, with the increase of QP the share of bins that were encoded with the ratio from range 0–0.8 increases rapidly, thus the proportion of bins for which lower efficiency was achieved (i.e. the ratio from range 0.8–1.0) decreases. In the cases of the control data, 40%–80% of bins are encoded with the ratio 0–0.8, and for 15%–50% of bins the ratio 0.8–1 was achieved. In the case of data representing motion information, the figures are 20%–80% for bins encoded with the ratio 0–0.8 and 10%–75% for bins encoded with the ratio 0.8–1.0. Detailed proportions of the different groups of bins depends on QP and can be read from Fig. 3 and 4. Previous studies of the authors showed that in the case of high QP the pool of coding modes which are actually used by the encoder is strongly reduced [8]. Because the quantization removes the details of the image, the encoder resigns from the use of small image blocks and uses more frequently the blocks of a larger size. Motion estimation performed in larger image blocks results in a more consistent motion field, whereby the prediction of motion data is more efficient (that is, prediction error of motion data often assumes a small value). As a result, the probabilities of data symbols (representing both the control and the motion data) are more varied at higher values of QP. With the increase of QP the entropy of data decreases, which results in higher efficiency of entropy coding.

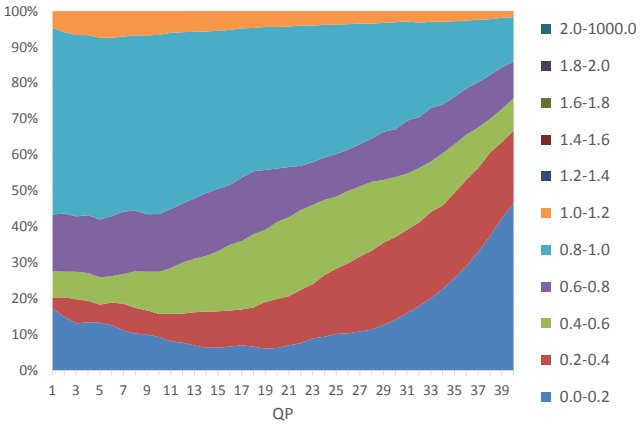


Figure 3. The relative proportion (in %) of the number of bins that have been encoded with the ratio belonging to a given range (0-0.2, 0.2-0.4, etc.). Only CABAC bins that are related to control data were considered in statistics.

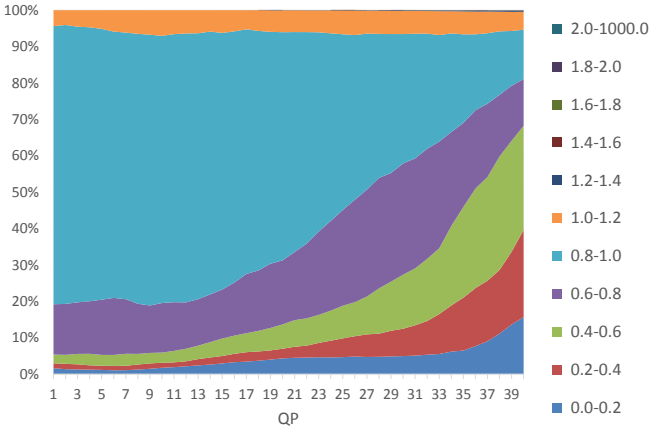


Figure 4. The relative proportion (in %) of the number of bins that have been encoded with the ratio belonging to a given range (0-0.2, 0.2-0.4, etc.). Only CABAC bins that are related to motion data were considered in statistics.

C. Results for data representing prediction error of image samples

In contrast to the results presented in subsections A and B, the results for prediction error of image samples do not show strong variation depending on the QP. On average, for half of the data from this group the ratio from range 0–0.8 is achieved. The remainder of the symbols are coded with the ratio 0.8–1.0 (about 40% of symbols), and symbols for which the number of bits for each and every bin encoded is greater than 1 i.e 1.5 bits per bin (less than 10% of such symbols for which the entropy coding is carried to the expansion of data). Detailed results were presented in Fig. 5.

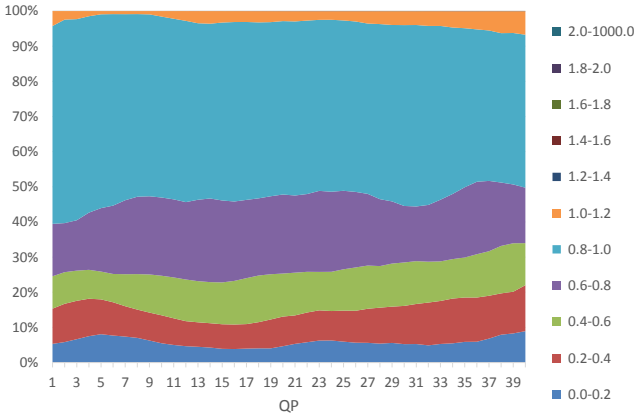


Figure 5. The relative proportion (in %) of the number of bins that have been encoded with the ratio belonging to a given range (0-0.2, 0.2-0.4, etc.). Only CABAC bins that are related to information about prediction error of image samples were considered in statistics.

Prediction of image samples is in the HEVC encoder performed in blocks of a certain size. In general, sizes of blocks from 4x4 to 64x64 are available. The used in the encoder selection criterion of a block size is to minimize the cost of a block encoding, which generally refers to minimization of prediction error of block samples. Thus, for each value of QP the encoder selects the coding mode that minimizes the entropy of image sample prediction error. For this reason, similar efficiency of entropy coding was observed for this group of data, regardless of the QP.

V. CONCLUSSIONS

The paper presents research on the efficiency of the CABAC entropy encoder in HEVC. For each of the 154 context models, performance has been measured as an fraction of a bit produced by each and every encoded bin. It has been shown that such data give detailed information about the efficiency of the CABAC entropy encoder in HEVC. For the sake of presentation, the data has been aggregated and analyzed in groups, which document the efficiency of the algorithm for the three types of video data. The results show that for the vast majority of the bins CABAC performs efficient compression. Although the detailed results depend on the type of encoded data and value of QP (as depicted in section IV), the averaged results show, that for half of the bins the ratio from the range of 0–0.8 bit was achieved, and for most of the other bins the ratio was 0.8–1 bit. What is interesting, there are also some bins, which encoding leads to expansion of the data. In the extreme case, these bins constituted 5% of all the bins. This gives a strong indication that some of CABAC context models can be coded more efficiently.

The results attained during the research may be a basis for future works on development of the CABAC improvements as well as fast mode decision algorithms in a video encoder.

A. *Fast caclulation of cost of application of a coding mode in a video encoder*

For each of image block the encoder selects the compression mode, which is a compromise between the number of bits that are sent to the decoder and the quality of reconstructed block. Thus, for each of the tested modes the abovementioned parameters must be calculated. From the amount of computations point of view determining the bit cost of a mode is more difficult. Accurate determination of the number of bits requires realization of full encoding path, including the entropy coding of data. Realization of all the CABAC functional blocks at the stage of coding modes testing gives accurate results, but such a solution is associated with very long encoding time.

For the purpose of calculating the bit cost for a mode, there is a possibility of significant simplification of the CABAC encoder, without losing on the accuracy of calculation of the cost. With the attained raw results for fraction of bit produced for each and every bin (averaged for each of the 154 CABAC context models) one can determine the bit cost of a mode with high accuracy, relying only on the data observed at the output of the binarizer block. So, we completely ignore the parts of statistical modeling and the core of arithmetic coding when testing the successive compression modes. The available studies on the complexity of the CABAC encoder show that the proposed solution will significantly shorten the encoding time of a sequence [9]. The authors are currently conducting research in the field of the proposed solution.

B. *Improvement of the CABAC entropy encoder*

The raw research results for fraction of bit per bin allow an accurate assessment of the efficiency of each of the 154 CABAC context models. This knowledge can be the basis for further improvements of the algorithm oriented to increase the efficiency of entropy coding. In this context, the improvement of those syntax elements is crucial, for which the fraction of bit per bin exceeds 0.8. In this group of bins, there are also those for which there is expansion (and not compression) of the data (see Fig. of section IV).

Presented in Fig. 3 averaged results for the control data highlights, that there can be even 60% of the bins that are encoded with the ratio of 0.8 or higher. In the case of the motion data, the same situation may take place for even 80% of the bins (see Fig. 4). The things look better for the prediction error of image samples data, for which there are about 50% of the bins that are encoded with the ratio 0.8 or higher. This indicates for great importance of the works on more efficiently encoding of the motion and control data.

ACKNOWLEDGMENT

Research project was supported by The National Centre for Research and Development, POLAND, Grant no. LIDER/023/541/L-4/12/NCBR/2013.

REFERENCES

[1] D. Marpe, H. Schwarz, and T. Wiegand: Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 620-636, July 2003.

[2] V. Sze, M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp. 1778-1791, Dec. 2012.

[3] ISO/IEC 14496-10, Generic Coding of Audio-Visual Objects, Part 10: Advanced Video Coding, March 2006.

[4] ISO/IEC 23008-2 (MPEG-H Part 2) / ITU-T Rec. H.265: High Efficiency Video Coding (HEVC), Apr. 2013.

[5] HEVC test model reference software – available online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[6] F. Bossen, Common test conditions and software reference configurations, Joint Collaborative Team on Video Coding (JCT-VC) of ITUT SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-J1100, Stockholm, Sweden, July 2012.

[7] F. Bossen, CE1: Table-based bit estimation for CABAC, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document: JCTVC-G763, Geneva, Switzerland, November 2011.

[8] J. Stankowski, T. Grajek, D. Karwowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, M. Domański, Analysis of Frame Partitioning in HEVC, Lecture Notes in Computer Science – Computer Vision and Graphics, pp. 602-609, Springer International Publishing Switzerland 2014.

[9] D. Karwowski, Computational complexity analysis of adaptive arithmetic coding in HEVC video compression standard, Image Processing and Communications, September 11-13, 2013, Bydgoszcz, Poland.

Video Quality in Multiple HEVC Encoding-Decoding Cycles

Jakub Stankowski, Tomasz Grajek, Krzysztof Wegner, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology
ul. Polanka 3, 60-965 Poznań
e-mail: {jstankowski,tgrajek,kwegner}@multimedia.edu.pl

***Abstract*—Paper concerns accumulation of video-quality losses in multiple HEVC encoding and decoding cycles. The results of extensive experiments with standard video test sequences are reported. General conclusions has been drawn about relative quality losses after each cycle, as well as about total quality loss with respect to first cycle. In general, quality losses after each cycle are decreasing and total quality loss is saturating after several cycles. Total quality losses after several cycles are hardly dependent on sequence content.**

Keywords—video; compression; video coding; multiple cycles; transcoding; HEVC; H.265; MPEG; MPEG-H

I. INTRODUCTION

Research on video coding is a very active topic due to the fact that every few years the new video compression standard is deployed on the market. In video editing workflow and postproduction multiple encoding and decoding of the same material is particularly important.

Currently (March 2013), new video coding standard HEVC (High Efficiency Video Coding) [1] is being finalized by Joint Collaborative Team on Video Coding (JCT-VC) the join body of MPEG (ISO/IEC JTC1/SC29/WG11) and VCEG (ITU-T SG16 WP3). HEVC is the result of many years of research on video compression and intensive collaboration between numerous research teams. The HEVC technology has been designed to be utilized in ultra-high resolution (UltraHD) systems as well in video streaming systems.

When compared to the widely used MPEG-4 AVC/H.264 technology, HEVC allows for about 50% bitrate reduction with preserving the same subjective image quality [2]. During development of HEVC special attention was paid on improving the compression efficiency for high resolution content (HD, FullHD) and future ultra-high resolution content (UltraHD, 4k, 8k). Transmission of high resolution video is still a challenge for telecommunication network and every possibility to reduce transmission burden is highly desirable [3]. Numerous advantages of HEVC over AVC imply that the new standard is likely to be widely adopted in the nearest future.

The HEVC compression efficiency increase over AVC has been achieved by improvements in most of the existing coding tools and by introducing new tools [4]. The most important improvements are focused on inter and intra prediction, interpolation filters, coefficient coding, flexible structure of coding block size (coding units, prediction units and transform units) and variety of unit sizes (4x4 to 64x64 for prediction and 4x4 to 32x32 for transform). The introduction of new compression standard with the rich set of new coding tools raises a question about consequences of multiple encoding and decoding of the same material. Thus, new coding scenarios involving HEVC technique will probably gain a lot of researchers attention.

Research about error accumulation due to multiple encoding-decoding cycles has been already investigated for such a compression standards as MPEG-2 or JPEG. For example in [5] MPEG-2 working in 4:2:2 profile is proposed as a basic standard in professional studios and post-processing industry due to its limited distortion accumulation. In [6,7] analysis of multiple compression cycles in JPEG and JPEG-2000 standards is considered. Only limited number of cycles has been investigated in the above mentioned literature.

II. PROBLEM DEFINITION

In the scenario when decoded material is again encoded, there is no possibility to avoid re-quantization of previously compressed (quantized) signal. The re-quantization causes inevitable and irreversible quality loss. Moreover, when encoding is performed several times on the same material, each cycle introduces new distortion and causes quality losses. Are the quality losses at each cycle additive or they saturate after several cycles of encoding and decoding? Maybe distortion caused by the next cycle cancels out the previous one? In this paper the accumulation of distortion caused by multiple encoding and decoding cycles of the same material is investigated. Additionally the total quality loss relative to the first coding, as well as the number of cycles after which relative quality loss is negligible will be estimated.

III. METHODOLOGY

Original video material was first encoded with HEVC encoder with several predefined QP values and decoded. Quality of decoded material was measured using PSNR metric for luma (Y) component. Next, the decoded material was again encoded by HEVC encoder with exact the same QP value and again decoded. This procedure was repeated multiple times. Each time the decoded video material quality (with respect to

original material) and resulting bitrate was measured. In this way data of multiple coding cycles using HEVC standard were gathered.

IV. EXPERIMENTAL MATERIAL

In our experiments we have used 12 SD TV (704x567) 25 Hz and 30 Hz video sequences. These sequences are recommended by an international expert group of ISO/IEC MPEG (Moving Picture Experts Group), which develops standards for coding audio and video, as the official video test sequences. They are also used worldwide by researches in processing, compression and quality evaluating. The used video sequences test set had covered wide range of content characteristics. The following sequences have been used: *bluesky*, *city*, *crew*, *harbour*, *ice*, *pedestrian*, *riverbed*, *rushhour*, *soccer*, *station2*, *sunflower* and *tractor*. As a HEVC codec we have used HEVC reference software in version 9.1 [8]. For each video sequence 4 different QP values of 22, 27, 32, 37 was used. Those QP values reflect commonly used bitrates for TV services. Encoding and decoding was repeated 80 times for each of those QP values resulting in total of 12*4*80 = 3840 coding points.

V. RESULTS

Fig. 1. shows rate distortion plot for 4 QP values (22, 27, 32, 37) for exemplary sequence *bluesky*. For all QP values a small quality degradation along with bitrate reduction can be observed. Additionally a detailed view of quality of decoded material and bitrate after each encoding-decoding cycle for QP=27 value has been shown (Fig. 1, bottom). The numbers of the encoding-decoding cycles are marked on the plot. After a couple of encoding-decoding cycles, change in video quality becomes negligible and change in the bitrate begins to be random (Fig. 1, bottom - zoomed part of the chart). The same can be observed for the rest of the test sequences and QP values.

Relative quality loss with respect to first encoding versus relative bitrate savings with respect to first encoding has been shown on Fig. 2. As we can see, after 80th encoding-decoding cycle for *bluesky* sequence, quality in terms of PSNR drops about -1.2 dB with respect to the first encoding. Detailed view for QP=27 is presented on Fig. 2, bottom. Significant quality loss and bitrate reduction occurs in first few encoding-decoding cycles. After 40 encoding-decoding cycles changes in both quality and bitrate are negligible. Similar observation could be taken for the rest QP values and test sequences.

Relative quality loss (ΔPSNR) after each encoding-decoding cycle for all tested QP values and video sequences has been shown on Fig. 3. Depending on video sequence, total quality loss after 80 encoding -decoding cycles is about:

- 0.70 – 1.80 dB for QP=22,
- 1.05 – 2.05 dB for QP=27,
- 1.10 – 1.90 dB for QP=32,
- 1.15 – 1.65 dB for QP=37,

along with about 5-20% reduction of the bitstream.

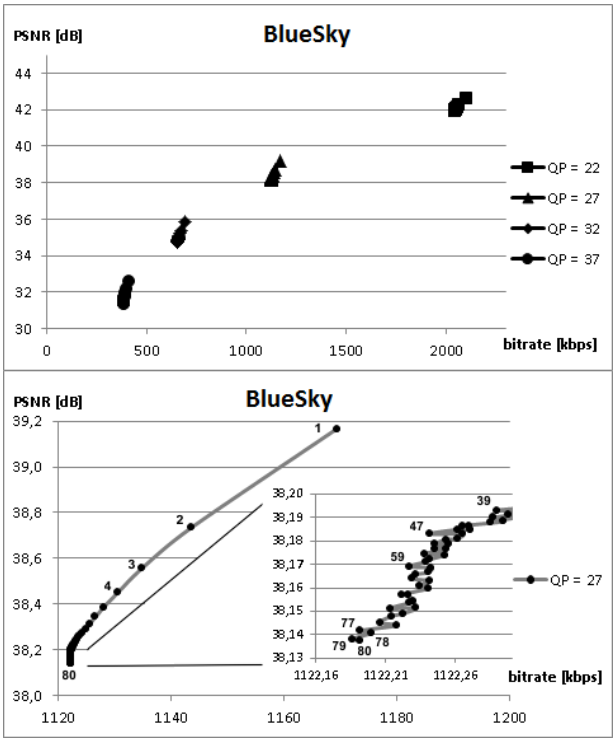


Fig. 1. Results of 80 encoding-decoding cycles of BlueSky sequence: for 4 different QP values (top) and detailed view for QP=27 only (bottom).

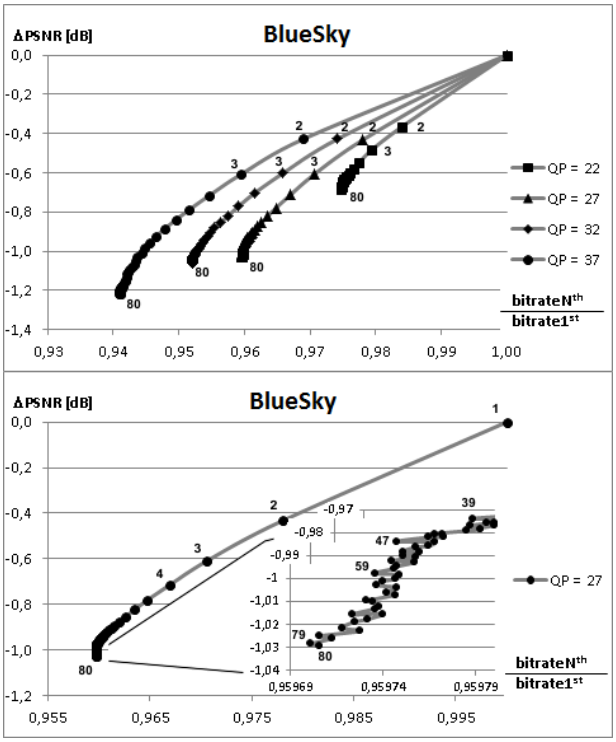


Fig. 2. Quality loss and bitstream reduction after each encoding-decoding cycle relative to first encoding for BlueSky sequence: for 4 different QP values (top) and detailed view for QP=27 only (bottom).

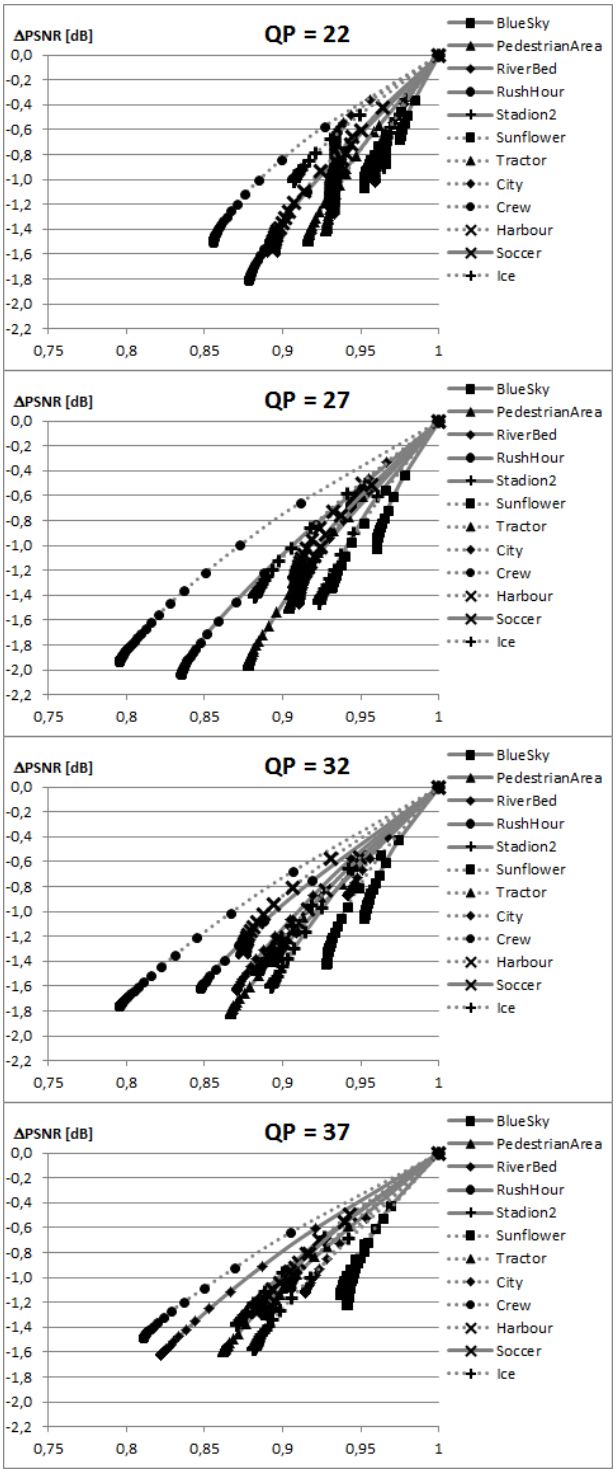


Fig. 3. Quality loss and bitstream reduction after each encoding-decoding cycle relative to first encoding for QP=22,27,32,37.

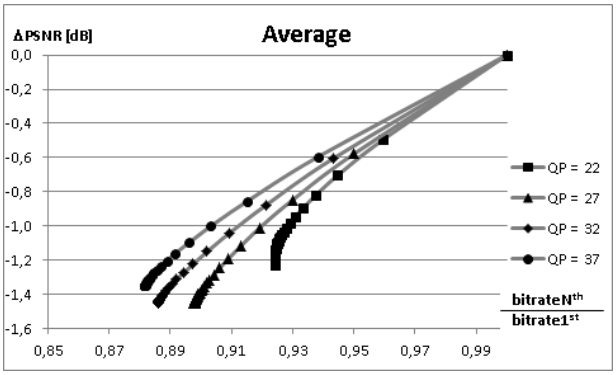


Fig. 4. Quality loss and bitstream reduction after each encoding-decoding cycle relative to first encoding averaged for all test video sequences for different QP values.

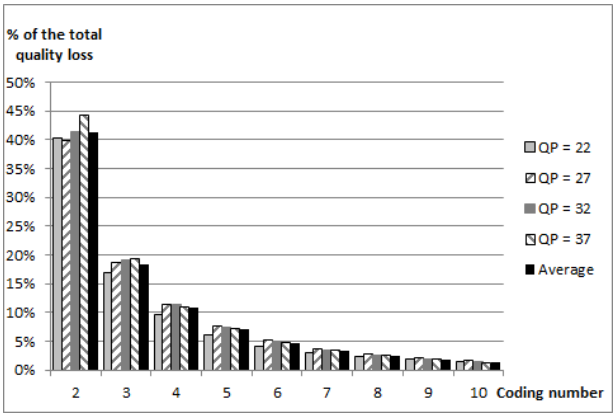


Fig. 5. Percentage of quality loss after each encoding-decoding cycle with respect to the total quality loss after 80 cycles.

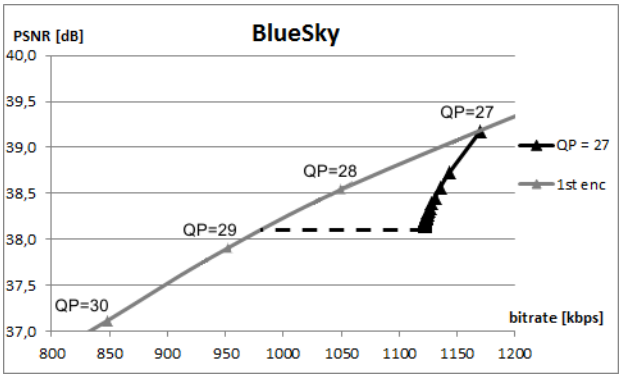


Fig. 6. Comparison of the video quality after 80 encoding-decoding cycles with QP=27 and video quality after first encoding with various QPs. Dashed line represents the same quality in both cases.

Fig. 4 presents averaged quality loss after each encoding-decoding cycle with respect to first encoding. On average quality degradation starts to saturate after 40 cycles and is almost fully saturated after 80 cycles. Total quality loss after 80 cycles is on average -1.5dB and bitrate reduction is on average 10%, both with respect to first encoding. What is interesting accumulation of the distortion (total quality loss) is rather independent on selected QP value.

Moreover, bitrate reduction resulting from multiple encoding-decoding cycles with the same QP value is smaller for high quality images (low QP value) and it increases as QP value increases. Average quality loss after 2nd cycle related to the 1st one and after 3rd related to the 2nd is about 0.5-0.6 dB and 0.2-0.3 dB respectively. More precisely, quality loss after 2nd cycle related to the 1st one is:

- 0.36 – 0.75 dB for QP=22,
- 0.33 – 0.82 dB for QP=27,
- 0.40 – 0.77 dB for QP=32,
- 0.42 – 0.74 dB for QP=37.

Fig. 5 presents quality loss at each encoding-decoding cycle as a percentage of the total quality loss (after 80 cycles with respect to first one) for the first few cycles. Second encoding-decoding cycle is responsible for 41% of the above mentioned total quality loss (on average).

We compared the quality (in terms of PSNR) of the sequence after 80 encoding-decoding cycles with the sequence quality after first encoding (Fig. 6). In order to encode a sequence with the same quality as after 80 encoding-decoding cycles, the QP value must be increased by 1.9 on average. In other words, decreasing QP value by 1.9 assures, that quality of the video after several encoding-decoding cycles will be no worst that it would be after first encoding with unmodified QP value.

VI. CONCLUSIONS

In the paper the accumulation of distortion caused by multiple encoding and decoding of the same material with constant QP value has been investigated. The main observation is that quality loss introduced by multiple encoding and decoding saturates after several cycles. On average, after 40 encoding-decoding cycles the quality losses and bitrate changes are negligible.

Moreover, total quality loss (i.e. quality loss after 80 encoding-decoding cycles with respect to the first encoding) is in range of 0.68 – 2.04 dB (1.4 dB on average). Total quality loss is almost independent of QP value tested. About 41% of the total quality loss is due to quality loss introduced by first encoding. Bitrate reduction after 80 cycles with respect to first encoding is about 10% and it is almost fully saturated.

Quality loss introduced by multiple coding cycle with constant QP value can be compensated by using lower initial QP value (higher initial quality). For 80 encoding-decoding cycles initial QP value has to be decreased by about 2.

ACKNOWLEDGMENT

Research project was supported by The National Centre for Research and Development, Poland. Grant no. NR02-0080-10/2010.

REFERENCES

[1] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, T. Wiegand, "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Last Call)", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, CH, Jan. 2013.

[2] J-R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, T. Wiegand: "Comparison of the coding efficiency of video coding standards - Including High Efficiency Video Coding (HEVC)", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[3] F. Bossen, B. Bross, K. Suhering, D. Flynn: "HEVC Complexity and Implementation Analysis", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1685–1696, Dec. 2012.

[4] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp. 1649-1668, Dec. 2012.

[5] C. Horne, T. Naveen, and A. Tabatabai, "Study of the characteristics of the MPEG2 4:2:2 profile-application of MPEG2 in studio environment", IEEE Trans. Circuits and Systems for Video Technology, Vol. 6, no. 3, pp. 251-272, 1996.

[6] R.L. Joshi, M. Rabbani, and M.A. Lepley, "Comparison of multiple compression cycle performance for JPEG and JPEG2000", Proc.SPIE, vol. 4115, pp. 492-501, 2000

[7] K. Rakowski, M. Domański, "Experimental analysis in multiple image coding-decoding cycles with color transformations," Proceedings of 10th International Workshop on Systems Signals and Image Processing, IWSSIP'2003, Prague, pp. 69-72.

[8] HEVC test model reference software – available online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 24, 2019, accepted June 22, 2019, date of publication July 12, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928423

Analysis of Video Quality Losses in Homogeneous HEVC Video Transcoding

TOMASZ GRAJEK, (Senior Member, IEEE), JAKUB STANKOWSKI, DAMIAN KARWOWSKI, KRZYSZTOF KLIMASZEWSKI, OLGIERD STANKIEWICZ, (Member, IEEE), AND KRZYSZTOF WEGNER, (Senior Member, IEEE)

Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology, 60-965 Poznań, Poland

Corresponding author: Tomasz Grajek (tomasz.grajek@put.poznan.pl)

This work was supported by the National Centre for Research and Development, Poland, under Grant LIDER/023/541/L-4/12/NCBR/2013.

ABSTRACT This paper presents a detailed, quantitative analysis of video quality losses in a homogeneous HEVC video transcoder together with the analysis of the origin of those quality losses and the influence of quantization step alignment on transcoding. With the use of HM15.0 reference software and a set of test video sequences, the cascaded pixel domain video transcoder (CPDT) concept has been used to gather all the necessary data needed for the analysis. This experiment was performed for a wide range of source and target bitrates. The essential result of the work is an extensive evaluation of CPDT, commonly used as a reference in works on effective video transcoding. So far, no such extensively performed study has been made available. The quality degradation between a transcoded video and a video that would be the result of direct compression of the original video at the same bitrate as the transcoded one has been reported. The dependences between quality degradation caused by transcoding and bitrate changes of the transcoded data stream are clearly presented in graphs.

INDEX TERMS Cascaded pixel domain video transcoder, CPDT, HEVC, H.265, video compression, video transcoding.

I. INTRODUCTION

Techniques of hybrid compression of digital video are the most common methods of video transmission used in contemporary IT networks. The hybrid technique that has been developed over the years, has become the subject of numerous standardization projects of expert groups of ISO/IEC MPEG and ITU-T VCEG (e.g. MPEG-2 [1], H.263 [2], AVC [3], [4], HEVC [5], [6] standards). The constantly changing requirements imposed on video compression algorithms, as well as the improving capabilities of hardware, have resulted in faster and faster changing generations of video compression technology.

Today, the state-of-the-art in the field of video compression is the High Efficiency Video Coding (HEVC) technology [6] that has been jointly developed by ISO/IEC and ITU-T and published in 2013 simultaneously as an international standard of ISO/IEC MPEG-H part 2 and recommendation ITU-T H.265 [5]. The new technique is the successor of the widely used and extremely successful Advance Video Coding (AVC) technology [4], [7]. Comparing these two techniques, HEVC allows up to a 2-fold reduction in the size of the encoded images without compromising the quality of video [8], and, more importantly, it supports the compression of ultra-high definition video, which is believed to be the future of video systems. For this reason, it is highly expected that within the next few years, the new HEVC technique will replace the currently used AVC technology in nearly all fields.

Similarly to the previous video compression standards (e.g. MPEG-2, AVC), HEVC has been developed with many different applications in mind, i.e., not only digital television but also mobile terminals, in which a codec has to run on devices with significantly different computing capabilities. The requirements differ for each of those applications and environments, therefore, the set of compression tools used also varies, accordingly. Very often, not only the set of tools is limited but also the parameters of the encoded video itself changed. For example, the target bitrate or resolution of a video must be suitable for the given application and target device. That is why different versions of the same

The associate editor coordinating the review of this manuscript and approving it for publication was Pradeep Kumar Gupta.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

video sequence with different parameters – resolution, bitrate, frame rate, etc. are needed. Taking into account that once a sequence is encoded, the original video signal is no longer available (thus, there is no possibility to repeat the process of encoding from the original signal), the only way to produce a required bitstream with new parameters is to use a video transcoder. Because in the considered case of video transcoding only some parameters of the encoded video are modified, and the general coding technology remains the same (in our case, it is HEVC), this type of transcoding is called homogeneous video transcoding. It is important to note here that during video transcoding generally, it is only possible to obtain a bitstream representing the video signal whose parameters, like image quality, are always lower than the parameters of a bitstream before transcoding. One of the goals of the research on transcoding is to minimize the loss of quality caused by re-encoding (transcoding) of a video in a case when the previously lossy compressed video is used to create a new bitstream at a different target bitrate. Especially, the quality (losses) of thus created video in comparison with the quality of a video encoded from the original sequence is an interesting issue.

It is commonly known that in the case of hybrid video compression transcoding always introduces some video quality loss, regardless of the choice of the target bitrate of a transcoded video [9]–[13]. This is a direct result of lossy operations performed each time by a video encoder. Although research on the impact of video transcoding on the quality of video, as well as its impact on the size of the encoded bitstream has already been carried out for the older compression techniques (e.g., MPEG-2 or AVC) [9], [12]–[16], no extensive study of video transcoding performance has been presented in the literature so far. Moreover, the conclusions of the existing partial studies cannot be directly transposed to the new HEVC technology. The main reasons are the differences between the older and the new compression techniques that strongly affect the compression process. A kind of exception here is the work [11] (published by some of the authors of this text), which shows the results of the accumulation of video quality losses in multiple HEVC encoding and decoding cycles. However, the above mentioned work also does not analyze in detail the mechanism of occurrence of video quality losses in the process of HEVC transcoding. Such an analysis is the subject of this work.

II. SPECIFIC OBJECTIVES OF THE WORK

Homogeneous HEVC transcoding has been the subject of research works for a few years now [17]–[25]. The main goal of the already published works was to reduce the computational complexity of HEVC re-encoding by exploiting the high similarity of selected parameters of the original (input) and the transcoded bitstreams. In these works, the performance (computational complexity and coding efficiency) of the proposed solutions in relation to the performance of the reference pixel-domain cascaded transcoder has been compared. The phenomenon of video quality degradation in the

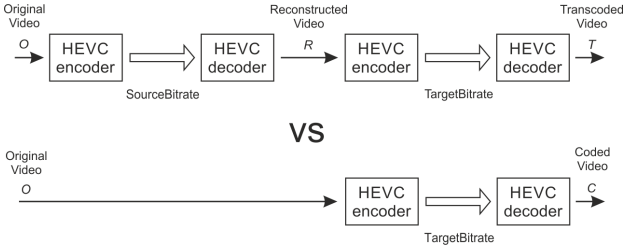


FIGURE 1. Methodology of comparing quality losses during video transcoding.

transcoder has not been the subject of a detailed analysis. For this reason, the authors of those works did not indicate the optimal conditions for HEVC video transcoding for which the loss of video quality is minimal.

The goal of this work is to perform a deep study of the impact of video transcoding on the quality of transcoded video material using the newest video compression standard. In this work, a cascaded pixel domain HEVC video transcoder is considered. That is, the process of transcoding is based on feeding a once HEVC-encoded and decoded video to another HEVC encoder (Fig. 1) and performing the full process of compression. As it was mentioned before, such a configuration is commonly used in the literature as a reference in research on the improvement of video transcoding. Therefore, our work may be used as a reference for any future improvements of the transcoding process for HEVC data. The cascaded method can also be used in practice as the simplest possible way of transcoding.

During the work described in this article the authors tried to answer two fundamental questions. First, what is the difference between the quality of an HEVC-transcoded video (transcoded with a specific value of target bitrate) and the quality of a video that would be achieved after a direct compression of the original video (assuming that such a video could be made available) at the same target bitrate? What factors affect this change of quality? Secondly, is there some optimal scenario of video transcoding (understood as a range of changes of bitrate in the transcoder), for which the loss of video quality is minimal? It must be stressed again that in this work ''the loss of video quality'' is expressed as the difference between the quality of a transcoded video (transcoded at some value of a target bitrate) and the quality of a video achieved after encoding the original sequence at the same target bitrate. In order to better understand the mechanism of quality losses, short remarks on this topic are highlighted in the next section.

III. LOSS OF QUALITY IN THE HEVC ENCODER – SHORT REMARKS

A detailed, quantitative analysis of the sources of video quality losses in the HEVC video encoder / transcoder is beyond the scope of this work. Therefore, only some remarks clearing out the reasons of such losses are outlined in this section.

The degradation of quality during transcoding is owed to the fact that generally HEVC is a lossy compression standard, i.e., some data are intentionally omitted in the process

of encoding. There are three separate sources of those losses (quality degradation) in the HEVC encoder. The first one is Discrete Cosine Transform (DCT)-like transformation of residual data, which is not fully reversible. The second is the quantization of the resultant transform coefficients. The third and the final source of losses is associated with the modification of reconstructed image samples by in-loop deblocking and sample-adaptive offset (SAO) filters. They aim to improve the visual (subjective) quality, but at the same time, they may introduce value shifts to pixel values, and thus decrease the objective quality of the image.

HEVC exploits finite precision approximation of DCT transformation [4]. Taking into consideration both the coding efficiency and complexity aspects, the HEVC standard requires 16-bit depth of values obtained at each stage of the transform process, including the sign bit (for 8-bit input image sample representation). In order to fulfill this requirement, there is a necessity of bit shifting and clipping operations at some stages of residual data transformation. As a result, a combination of forward (performed in the encoder) and inverse (calculated in the decoder) DCT-like transformations gives only near-lossless reconstruction of input image samples, and not fully lossless. So, there is some (in practice very minor) loss of quality. It should be noted, however, that this quality loss happens twice in the transcoding procedure, when a sample goes through the DCT-like transformation chain: first, during the encoding of the original material, and for the second time, during the transcoding itself.

The step that introduces losses in all hybrid video codecs is the quantization of transform coefficients of residual data [6], HEVC not being an exception from this rule. The process of quantization is equivalent to (however, from the mathematical point of view, not exactly the same as) dividing transform coefficients by the quantization step size (so called *QStep*) together with subsequent rounding operations. This leads to an irreversible loss of part of information about the signal – the reconstructed images differ significantly from the original ones. The level of quality loss depends directly on the quantization step size (the value of *QStep*) – the higher the value, the coarser the representation of transformed residual data. Coarse representation results in a lower quality of reconstructed images, but gives a higher compression ratio of the video.

Due to the nature of block-based encoding used in the HEVC coder, the reconstructed image can exhibit some distortions, usually called block artifacts, which are very annoying for viewers, especially at lower bitrates. In order to improve the subjective quality of the reconstructed images, deblocking and SAO filters are used [6], [26], [27]. This helps to improve the subjective quality, but, at the same time, can degrade the objective quality metrics, such as still commonly used PSNR.

In order to reduce the blocking artifacts (by the use of a deblocking filter) and attenuate the ringing artifacts (with the help of a SAO filter) that are present in the reconstructed images after decoding, the values of image samples are appropriately modified before writing them to a file or showing them on a screen. These filtering operations, when carried out several times (as is in transcoding), may change the values of image samples essentially. Even though the goal of those filters is to improve the subjective quality, the effect of these operations on the PSNR metric (used to measure the quality in transcoding) is difficult to predict and the use of those filters may lead to a further reduction of the objective quality metric value.

IV. LOSS OF QUALITY IN HOMOGENEOUS HEVC TRANSCODING – METHODOLOGY OF EXPERIMENT

Suppose that we have an HEVC-encoded video material, compressed at *SourceBitrate*. We can transcode the bitstream by decoding it and re-encoding at a different *TargetBitrate*, as seen in Fig. 1. The question is: what is the difference between the quality of a transcoded video T (obtained by re-encoding a previously encoded video R to a new bitstream with a rate equal to the *TargetBitrate*) and the quality of a video C achieved by a direct encoding of the original video to a bitstream whose rate is again equal to the *TargetBitrate*? So, what loss of video quality is actually introduced by using the previously encoded video R instead of the original one O. What loss of quality is introduced by video transcoding?

In order to answer this question, an extensive and detailed experiment was conducted. At first, original video materials O have been encoded by an HEVC encoder with the use of all possible quantization parameter values *QP* (*QP* ranging from 0 to 51) resulting in a number of bitstreams with different *SourceBitrates*. The resultant bitstreams have been decoded and the obtained video R has been encoded again by an HEVC reference encoder in order to simulate a cascade pixel domain transcoder. Re-encoding of a single decoded video R has been performed with the use of all possible quantization parameters (*QP* ranging from 0 to 51) in order to simulate a wide range of requested bitrate changes. As a result, more than 2500 video bitstreams have been prepared.

The quality of each decoded video, both after the first encoding (video R) and after transcoding (video T), has been measured by the luminance PSNR metric. Also, the bitrates of each encoded stream have been gathered for further analysis.

The reference HEVC software has been configured according to the ''main_randomaccess'' test conditions [28], [29] with the in-loop deblocking and samples adaptive offset (SAO) filters turned ON. Such a scenario of video encoding is well recognized as a high-efficiency video compression scenario and corresponds to HEVC encoder configuration which is very commonly used in practice. The widely used *BlueSky*, *PedestrianArea*, *Riverbed*, *RushHour*, *Station*2, *Sunflower* full HD test sequences were taken as the test material (1920x1080 spatial resolution, 30 frames per second). In each case, only the first 200 frames from each sequence were encoded.

In order to determine the loss of image quality that results from cascaded video transcoding, ΔPSNR was calculated as the difference between the PSNR of transcoded video T and

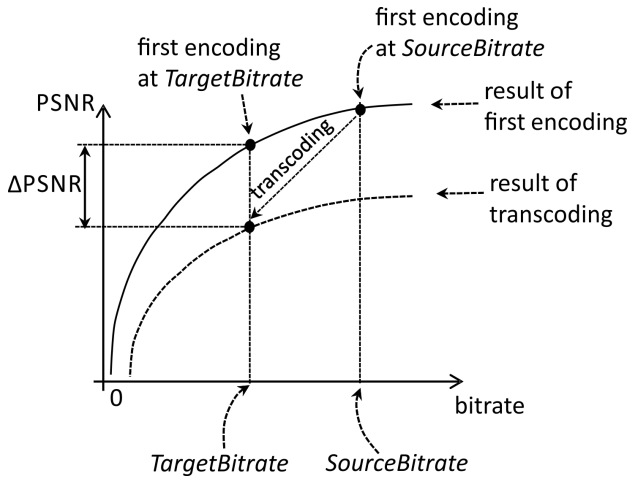


FIGURE 2. The way of determining ΔPSNR in a homogeneous HEVC transcoder.

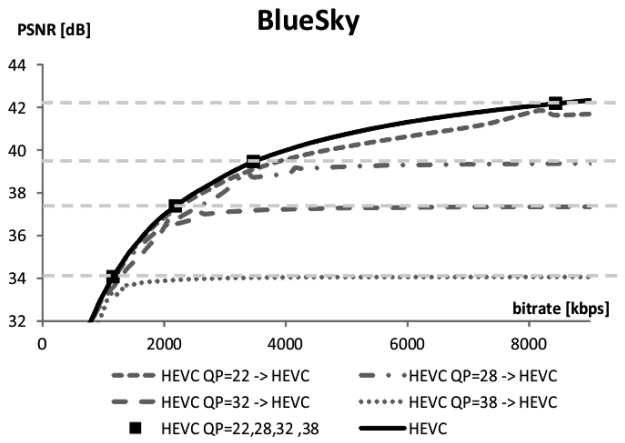


FIGURE 3. Exemplary results of video transcoding obtained for *BlueSky* test sequence.

the PSNR of encoded video C, both encoded at the same *TargetBitrate* (Fig. 1). The procedure of determining ΔPSNR is also illustrated in Fig. 2.

In case of each transcoding, the relationship between the *SourceBitrate* (i.e., bitrate of the first video encoding) and the *TargetBitrate* (i.e., final bitrate of transcoded data) was additionally analyzed (see Fig. 2). The obtained partial results were used to determine the ΔPSNR in the function of the transcoding ratio (i.e., *TargetBitrate*/*SourceBitrate* ratio) for each of the test sequences, and each value of source and target *QP*.

In the authors' opinion, the methodology used in this research correctly shows the results of video transcoding. This methodology has also been recently used by the experts of ISO/IEC MPEG in their new standardization activity on network distributed video transcoding.

V. VIDEO QUALITY IN HOMOGENEOUS HEVC TRANSCODING – RESULTS AND DISCUSSION

A. EXPERIMENTAL RESULTS – GENERAL DATA

As evidenced by the results obtained during our experiments, transcoding HEVC bitstreams leads to a noticeable degradation of the quality of video. Figure 3 presents exemplary but representative data of video transcoding that were obtained for the *BlueSky* test sequence. In this figure, the solid line represents the "PSNR – bitrate" curve which is a result of the first step of the experiment: direct encoding of an original test sequence with various *QP*s. A few selected values of $QP = 22, 28, 32, 38$ have been directly marked on the plot using rectangles. This line represents the quality of video C (from Fig. 1) encoded at different *TargetBitrates*.

The dotted lines give information about the results of transcoding of the source bitstream to new bitstreams with different rates. Each dotted line represents the "PSNR-bitrate" curve of the transcoded video produced based on a single selected source bitstream encoded at some *SourceBitrate*. Those source bitstreams were the bitstreams

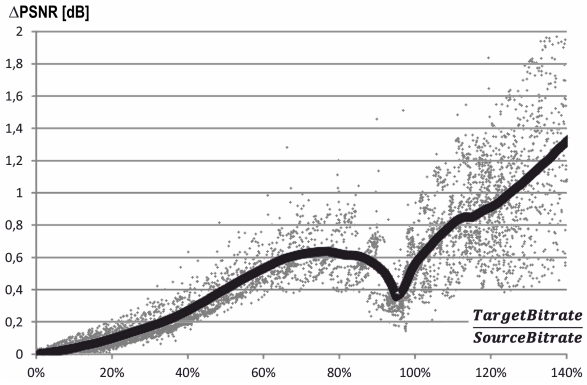


FIGURE 4. ΔPSNR (quality loss) with respect to transcoding ratio (i.e. *TargetBitrate*/*SourceBitrate* ratio). The result of averaging the partial data obtained for individual test sequences and different values of *SourceBitrate*.

obtained in the first step of encoding for the *QP* values of 22, 28, 32, 38.

The reasons why the quality of the transcoded video is always worse than the quality of the source video, have already been outlined in Section III.

It must be emphasized here that although the overall nature of changes of video quality follow the trends seen for the exemplary sequence (see Fig. 3), the exact results of transcoding also depend on the content of the video to some extent.

The general relationship of quality differences ΔPSNR (quality loss) in the function of the transcoding ratio - *TargetBitrate*/*SourceBitrate* is presented in Fig. 4. In this figure, each dot represents the result of a single transcoding, from a bitstream encoded at the *SourceBitrate* to a bitstream encoded at the *TargetBitrate*. The solid line is the result of averaging the results that are marked with dots. As it can be easily seen, the value of the *TargetBitrate*/*SourceBitrate* ratio affects the results to a large extent.

The main conclusion that can be drawn here is that the lower the value of the transcoding ratio (which translates into a bigger difference between *TargetBitrate* and

SourceBitrate values), the lower the loss of quality expressed by ΔPSNR, as compared to a direct encoding of original data at the *TargetBitrate*.

In the extreme case, in which the *TargetBitrate* is very low, while the *SourceBitrate* is significantly higher (so, the value of the ratio is low as well), transcoding will not lead to a significant degradation of video quality (understood as ΔPSNR). This can be explained by the fact that encoding video material at a very low *TargetBitrate* causes large quality losses. In such conditions, it is not relevant whether we use original video material or one previously slightly degraded by encoding at a relatively high *SourceBitrate*.

In the opposite case of high values of the *TargetBitrate* (which also means higher values of the transcoding ratio), the quality of the video material used (original or previously encoded at a rate similar to the *TargetBitrate*, *SourceBitrate*) makes a significant difference. This highlights a greater loss of video quality during transcoding with a high transcoding ratio, when compared to the scenario with a low transcoding ratio.

Summarizing, we can say that an average maximum PSNR decrease of 1.4dB was observed (in comparison to video encoded directly from original material, at the same *TargetBitrate*) for the range of transcoding ratios considered in the research. One has to remember, though, that this average maximum value also includes results for cases where the target *QP* was lower than the source *QP*, so the *TargetBitrate* was higher than the *SourceBitrate* – a scenario which rarely finds any practical justification. In cases which have the greatest practical use (that is $TargetBitrate/SourceBitrate < 100\%$), the average maximum decrease of PSNR is below 0.7 dB.

Another observation concerns the results of a moderate reduction of the bitrate in the transcoder. What is very interesting, there is an optimal point of transcoding, i.e. an optimal value of the transcoding ratio for which the loss of video quality finds its local minimum in this transcoding scenario. The averaged results (see Fig. 4) show that this point is $TargetBitrate/SourceBitrate = 95\%$ and the loss of video quality is about 0.35 dB for this point. A deeper analysis of this phenomenon will be the topic of subsection V-C.

Apart from the point that ''minimizes'' the loss of video quality (for a moderate reduction of the bitrate), there is also a point of transcoding which forms a local maximum of the loss of video quality. This is for the case of the transcoding ratio approximately equal to 75%, which gives the loss of video quality ΔPSNR = 0.63 dB. Consequently, a reduction of the bitrate by 25% (more or less) results in the highest loss of video quality, when compared to the reference direct encoding results.

Another obvious observation that arises from the analysis of the graph is that the ratios above 100%, for which the *TargetBitrate* is higher than the *SourceBitrate*, are, on average, characterized by a steadily increasing quality loss. The higher the ratio, the higher the loss. The explanation here is that for a higher *TargetBitrate*, transcoding is not able

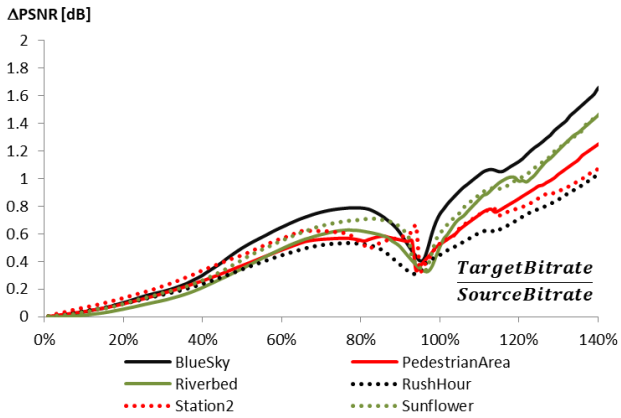


FIGURE 5. The results of averaging the partial data obtained for the considered values of *QP*. Results of individual test sequences.

to compete with direct encoding, since direct encoding of original data is able to preserve the details that are simply not presented in the encoded data (video *R* from Fig. 1) fed to the transcoding chain. Therefore, being unable to recover those details, the transcoder is not able to produce a video of similar quality as in the reference scenario.

B. EXPERIMENTAL RESULTS - MORE DETAILED DATA

The exact relationship between the expected ΔPSNR and the transcoding ratio depends on the content of a sequence to some extent but, in general, the overall properties of data remain unchanged. The results that were obtained for individual test sequences are presented in Fig. 5.

A deeper analysis of the presented relationships allowed us to conclude that higher losses of quality occur in more complex (in terms of image content) video sequences. In this case, it is more difficult to predict the content of image blocks which results in higher values of prediction error of image samples that undergo quantization. In the described scenario, the quantization of a prediction error signal causes a larger loss of video quality compared to sequences with a simple content. In the latter case, a smaller prediction error signal is obtained, and that results in smaller losses of information observed in the process of data quantization. It is also worth noting that for complex sequences, even small *QP* (high *SourceBitrate*) used for the first compression in the process, removes a significant amount of data, which are impossible to restore before the second encoding in the cascade transcoder.

The relationship between expected quality loss ΔPSNR and the transcoding ratio depends even stronger on the value of the *SourceBitrate* (i.e., on the value of *QP* used in the first stage of cascade transcoder) than on the content of a video itself. As it can be seen from exemplary results presented for the *BlueSky* sequence (see Fig. 6), the lower the *SourceBitrate* (the lowest bitrate for encoding with $QP = 38$ – round dots, the highest for $QP = 22$ – short dense dashes), the higher the expected losses of video quality when transcoding a video. This is particularly visible for higher values of the transcoding ratio. A more detailed analysis of this property revealed

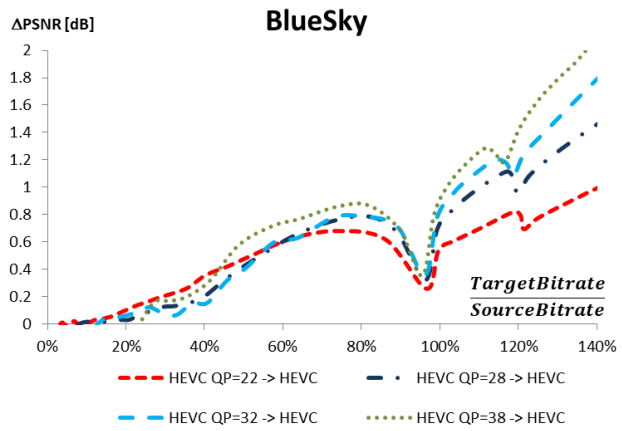


FIGURE 6. ΔPSNR with respect to transcoding ratio for four scenarios of encoding the original video: $QP = 22, 28, 32, 38$.

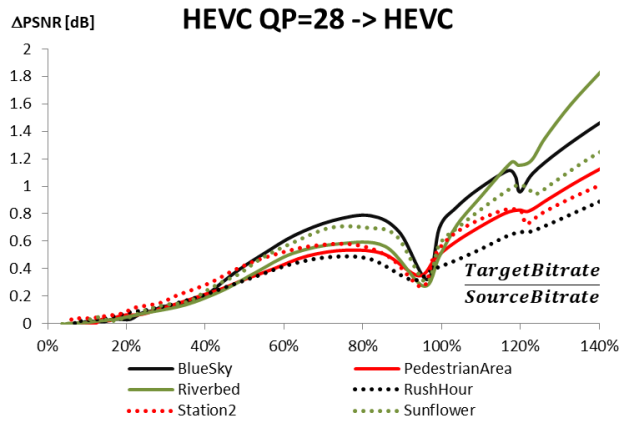


FIGURE 7. ΔPSNR with respect to transcoding ratio for individual sequences. Results for first step compression with $QP = 28$.

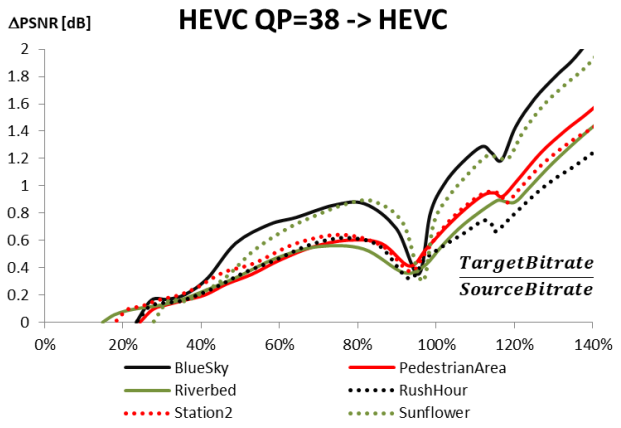


FIGURE 8. ΔPSNR with respect to transcoding ratio for individual sequences. Results for first step compression with $QP = 38$.

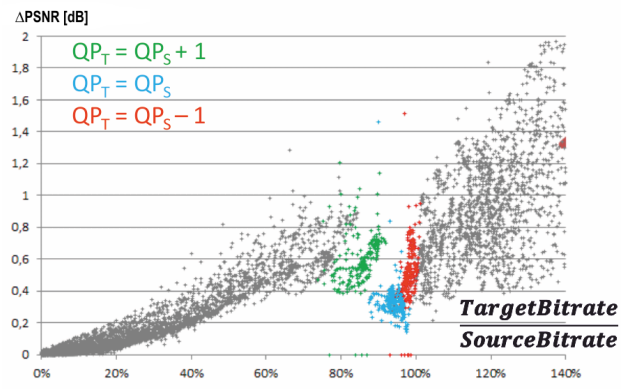


FIGURE 9. ΔPSNR with respect to transcoding ratio (i.e. *TargetBitrate*/*SourceBitrate* ratio). Figure presents partial data obtained for individual test sequences. With colors results for specific relations of QP_T and QP_S were additionally highlighted.

that underlying reasons can be found in the nature of the "PSNR-bitrate" curve (see solid line in Fig. 3). This curve is very steep for low values of bitrates (which corresponds to higher values of QP), and relatively flat for high bitrates (so, for low QP values). As a result, transcoding of low bitrate bitstreams will result in a greater change (loss) of video quality, than in the case of higher values of bitrates. Complementary to the results of Fig. 6 are data presented in Figs. 7 and 8, presenting the results obtained for individual test sequences for two strongly different values of QP: 28 and 38. The presented results confirm all the conclusions that were formulated before.

C. CONDITIONS FOR MINIMUM LOSS VIDEO TRANSCODING

As it was pointed out in section V-A, there exists a characteristic point of video transcoding which minimizes the loss of video quality for a scenario of a moderate reduction of bitrate in a video transcoder. According to the averaged experimental data, this point is $TargetBitrate/SourceBitrate = 95\%$. In order to investigate the causes of the existence of such a point, each portion of the partial data has been carefully analyzed in terms of settings of the encoder, which led to individual results. In particular, the relationship between the value of quantization parameter QP_S used for encoding the original video (that produces video R from Fig. 1) and the value of this parameter QP_T used later in the transcoder, was studied. The results of this analysis achieved for the range of the transcoding ratio from 75% to 100% were presented in Fig. 9.

As it turned out, the local minimum in quality loss is obtained for the case when encoding in the second step of transcoding uses the same QP value as the one used in the first step of encoding. The points closest to this minimum correspond to the cases when the second-step encoding was performed with the QP value higher or lower by 1 than the QP used in the first step. Partial results obtained for the various scenarios (mutual relationship of values of QP_T and QP_S) have been marked with appropriate distinctive colors in Fig. 9.

The important thing to note is that the bitrate ratio of 100% is obtained almost exclusively for the cases when encoding in the second step is performed with QP lower by one than the

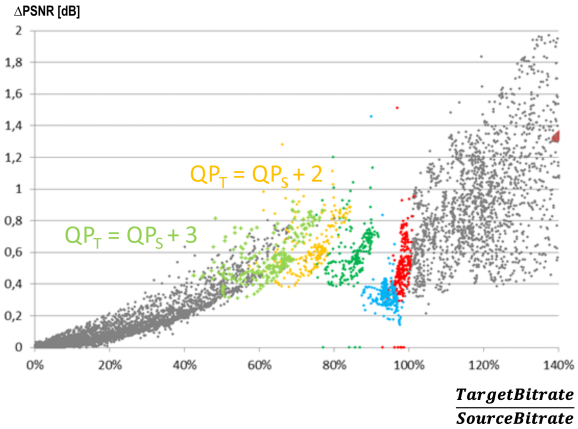


FIGURE 10. ΔPSNR with respect to transcoding ratio (i.e. *TargetBitrate*/*SourceBitrate* ratio). With green and yellow colors results for specific relations of QP_T and QP_S were additionally highlighted.

value used in the first step, while causing a loss of 0.5 dB in quality on average.

D. CONDITIONS FOR MAXIMUM LOSS VIDEO TRANSCODING

In the scenario of video transcoding when the ratio *TargetBitrate*/*SourceBitrate* is in the range of 50% to 100% one can easily see cases when quality losses are particularly high. As it can be observed in figures of this section, the maximum losses of video quality were reported for the *TargetBitrate*/*SourceBitrate* ratio approximately equal to 75%. Although it is very difficult to accurately point out the reasons of this phenomenon, we present our thoughts on this matter.

First of all, we noted that the results tend to cluster with respect to the difference between QP_S and QP_T. The maximal loss of transcoding is observed mainly for $QP_T = QP_S + 2$ and $QP_T = QP_S + 3$. This is an important conclusion of the research. The partial results obtained for these cases are shown in the Fig. 10.

Our thoughts on the origin of the presence of the maximum of quality loss is the following.

We observed that when the *TargetBitrate* varies between 80% to 120% of the *SourceBitrate*, the farther the *TargetBitrate* from the point of the minimal losses is, the greater the transcoding loss becomes. This seems to be true regardless of whether the *TargetBitrate* increases or decreases. The mechanism of this increase of loss stems from the misalignment of the quantization steps and does not seem to be surprising. Unfortunately, for *TargetBitrate* lower than 80% of the *SourceBitrate*, we observe an extreme of the quality loss in terms of ΔPSNR.

We understand that this is the result of the fact that the ΔPSNR metric is a superposition of two important factors. One of these factors is the size and overlap of the quantization steps. This mechanism is also used in the transcoder to decrease the bitrate. The other, however, is of a different nature.

For low *Target* to *Source* bitrate ratio, either the *TargetBitrate* itself needs to be low in general, or the *SourceBitrate* needs to be high.

For the first case, the quality of the reference bitstream (coded directly from the original sequence to the *TargetBitrate*) is low, and the difference between the quality of transcoded and reference bitstreams is low (the quality will not decrease much lower than the reference).

For the second case, for which the *TargetBitrate*/*SourceBitrate* is low due to a very high *SourceBitrate*, the points are close to the horizontal axis (i.e., the transcoding loss is low) because the source quality is very high and close to the quality of the original sequence. Then, the transcoding results are very much like the reference coding results, and the transcoding process is not much different (hence, it does not introduce significant loss) than the encoding of the original sequence.

Both of these phenomena cause the transcoding curve to drop (i.e., the transcoding loss is decreasing) after the *TargetBitrate*/*SourceBitrate* decrease below approximately 75%. To sum up, the following seems to be true for the points with bitrate ratio below 75%:

1) The *TargetBitrate* (numerator) is low. Then, both the reference and transcoded sequence are of similar, low quality and ΔPSNR is low.
2) The *SourceBitrate* is high. Then the source stream is of high quality, similar to original sequence. Transcoding in such a case produces similar quality to the quality of the reference encoded bitstream, therefore ΔPSNR is, again, low.

We did not investigate the reason why all these phenomena superimpose to reach the maximum loss at approximately 75% of bitrate ratio, but one important observation is to avoid transcoding with $QP_T = QP_S + 2$ and $QP_T = QP_S + 3$.

VI. VIDEO QUALITY IN HOMOGENEOUS HEVC TRANSCODING - FURTHER CONSIDERATIONS

The results presented in the previous section clearly show that when transcoding a video, the mutual relationship of the quantization parameters of source bitstream QP_S and target bitstream QP_T determines directly how big the loss of quality of a signal will be. Figure 9 presents the empirical data of video quality loss for cases when $QP_T = QP_S$ and $QP_T = QP_S \pm 1$ in a homogeneous HEVC transcoder. But at this point there are still open questions. First, what is the justification of the nature of obtained results? And secondly, how big would the loss of video quality be if the mutual relationship of values of QP_S and QP_T was different from the one already analyzed in section V-C? These issues will be the subject of a more detailed discussion in the following paragraphs.

A. THE NATURE OF OBTAINED RESULTS – ANALYSIS OF REASONS

In the context of the first question, a theoretical analysis has been performed to explore the impact of data quantization on the size of quantization error. This analysis has been carried

out assuming a special case of two consecutive stages of quantization. In the considered scenario, the first stage of data quantization is realized with parameter QP_S, while the second stage with parameter QP_T. The following special cases were considered in the analysis:

- Case A: $QP_T = QP_S$,
- Case B: $QP_T < QP_S$,
- Case C: $QP_T > QP_S$.

Conclusions from the analysis have been presented in illustrations of Fig. 11, from which the following general conclusions can be drawn.

The use of two identical quantizers in both stages (thus, $QP_T = QP_S$) will lead to a value of the final quantization error 'e' which will never be greater than the maximum error 'e_{max}' for the second quantizer (see Case A in the Fig. 11). This is because of the equal quantization step sizes used in both the first and second quantizer. Of course, this will be the case when the output value of the first quantizer is directly put onto the second quantizer (such a scenario was adopted in Fig. 11). However, in a real transcoder, the situation is somewhat different. The data that are quantized in a transcoder (with quantization parameter value QP_T) may be different from those quantized in the first stage of video encoding (that is realized with parameter QP_S), due to other steps of processing, like nonlinear filtering and prediction that will influence the final value of the sample. This change of value explains the non-zero loss of video quality for the $QP_S = QP_T$ scenario (as presented in Fig. 9). But the general theoretical conclusion highlighted above that the full compliance of the two quantization step sizes ($QP_S = QP_T$) gives the minimum error of video transcoding is still true, compared with other relationships of values of QP_S and QP_T.

In the case when the two quantizers are different from each other (i.e., QP_S differs from QP_T), the final quantization error 'e' can be bigger than the maximum error 'e_{max}' of the second quantizer. The discussed situation for the two separate cases in which $QP_S < QP_T$ or $QP_S > QP_T$ has been illustrated with Cases B and C in Fig. 11. The presented conclusion results from a different allocation of decision boundaries (on the axis of quantized values) in the first and second quantizer, which comes directly from different widths of quantization intervals. Thus, in these cases the resulting error of transcoding is generally greater than in the case when $QP_S = QP_T$.

B. RELATIONSHIP OF QP_T and QP_S AND ITS IMPACT ON ERRORS OF TRANSCODING

The illustrations from the previous subsection clearly present the mechanism of introducing transcoding errors in a video transcoder when QP_S differs from QP_T, but there is still the unsolved question about the magnitude of this error in practical situations. In order to answer the question of error magnitude for certain cases, the range of values of QP_S and QP_T were tested. This way, the influence of the width difference

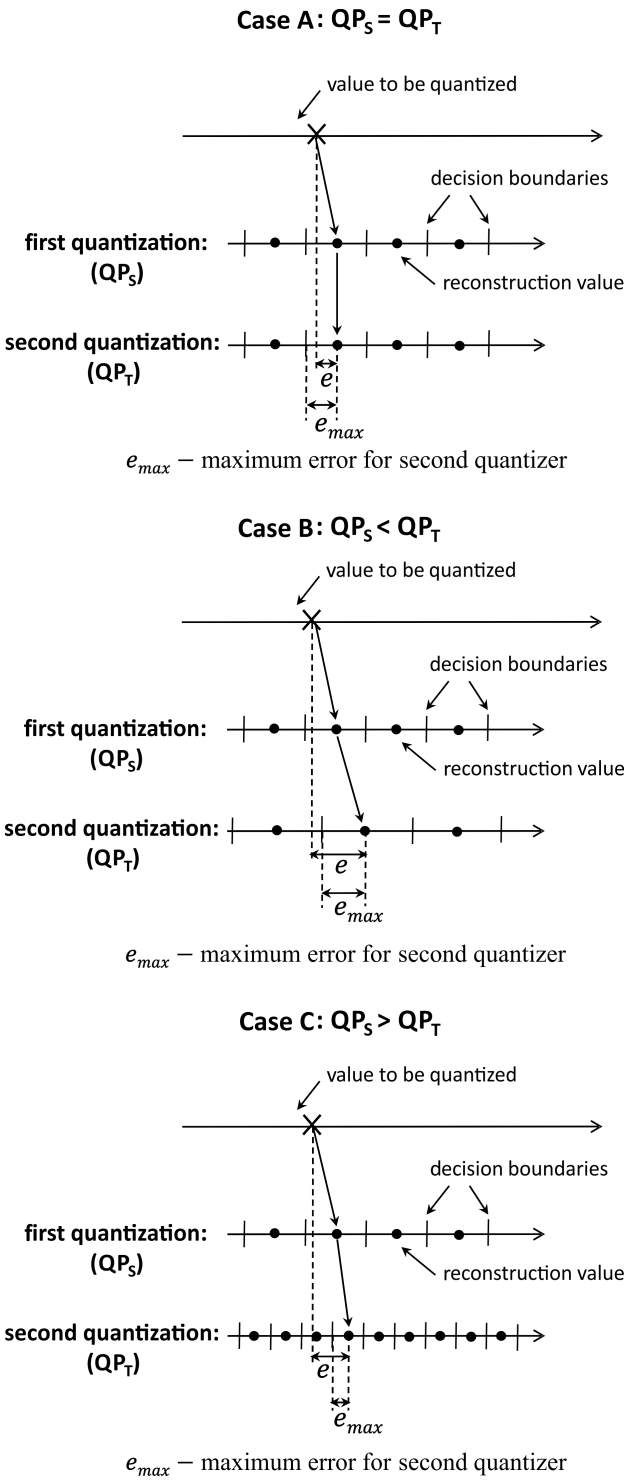


FIGURE 11. Impact of the relation of values of QP_S and QP_T on the size of quantization error. Analysis of three cases: Case A – $QP_S = QP_T$, Case B – $QP_S < QP_T$, Case C – $QP_S > QP_T$.

of the quantization step at different stages of encoding was tested in an additional experiment.

In this experiment, a uniform quantizer with a quantization step equal to $QSteps_S$ was used to quantize every possible transform coefficient value from an HEVC encoder.

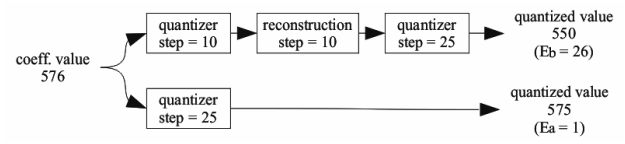


FIGURE 12. An example result of the experiment with varying quantization steps.

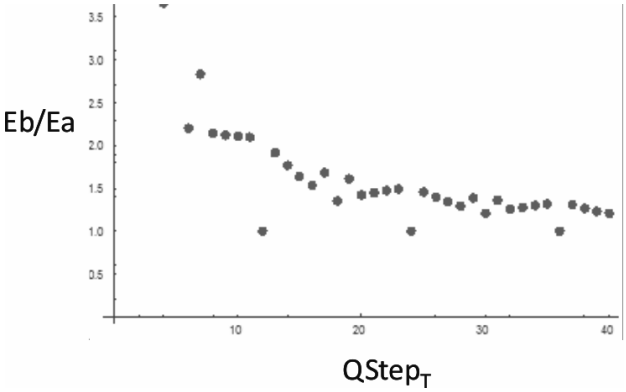


FIGURE 13. The error ratio as a function of $QStep_T$. $QStep_S$ was kept constant at the value of 12.

Then, the dequantized value was quantized once again using a quantization step equal to $QStep_T$. For reference purposes, the same initial value of the transform coefficient was directly quantized using the quantization step equal to $QStep_T$. For both cases, quantization error E was calculated – E_a and E_b respectively for those two cases. The example calculation is shown in Fig. 12.

The average values of errors E_a and E_b for the example case ($QStep_S = 10$ and $QStep_T = 20$) (averaged over all possible coefficient values) are 12 and 14.5, respectively. Therefore, it can be seen that transcoding introduces an additional error when compared to a case with direct compression with a given width of the quantization step, which agrees with the observations described in the previous chapters. The ratio of mean error values (that is Eb/Ea) is 1.2.

To examine the influence of the quantized step width change on the results, a different experiment was conducted, for which $QStep_S$ was kept constant at the value of 12, while $QStep_T$ was changed from 2 to 40 with a step of 1. The ratio of mean error value Eb/Ea is given as a function of $QStep_T$ in Fig. 13.

What is evident from this figure, is the following. The minimum error can be obtained only when the $QStep_T$ width is an integer multiple of the $QStep_S$ width. For such cases the error remains the same in both scenarios, therefore $Eb/Ea = 1$. For every other case, the scenario with requantization gives a higher average error, even when $QStep_T$ changes by 1 from the multiple of $QStep_S$. For such a case, the error can be over twice as high.

Another observation also confirms the previously presented results – the higher the $QStep_T$, the lower the Eb/Ea ratio, and therefore the requantization loss is smaller.

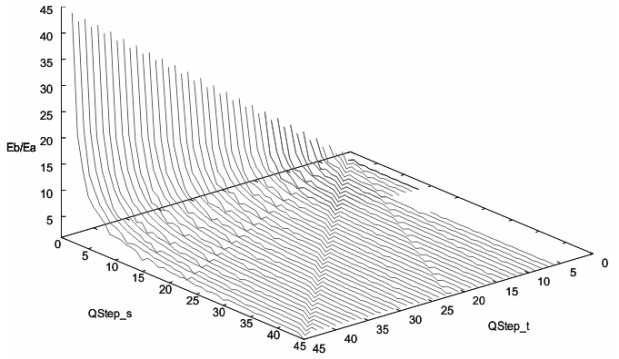


FIGURE 14. The error ratio as a function of $QStep_T$ and $QStep_S$.

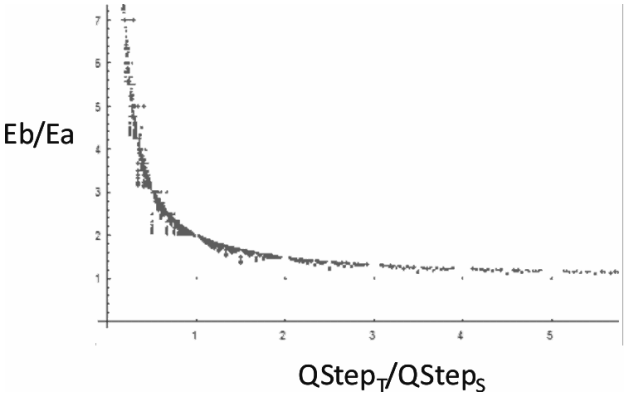


FIGURE 15. The error ratio as a function of $QStep_T/QStep_S$ ratio.



FIGURE 16. The quantization step limits overlap for different $QStep_T/QStep_S$ ratios.

In Fig. 14, the relationship between the Eb/Ea coefficient and both quantization steps, $QStep_S$ and $QStep_T$, is visualized.

The minimum values for $QStep_S = QStep_T$ are clearly noticeable in this figure, as well as a less noticeable minimum for $QStep_T/QStep_S = 1.5, 2.5, 3.5$ and so on.

In Fig. 15, the same data are plotted in 2D – again, the minimums mentioned above are clearly visible.

The reason for the minimum to occur for $QStep_T$ being a multiple of $QStep_S$ is obvious: the quantization step limits (decision boundaries) are overlapping for those cases, therefore the samples are always represented by the same quantized value, regardless of whether they underwent a single quantization with $QStep_T$ or two consecutive quantizations – first with $QStep_S$ and then with $QStep_T$. This is presented in Fig. 16.

The reason for the Eb/Ea ratio for $QStep_T = 2.5 \cdot QStep_S$ not being one is that only every other of the decision

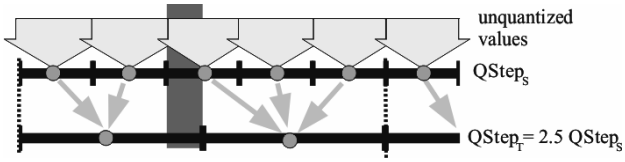


FIGURE 17. Requantization introduced additional error – marked in dark gray.

boundaries overlap. This means that two out of three steps will overlap entirely, but every third one will get requantized and will be assigned to a certain step of $QStep_T$, with a maximal quantization error equal to $0.5 \cdot QStep_S$, as it is shown in Fig. 17. The dark gray area marks the range of input values that are assigned to a different step in the case of re-quantization.

VII. FINAL REMARKS AND CONCLUSIONS

A quantitative analysis of the expected quality losses in a homogeneous HEVC transcoder was presented in the paper. The obtained results can be used as a reference for future research on video transcoder improvements. The results have shown that, although video transcoding always leads to some loss of video quality, there are transcoding conditions for which the loss of quality is minimal - with respect to the quality of video encoded from the original video at the same bitrate as the transcoded video. The minimum quality losses of $\Delta\text{PSNR} = 0.35$ dB are achieved with a moderate reduction of the bitrate, for the transcoding ratio $TargetBitrate/SourceBitrate = 95\%$. It corresponds to the case of using the same value of QP in the first and second video encodings. On the other hand, transcoding scenario in which $TargetBitrate/SourceBitrate = 75\%$ gives maximum loses of video quality. This scenario corresponds to a $QP_T = QP_S + 2$ configuration in most cases. However, it should be further noted that the exact relationship of ΔPSNR versus transcoding ratio depends significantly on the value of the *SourceBitrate*, as well as the content of a video sequence.

The results presented in this work, apart from their scientific value, are also of high practical importance. In common industrial practice, content providers prepare multiple versions of the same video material encoded at different bitrates. Currently, the most widely spread-out approach is to prepare video material at halved bitrates (e.g., 2Mbps, 1Mbps, 500kbps, 250kbps), and for small bitrate adjustments of those bitrates video transcoding is utilized. Our analysis clearly shows that such an approach is not always optimal, as the biggest quality losses may occur in the case of such small bitrate adjustments. Instead, based on the presented results, the optimal approach would be to prepare several closely distributed (in bitrates) video streams directly from original material, and to perform video transcoding in a strictly defined scenario of video encoding. The cases when it is worth to use video transcoding (instead of using multiple versions of the encoded video) were indicated in the Fig. 18.

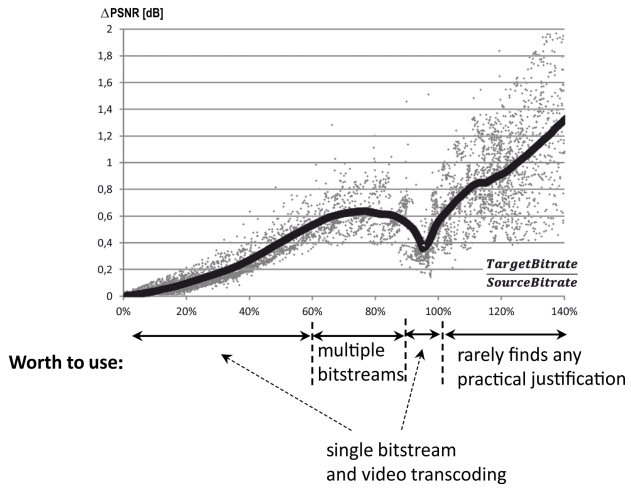


FIGURE 18. Ranges of values *TargetBitrate*/*SourceBitrate* when it is worth to use video transcoding instead of the solution with multiple bitstreams.

Due to the wide use of cascaded pixel domain video transcoders (CPDT) in industrial practice this technique of transcoding has been considered in the work. The subject of the study was the classical CPDT transcoder that performs the full encoding of a video material. According to the best knowledge of the authors, the obtained results will apply to all of the transcoders that are based on the idea of CPDT, including the state-of-the-art ones. As we suppose, every other type of CPDT transcoder will have the same limitations (those discussed in this work) and possibly, some additional ones, imposed by the method proposed by the authors. Thus, any CPDT transcoder will exhibit the same behavior as described in this paper.

REFERENCES

[1] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video. (MPEG-2)*, document ISO/IEC 13818-2 and ITU-T Rec. H.262, Nov. 1994.
[2] *Video Coding for Low Bit Rate Communication*, document ITU-T Rec. H.263, Aug. 2005.
[3] *Generic Coding of Audio-Visual Objects, Part 10: Advanced Video Coding*, document ISO/IEC 14496-10, Mar. 2006.
[4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[5] *High Efficiency Video Coding—MPEG-H Part 2*, document ISO/IEC IS 23008-2:2015, ITU-T Rec. H.265 Edition 3.0, 2015.
[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[7] A. Puri, X. Chen, and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Signal Process., Image Commun.*, vol. 19, no. 9, pp. 793–849, Oct. 2004.
[8] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—Including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.
[9] T. Grajek, J. Stankowski, K. Wegner, and M. Domański, "Video quality in AVC homogenous transcoding," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Dubrovnik, Croatia, May 2014, pp. 211–214.
[10] O. Alay and H. Stephansen, "The effect of multi-generation encoding in broadcast contribution on the end-user video quality," in *Proc. 19th Int. Packet Video Workshop (PV)*, Munich, Germany, May 2012, pp. 113–118.

[11] J. Stankowski, T. Grajek, K. Wegner, and M. Domański, "Video quality in multiple HEVC encoding-decoding cycles," in *Proc. 20th Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Bucharest, Romania, Jul. 2013, pp. 75–78.

[12] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: An overview," *IEEE Signal Process. Mag.*, vol. 20, no. 2, pp. 18–29, Mar. 2003.

[13] I. Ahmad, X. Wei, Y. Sun, and Y.-Q. Zhang, "Video transcoding: An overview of various techniques and research issues," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 793–804, Oct. 2005.

[14] J. De Cock, S. Notebaert, P. Lambert, and R. Van de Walle, "Requantization transcoding for H.264/AVC video coding," *Signal Process., Image Commun.*, vol. 25, no. 4, pp. 235–254, Apr. 2010.

[15] S. Notebaert, J. De Cock, K. Vermeirsch, P. Lambert, and R. Van de Walle, "Quantizer offset selection for improved requantization transcoding," *Signal Process., Image Commun.*, vol. 26, no. 3, pp. 117–129, Mar. 2011.

[16] S. Moiron, S. Faria, A. Navarro, V. Silva, and P. Assunção, "Video transcoding from H.264/AVC to MPEG-2 with reduced computational complexity," *Signal Process., Image Commun.*, vol. 24, no. 8, pp. 637–650, Sep. 2009.

[17] L. P. Van, J. De Cock, G. Van Wallendael, S. Van Leuven, R. Rodriguez-Sánchez, J. L. Martínez, P. Lambert, and R. Van de Walle, "Fast transrating for high efficiency video coding based on machine learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Melbourne, VIC, Australia, Sep. 2013, pp. 1573–1577.

[18] L. P. Van, J. De Cock, A. J. Diaz-Honrubia, G. Van Wallendael, S. Van Leuven, and R. Van de Walle, "Fast motion estimation for closed-loop HEVC transrating," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2492–2496.

[19] L. P. Van, J. De Praeter, G. Van Wallendael, S. Van Leuven, J. De Cock, and R. Van de Walle, "Efficient bit rate transcoding for high efficiency video coding," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 364–378, Mar. 2016.

[20] V. A. Nguyen and M. N. Do, "Efficient coding unit size selection for HEVC downsizing transcoding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Lisbon, Portugal, May 2015, pp. 1286–1289. doi: 10.1109/ISCAS.2015.7168876.

[21] L. P. Van, J. De Praeter, G. Van Wallendael, J. De Cock, and R. Van de Walle, "Performance analysis of machine learning for arbitrary downsizing of pre-encoded HEVC video," *IEEE Trans. Consum. Electron.*, vol. 61, no. 4, pp. 507–515, Nov. 2015. doi: 10.1109/TCE.2015.7389806.

[22] L. P. Van, J. De Praeter, G. Van Wallendael, J. De Cock, and R. Van de Walle, "Machine learning for arbitrary downsizing of pre-encoded video in HEVC," in *Proc. IEEE Int. Conf. Consumer Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2015, pp. 406–407. doi: 10.1109/ICCE.2015.7066464.

[23] M. Sung, M. Kim, M. Kim, and W. W. Ro, "Accelerating HEVC transcoder by exploiting decoded quadtree," in *Proc. 18th IEEE Int. Symp. Consumer Electron. (ISCE)*, JeJu Island, South Korea, Jun. 2014, pp. 1–2. doi: 10.1109/ISCE.2014.6884329.

[24] Z. Lin, Q. Zhang, K. Chen, J. Sun, and Z. Guo, "Efficient arbitrary ratio downscale transcoding for HEVC," in *Proc. 30th Vis. Commun. Image Process. (VCIP)*, Chengdu, China, Nov. 2016, pp. 1–4. doi: 10.1109/VCIP.2016.7805465.

[25] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Fast HEVC to SCC transcoder by early CU partitioning termination and decision tree-based flexible mode decision for intra-frame coding," *IEEE Access*, vol. 7, pp. 8773–8788, 2019. doi: 10.1109/ACCESS.2018.2890720.

[26] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, "Sample adaptive offset in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[27] K. Yang, S. Wan, Y. Gong, H. R. Wu, and Y. Feng, "Perceptual based SAO rate-distortion optimization method with a simplified JND model for H.265/HEVC," *Signal Process., Image Commun.*, vol. 31, pp. 10–24, Feb. 2015.

[28] HM 15.0 Reference Software. *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*. Accessed: Oct. 2018. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn

[29] F. Bossen, *Common Test Conditions and Software Reference Configurations*, document JCTVC-J1100, Joint Collaborative Team on Video Coding (JCT-VC) of ITUT SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Stockholm, Sweden, Jul. 2012.

Journal of Electronic Imaging

JElectronicImaging.org

Fast mode selection in the high-efficiency video coding intravideo encoder based on statistics of modes

Krzysztof Wegner
Damian Karwowski
Jakub Stankowski
Tomasz Grajek
Krzysztof Klimaszewski
Olgierd Stankiewicz

SPIE. IS&T imaging.org

Krzysztof Wegner, Damian Karwowski, Jakub Stankowski, Tomasz Grajek, Krzysztof Klimaszewski, Olgierd Stankiewicz, "Fast mode selection in the high-efficiency video coding intravideo encoder based on statistics of modes," *J. Electron. Imaging* **27**(4), 043051 (2018), doi: 10.1117/1.JEI.27.4.043051.

Journal of Electronic Imaging 27(4), 043051 (Jul/Aug 2018)

Fast mode selection in the high-efficiency video coding intravideo encoder based on statistics of modes

Krzysztof Wegner, Damian Karwowski,* Jakub Stankowski, Tomasz Grajek, Krzysztof Klimaszewski, and Olgierd Stankiewicz
Poznan University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznan, Poland

Abstract. A method of fast coding mode selection for high-efficiency video coding (HEVC) intravideo coding to exclude from the testing procedure coding modes that are statistically rarely chosen by the HEVC encoder is presented. To achieve the best compromise between computational complexity and compression performance, detailed conditional statistics of modes are used in the method. The proposed method enables (1) fast partitioning of each coding tree unit into coding unit, prediction unit (PU), and transform unit blocks of a proper size and (2) faster selection of an appropriate intrapredictor for each PU block. Due to the proposed method, it is possible to significantly scale down the computational complexity of the encoder. Simulation results indicate that, in the case of 1920×1080 sequences, there is a threefold and twofold reduction in the encoding time, while the BD-rate increases by 4% and 1.5%, respectively, compared with HEVC test model version 16.9 of the reference software. Considering the principle of the proposed method, it is evident that it can complement other proposals concerning fast mode selection in the HEVC video encoder. *© 2018 SPIE and IS&T* [DOI: 10.1117/1.JEI.27.4.043051]

Keywords: video compression; high-efficiency video coding; video encoder control; fast mode selection in high-efficiency video coding.

Paper 170945 received Nov. 1, 2017; accepted for publication Jul. 24, 2018; published online Aug. 23, 2018.

1 Introduction

In recent years, there has been a growing interest in services related to the transmission and storage of high and ultrahigh definition videos. This has given rise to the need for stronger video compression than ever before to further reduce transmission and storage costs. Research on this subject has resulted in the expansion of new high-efficiency video coding (HEVC) technology. This technology was described in international standards of ISO/IEC and ITU-T in 2013,[1] and it doubles the rate-distortion (R-D) performance compared with the widely used MPEG-4 advanced video coding (AVC)/H.264 video encoder.[2]

Similar to older video compression technologies (e.g., MPEG-4 AVC/H.264),[3] HEVC is based on a hybrid scheme of coding image blocks, which uses intra- and interframe prediction coding together with transform coding of residual data.[4] To more efficiently compress high and ultrahigh definition videos, many improvements have been made to the existing methods, and new coding tools have been introduced in HEVC.

In comparison with the previous AVC standard, a major change in the HEVC is the replacement of the macroblock structure (of size 16×16) with the new coding tree unit (CTU) structure. CTUs are four times larger than macroblocks and consist of three types of specialized blocks: the coding unit (CU), the prediction unit (PU), and the transform unit (TU).

According to the general concept of CTU blocks, a block can contain a single CU block (size 64×64) or may be divided into multiple, smaller CU blocks (32×32, 16×16, or 8×8). This division is implemented using the concept of quad-tree partitioning.[5] At the CU level, a decision is made about the coding mode of the block (intra- or interframe coding) and the degree of data quantization.

Predictive coding of image samples is conducted using the selected method at the level of PU blocks. Compared with other video coding standards, HEVC offers a wider range of (PU) block sizes [from 64×64 to 4×4 for intracoding and from 64×64 to 8×4 (4×8) for intercoding]. Additionally, the PU blocks can be either square or rectangular in shape, and it is possible to achieve an asymmetric division for rectangular blocks, although nonsquare blocks are only allowed in intercoding. With regard to intraframe prediction, the number of directional predictors has been significantly increased (compared with AVC). Moreover, image samples are predicted with greater accuracy (due to samples filtering and better interpolation filters for reference frames).

Both transform coding and entropy coding of residual data are executed in TU blocks. Each CU can be independently divided into a number of TU blocks to exploit quad-tree partitioning. Sizes 32×32, 16×16, 8×8, and 4×4 of TU blocks are all possible, meaning that more available solutions are offered than in the older standards.

All modifications made in HEVC offer incomparably more possibilities of encoding image blocks. A thorough verification of all the possible ways to encode an image block results in substantial computational complexity. The application of such an approach, especially in the case of high and ultrahigh definition video, is impossible in practice. It is therefore necessary to find solutions that optimize the operation of the HEVC encoder. Indeed, such works have been conducted since the beginning of the development of a new HEVC standard. Due to the differences between the new HEVC technology and the older compression

*Address all correspondence to: Damian Karwowski, E-mail: dkarwow@multimedia.edu.pl

1017-9909/2018/$25.00 © 2018 SPIE and IS&T

techniques (e.g., AVC), the works primarily focused on the development of new methods for fast coding mode selection, which concern both intra[6–22] and inter[23–31] HEVC video encoders. This paper is devoted to the fast selection of intracoding modes in the HEVC encoder.[32]

Among the methods of fast intracoding known within the literature, those presented in Refs. 13 and 14 are of special importance due to their practical application in the HEVC test model (HM) reference software of the HEVC codec.[33] The basis of these methods is the selection of coding modes for blocks when using the RD optimization that is only performed on a limited (and not full) set of compression modes. According to the methods, successive sizes of CUs/PUs are verified (starting with the largest CUs/PUs). For each size of PU within a CU, the sum of absolute transformed differences cost is evaluated for each of the available compression modes [rough mode decision (RMD) mechanism].[13] The information on compression modes that were used in previously coded neighboring blocks of the image are also used [most probable mode (MPM) mechanism].[13] The RMD and MPM mechanisms allow us to quickly determine the limited set of compression modes for which a full RD optimization is then performed (assuming the maximum possible TU size). Finally, the optimal TU size is selected for the best compression mode (also using RD optimization).

Other methods described within the literature can be categorized into two groups: the first group includes methods that improve the solutions used directly in the HM reference software, i.e., methods that allow a rapid selection of the prediction mode for a CU/PU of a given size. In the new proposals of this group, the limited set of compression modes (for which full RD optimization is performed) is merely determined in a different way. Typically, the choice is done through the analysis of texture complexity and texture directionality (intra- and/or interframe compression),[8,10–12,15,16,19] and the use of the spatial and temporal correlation of modes (intra and inter).[9,10,15] The second group of methods proposes not only a rapid mode selection for a given CU/PU but also a selective CU/PU search of a certain size. This is performed based on various attributes of the image and/or encoding process.[10,25,20–22] These attributes provide information about texture complexity and coding modes used in neighboring image blocks,[10] information about the cost of using modes in adjacent blocks (e.g., Ref. 20), data on the statistical relationships of the sizes of neighboring CU/PU blocks within an image,[21] and information about the relationship between the costs of modes obtained for the CU/PU of a given size.[20,22]

The methods developed recently reduce the encoding time by 20% to 60% when compared with the complexity of the original HM reference software; however, there is some loss of compression performance.

This paper presents a different perspective on the problem of fast mode selection in HEVC intra. It introduces the authors' method of accelerating the HEVC encoder. Instead of checking the performance of successive compression modes in a large number of variants (sizes) of CU, PU, and TU (which is, computationally, extremely complex), this method proposes that the least probable modes be omitted. They can be identified after determining the statistics of modes in the encoder. Thus, it is possible to significantly accelerate the video encoder at the risk of losing some compression efficiency. By exploiting the detailed statistics of mode selection, which is dependent on several factors [CU, PU, and TU sizes, spatial image resolution, and the value of the quantization parameter (QP)], it is possible to minimize losses related to compression efficiency when following such an approach.

The amount of detail of the mode statistics will obviously determine the efficiency of the method. In fact, previous studies have presented statistical analyses of modes in the HEVC, e.g., Refs. 15, 34, and 35. However, the level of detail of those statistics is limited. In the case of intracompression, the statistics indicate the distribution of CU sizes for selected small sets of 3 or 4 values of QP (rather than a wide set of values) and only for single test sequences[15] or present the general use of individual intrapredictors (from a set of 35 predictors), without considering the size of PU as well as the value of QP parameters.[34] Thus, it is clear that no particularly detailed conditional statistics of HEVC modes have yet been published. Therefore, for the purpose of the presented speedup method that omits checking of some modes, highly accurate statistics of encoding modes have been gathered. A portion of the gathered data is presented in the following sections. The statistics enable rapid partitioning of CTUs into CUs, PUs, and TUs, as well as make it possible to bypass some intrapredictors during mode selection in the encoder. These statistics extend the more general data that have been previously published. Contrary to approaches mentioned within the literature, which primarily rely on a much simpler data analysis, our approach is based on a significantly broader set of statistics. The large amount of data gathered in the initial stage of research allows us to calculate statistics that can more precisely describe the interdependency of the modes selected by the encoder by providing conditional probabilities of encoder mode selection. Therefore, we are able to perform a significantly more complex conditional probability analysis and improve the mode selection process in HEVC, compared with other methods from the literature.

The method is implemented in a way that enables us to establish the cutoff "th" threshold value for the testing modes. Only the compression modes (understood as a way of CTU partitioning together with intrapredictors for PUs) that are used in more than "th"% of the image area are examined further by the encoder in the RD optimization procedure. No calculations are performed for the other coding modes.

The proposed method is examined in terms of the time of video encoding and efficiency of video compression (measured relative to results obtained for the original HM software of HEVC) to assess its suitability in fast mode selection in HEVC.

The rest of the paper is organized as follows. Section 2 presents the methodology of research. Section 3 outlines the subset of the mode statistics that have been gathered during the research process. In Sec. 4, the proposed fast mode selection in the HEVC encoder is introduced and the experimental results are discussed. Section 5 concludes the paper and offers final remarks.

2 Methodology of Research

The first step of the research is to perform a complex conditional probability analysis of the modes to determine

the conditional statistics of the coding modes in the HEVC video encoder. The statistics are based on the percentage of image samples that have been encoded using a given compression mode. As previously known, the details of encoder implementation (*inter alia*, the method of compression mode selection) determine the statistics and affect the efficiency of video compression. The intention of the authors is to develop the statistics of the modes for the efficient implementation of the HEVC encoder, which does not contain essential simplifications in the mechanism of coding mode selection. Therefore, the statistics of the modes were gathered using the reference software HM 16.9 of the HEVC encoder, in which the existing simplifications of mode selection were previously switched off by the authors (the RMD and MPM mechanisms were switched off). This allowed us to gather data for a very HEVC scenario.

The statistics of the modes were studied using a comprehensive set of test video sequences, with six different spatial resolutions of the images: 2560×1600 (class A: four sequences—Traffic, PeopleOnStreet, Nebuta, and SteamLocomotive), 1920×1080 (class B: five sequences—Kimono1, ParkScene, Cactus, BQTerrace, and BasketballDrive), 832×480 (class C: four sequences—RaceHorses, BQMall, PartyScene, and BasketballDrill), 416×240 (class D: four sequences—RaceHorses, BQSquare, BlowingBubbles, and BasketballPass), 1280×720 (class E: three sequences—FourPeople, Johnny, and KristenAndSara), and 832×480, 1024×768, and 1280×720 (class F: four sequences—BasketballDrillText, ChinaSpeed, SlideEditing, and SlideShow). To obtain results with a wide range of practical applications, the full range of QP values ($QP = 0$ to 51, as defined in the HEVC standard) was considered in this research.

The resulting statistics of the modes were then used in the second step of the work, which involved introducing necessary modifications to the HM 16.9 encoder to allow for the use of the statistics during the encoding process. Thus, a modified version of the HEVC encoder was obtained. In this modified version, rarely chosen coding modes are not tested by the encoder and the number of omitted modes can be controlled by a certain threshold value to create a trade-off between speedup and quality loss. At this point, two different encoders were used as the anchor to thoroughly test the solution proposed in the paper.

Further details of the methodology are presented in the following sections of the paper.

3 Statistics of Coding Modes in the HEVC Intraencoder

3.1 *Details of Experiment Methodology*

The statistics of the compression modes were researched to evaluate the following relationships:

- Percentage $P\{\mathrm{CU}_{\mathrm{size\_CU}}\}$ of image samples that are associated with CU blocks of size size_CU. All the possible CU sizes ($\mathrm{size\_CU} = 64 \times 64$, 32×32, 16×16, and 8×8) have been considered in this study.
- Percentage $P\{\mathrm{intra\_pred}/\mathrm{CU}_{\mathrm{size\_CU}}\}$ of image samples that are part of the image for which a specific intrapredictor has been used (35 available intrapredictors, so "intra_pred" = 0 to 34), while assuming the size size_CU of the CU is 64×64, 32×32, 16×16, and 8×8.
- Percentage $P\{\mathrm{TU}_{\mathrm{size\_TU}}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{intra\_pred}\}$ of image samples that are associated with parts of the image for which the size size_TU of TU was used ($\mathrm{size\_TU} = 32 \times 32$, 16×16, 8×8, and 4×4) for a given size size_CU of CU ($\mathrm{size\_CU} = 64 \times 64$, 32×32, 16×16, and 8×8) and the number of intrapredictors used ("intra_pred" = 0 to 34) for the considered part of the image.

Preliminary studies have shown that, in the case of sequences of natural scene images, the statistics of the modes are strongly dependent on the QP value, as well as the spatial resolution of the images. For this reason, relevant experiments were performed for the entire possible range of QP values ($QP = 0$ to 51) and for the six classes of test video sequences: classes A, B, C, D, E, and F. As indicated above, the individual classes vary with regard to the spatial resolution "img_res" of the images. The sequences used in the experiments are recommended by the ISO/IEC and ITU-T for research on video compression techniques.[36] During the tests, the following conditional statistics of the coding modes were recorded (the meaning of individual probabilities was explained earlier):

- $P\{\mathrm{CU}_{\mathrm{size\_CU}}/\mathrm{QP} \wedge \mathrm{img\_res}\}$,
- $P\{\mathrm{intra\_pred}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}$, and
- $P\{\mathrm{TU}_{\mathrm{size\_TU}}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{intra\_pred} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}$.

As can easily be calculated, 1248, 43,680, and 141,960 values of probabilities were collected in the three mentioned groups of statistics, respectively. In the last group, it is important to remember that a TU cannot be larger than the size of the CU itself. This portion of the experiment was performed using the HM 16.9 software of HEVC, with RMD and MPM speedup tools switched off.[13,14] The goal was to collect statistics from the encoder, in which the encoder does not skip a significant portion of modes during the mode selection procedure in advance. The rest of the encoders' settings were configured in accordance with the recommendations of MPEG and VCEG experts (known as "common test conditions—All Intra").[36] The adopted test scenario, thus, provides a mode analysis of video compression conditions oriented toward achieving very high coding efficiency in case of intracompression.

The method for measuring the statistics was a direct result of the way in which the HM software tests the individual coding modes. Initially, the division of CTU into large CU blocks was considered. Subsequently, the division of CTU into a larger number of smaller CUs was tested. Each CU block (of $\mathrm{size\_CU} = 64 \times 64$, 32×32, 16×16, and 8×8), as well as 4×4 image blocks, was encoded using one of each of 35 intrapredictors, taking into account different sizes size_TU of TU (size_TU from 32×32 to 4×4). The final coding mode is selected on the basis of the RD costs of the tested modes.

At the lowest level of the presented procedure of coding mode selection, the image samples encoded using TU units of a given size were counted, and the size of the CU block

and the intrapredictor used were recorded. This provided the basis for determining the percentage of samples that corresponds to the probability

$$P\{\mathrm{TU}_{\mathrm{size\_TU}}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{intra\_pred} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}.$$

Other conditional statistics were calculated on the basis of the following mathematical relationships:

- $P\{\mathrm{intra\_pred}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\} = \sum_{\mathrm{TU}_{\mathrm{size\_TU}}} P\{\mathrm{TU}_{\mathrm{size\_TU}}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{intra\_pred} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}$,
- $P\{\mathrm{CU}_{\mathrm{size\_CU}}/\mathrm{QP}\wedge\mathrm{img\_res}\} = \sum_{\mathrm{intra\_pred}} P\{\mathrm{intra\_pred}/\mathrm{CU}_{\mathrm{size\_CU}}\wedge\mathrm{QP}\wedge\mathrm{img\_res}\}$.

3.2 *Experimental Results*

Because of the assumed methodology of the experiments (see Sec. 3.1), the obtained set of statistics of the coding modes is substantial. Although the method proposed in this paper actually uses the full set of the evaluated statistics, the presentation of all results of the research is simply impossible in this paper. Therefore, only a portion of the results, along with the most important conclusions of this study, is presented here. Due to the expected areas of application of the proposed method (mainly the reduction of encoding time of high-resolution video sequences), only the data obtained for class B are presented as an example. The average results of all classes of sequences are also presented within this paper.

3.2.1 *Results of statistics $P\{CU_{size\_CU}/QP \wedge img\_res\}$*

While the statistics of CU sizes can be found in many publications, e.g., Refs. 15 and 35, such an analysis is typically conducted using certain assumptions, e.g., a limited set of considered values of QP or a small number of test sequences. Therefore, for consistency of all the results presented in this paper, the authors' statistics of CU sizes are presented in this section. The results, as a function of the QP value, are presented in Fig. 1. When analyzing the partial results, the following general conclusions can be drawn.

For low values of QP, a high (above 60%) percentage of the image area uses the smallest CUs (8×8). This percentage decreases as QP increases, and, in the case of a high QP (44 or more), <10% of the image area uses the smallest CUs. Thus, as QP increases, so does the participation of larger CUs (sizes 16×16, 32×32, and 64×64).

The spatial resolution of a test sequence affects the results. In the case of a lower resolution of images, a relatively high percentage of 8×8 CU blocks is observed in a wider range of QP values. For the 1920×1080 image resolution (class B), 8×8 blocks cover more than half of the image area for QP < 22. In the case of the 416×240 resolution (class F), the same is observed for a range of QP < 34. A more precise nature of the statistics is presented in the graphs below.

A small QP preserves the fine details of the encoded images (e.g., edges in images). Encoding such content using large blocks would negatively affect the efficiency of the intraprediction process. Therefore, the encoder usually selects small 8×8 blocks in the areas with fine details.

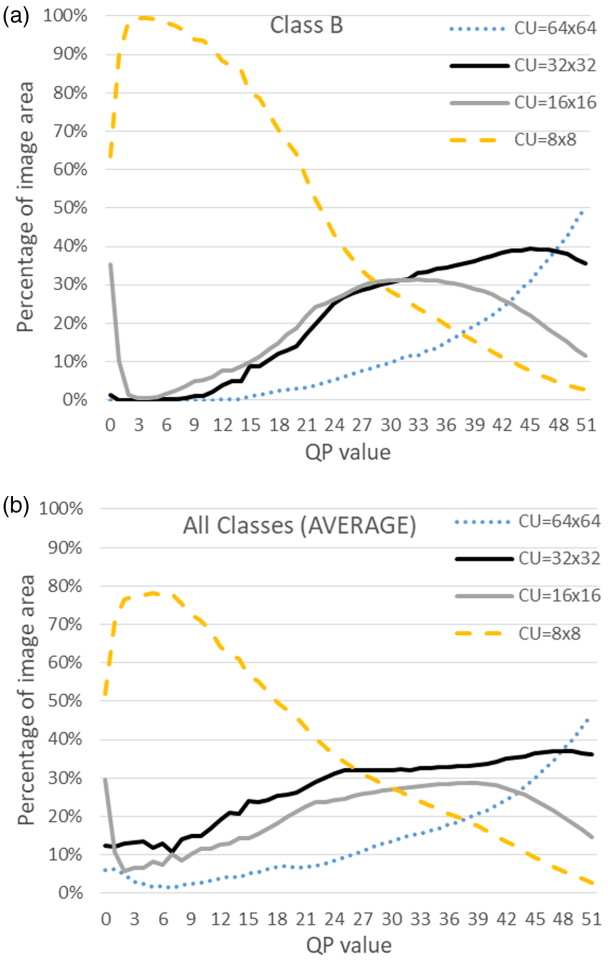


Fig. 1 (a) Statistics of CU sizes achieved for class B and (b) averaged results of all classes (A, B, C, D, E, and F).

Analogously, in the case of large QP, the fine image details are practically eliminated by the quantizer. As a result, there are more smooth areas in the encoded images. These areas can be well predicted through the predictive coding of image samples in blocks larger than 8×8 points. Image division into smaller blocks would, in general, decrease the coding efficiency because of the greater (than in the case of image division into large blocks) amount of control data that must be sent to the decoder.

3.2.2 *Results of statistics $P\{TU_{size\_TU}/CU_{size\_CU} \wedge QP \wedge img\_res\}$*

The following main conclusions can be drawn from this part of the research.

In the HEVC standard, the TU blocks cannot be larger than their respective CUs. Therefore, the statistics of CU size in practice determine the majority of the statistics of TU size. Due to a high percentage of 8×8 CUs when QP is low, a high percentage of 4×4 and 8×8 TUs is observed in the image for this scenario, since 8×8 CUs can be encoded only by the use of 4×4 or 8×8 TUs. This is particularly evident in the case of high-resolution sequences. For the range of QP < 20, the 4×4 TU units cover 20%

to 40% of the image area, while for the 8×8 TU, they cover 20% to 50%.

Since the contribution of larger CUs (16×16, 32×32, and 64×64) increases as QP increases, there is a growing possibility of choosing larger than 4×4 and 8×8 TUs. According to the experimental results, 16×16 and 32×32 TU blocks cover more than 65% of the image area when QP > 40. The nature of the quantized residual data obtained in this case (numerous miniscule or zero values) does not justify the division of large CU blocks into a high number of very small TU blocks. This would not improve the efficiency of the same transform coding as more small TU blocks create a larger amount of control data (tree of splitting the CU into TUs), which must also be sent to the decoder.

An important finding of the study is that for a given size of CU/PU, the same size TU is most often used, or the CU/PU is encoded in four equal-sized TUs. Thus, the division of CU/PU into more than four TUs is relatively rare. It must be emphasized that although both QP and spatial image resolution affect the detailed statistics, the majority of TUs have the size of CU/PU or a quarter of its size. More detailed results are displayed in the graphs of Fig. 2.

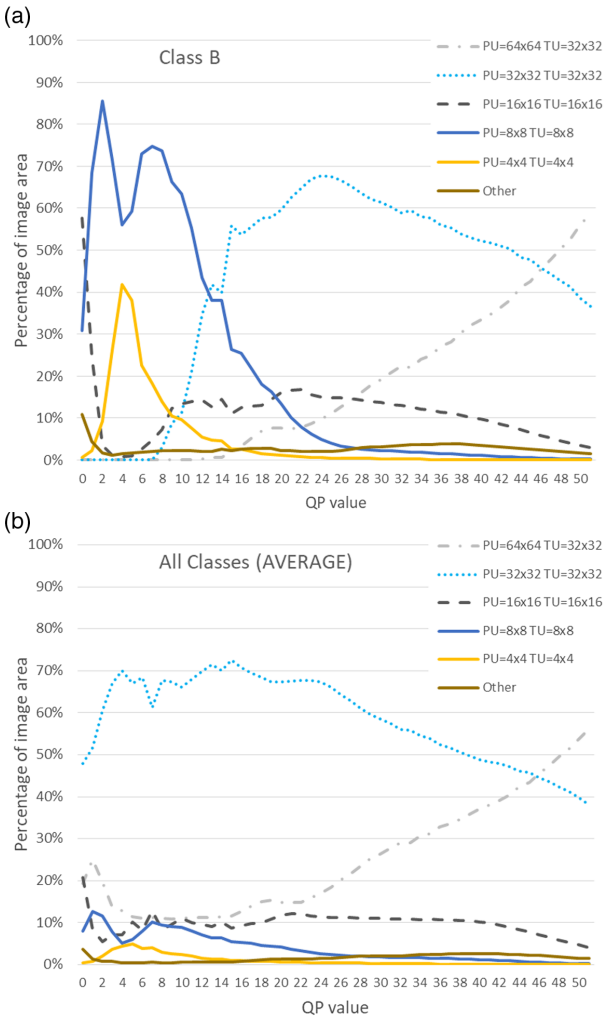


Fig. 2 (a) Statistics of TU and CU sizes achieved for class B and (b) averaged results of all classes (A, B, C, D, E, and F). Curve named "other" sums up all remaining possibilities of PU and TU divisions.

3.2.3 *Results of statistics P{intra_pred / CU_{size_CU} ∧ QP}*

The analysis of the partial results leads to the following general conclusions regarding the intrapredictor. While the intrapredictor number appears to be the most content-dependent data, several general observations can be advantageously used to reduce computational complexity.

Statistically, the use of 0-planar and 1-DC predictors is the most probable. These predictors cover 25% and 27% of the image area, respectively (average for all test sequences, tested values of QP and different sizes of CU/PU). It is also worth noting that the use of 10-horizontal, 26-vertical, and 9-near horizontal predictors (4%, 8%, and 3%, respectively) is noticeable. In the remaining the cases, the predictors outside the previously stated group of five predictors were used, where the probability of using each of them is relatively low. However, these remaining predictors cover between 30% and 40% of the total image area. Since smooth areas dominate in images of natural scenes, there is a low percentage of the use of directional predictors (2 to 34

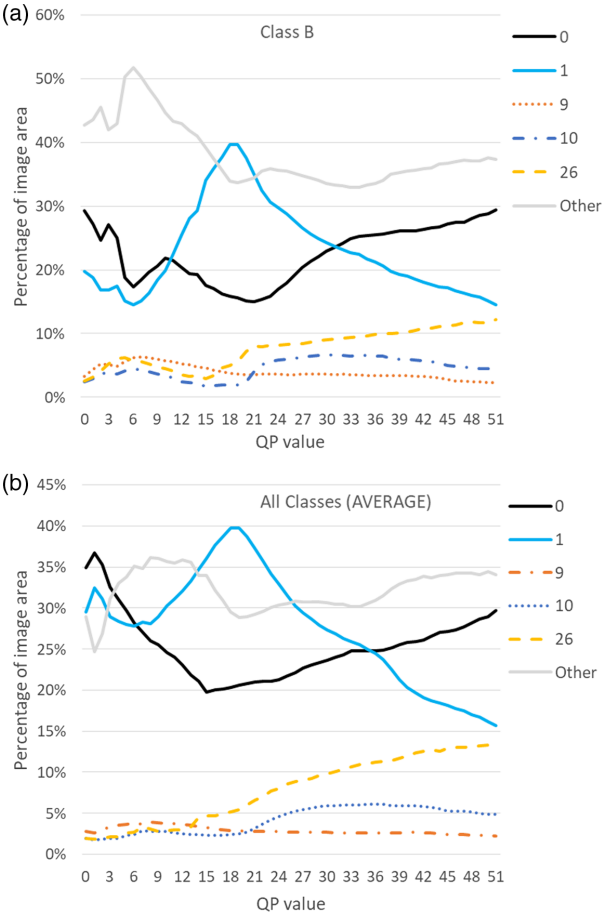


Fig. 3 (a) Statistics of intrapredictors for class B and (b) average results for all classes (A, B, C, D, E, and F). Curve named "other" sums up all remaining intrapredictors, which are not listed in the legend.

predictors) and an especially high percentage of the use of 0 and 1 predictors.

Both the size of CU/PU and the value of QP have a significant impact on the statistics of predictors; therefore, an increase in CU/PU size increases the total percentage of the use of 0 and 1 predictors. During this research, the total rate of utilization of these predictors was as follows: 60% (for PU 64×64), 55% (for PU 32×32), 40% (for PU 16×16), 35% (for PU 8×8), and 30% (for PU 4×4). For the rest of the image area, directional predictors were used. The quoted percentage of predictor usage results from the fact that large blocks of CU/PU are mainly used in the smooth parts of the image (i.e., without fine details), which can be predicted without the use of directional predictors (numbers 2 to 34). The inverse situation occurs in the case of small CU/PU blocks, which are commonly used in the parts of the image with more details. As a result, the degree of the use of directional predictors increases in smaller CU/PU blocks. For the selected predictor usage (in particular, for predictors of numbers 0, 1, and 26), a strong influence of QP was observed.

Predictor usage is also dependent on the spatial resolution of the images. In particular, this applies to predictors of numbers 0, 1, 10, and 26. Based on the detailed results, it is rather difficult to draw general conclusions. The exact statistics of predictor usage are presented in Fig. 3.

4 Fast Mode Selection in the HEVC Intraencoder

4.1 *Details of the Proposed Fast Mode Selection Method*

All the conditional statistics recorded during the tests were applied to the HM 16.9 software of the HEVC encoder, resulting in the modified HM software. The knowledge of the statistics in the encoder significantly reduces the number of tested coding modes by excluding the modes that are rarely chosen in the RD optimization procedure. In the modified software, the coding modes that have a probability of being selected by the encoder (measured as the percentage of image samples that are encoded with the coding mode), lower than a user-defined threshold "th," are not tested during coding. Early exclusion of modes in the developed method is realized at three levels:

1. At the level of partitioning of CTU into CU blocks, on the basis of the values of conditional probabilities $P\{\mathrm{CU}_{\mathrm{size\_CU}}/\mathrm{QP} \wedge \mathrm{img\_res}\}$ of CU sizes.
2. At the level of selecting the intrapredictor in PU blocks, on the basis of the values of conditional probabilities $P\{\mathrm{intra\_pred}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}$ of the intrapredictor.
3. At the level of CU partitioning into TU blocks, based on the values of conditional probabilities $P\{\mathrm{TU}_{\mathrm{size\_TU}}/\mathrm{CU}_{\mathrm{size\_CU}} \wedge \mathrm{intra\_pred} \wedge \mathrm{QP} \wedge \mathrm{img\_res}\}$.

In practice, the developed statistics of modes, together with the value of threshold "th," control the procedure of omitting the whole branches on the tree of possible ways of encoding CTU in the encoder (see Fig. 4.). Excluding some parts of this tree from testing may significantly reduce the video encoding time. The time of video encoding, as well as compression performance, depends directly on the number of skipped modes, and this number is affected by the value of "th" and the statistics of the modes.

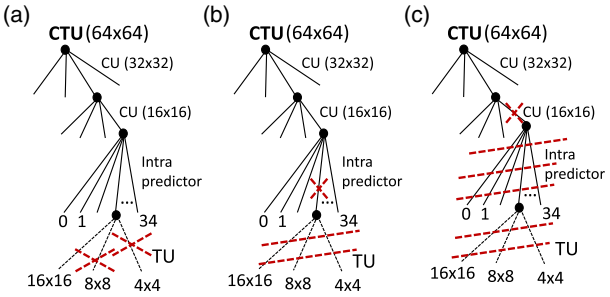


Fig. 4 Three levels of coding modes excluding: (a) TU size level, (b) intraprediction direction level, and (c) CU size level.

In the modified HM software, the order in which the coding modes are tested does not differ from the one used in the original HM implementation, as can be seen in Fig. 5. Select modes are skipped in the software by program instructions, which compare the values of the mode statistics with the value of "th." Only those coding modes for which the probability of being used is not lower than "th" are subjected to further analysis in the encoder.

4.2 *Parameters of the Proposed Method—Details of Experiment Methodology*

The modified HM 16.9 software that realizes the developed method of fast mode selection in the encoder was tested based on the following parameters:

1. total video encoding time and
2. compression performance of the video encoder.

These parameters were referenced to those used for the HM 16.9 encoder, in which RMD and MPM tools are switched on. Thus, the original HM 16.9 encoder implementation (with RMD and MPM tools) provides a reference for the results of our proposal, which are presented in this section.

The performance of the two encoders (modified and the original HM) was evaluated using six classes of test sequences, mentioned in the "common test conditions" recommendation[36] of ISO/IEC MPEG and ITU-T VCEG expert groups: classes: A, B, C, D, E, and F—the same that were used in the research on the statistics of the modes. To prove the generality of the proposed solution, this part of the research was also conducted using an additional set of 1920×1080 video sequences: BlueSky, PedestrianArea, RiverBed, RushHour, Station2, SunFlower, and Tractor. These sequences are also recommended by MPEG and VCEG experts. At this point, the goal was to confirm the performance of the solution for sequences that were not used in determining the statistics of the modes. Both the configuration of the encoders and the methodology of the experiments were the same and fully compatible with the "common test condition—All Intra" recommendation.[36] Since the parameters (complexity and compression performance) of the modified encoder depend directly on the values of the "th" threshold (expressed as a percentage), the proposed modified encoder was tested for a comprehensive set of 24 values of "th": 0, 0.05%, 0.1%, 0.15%, 0.2%, 0.25%, 0.3%, 0.35%, 0.4%, 0.45%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 4%, 6%, 8%, 10%, and 20%. In each case, the total

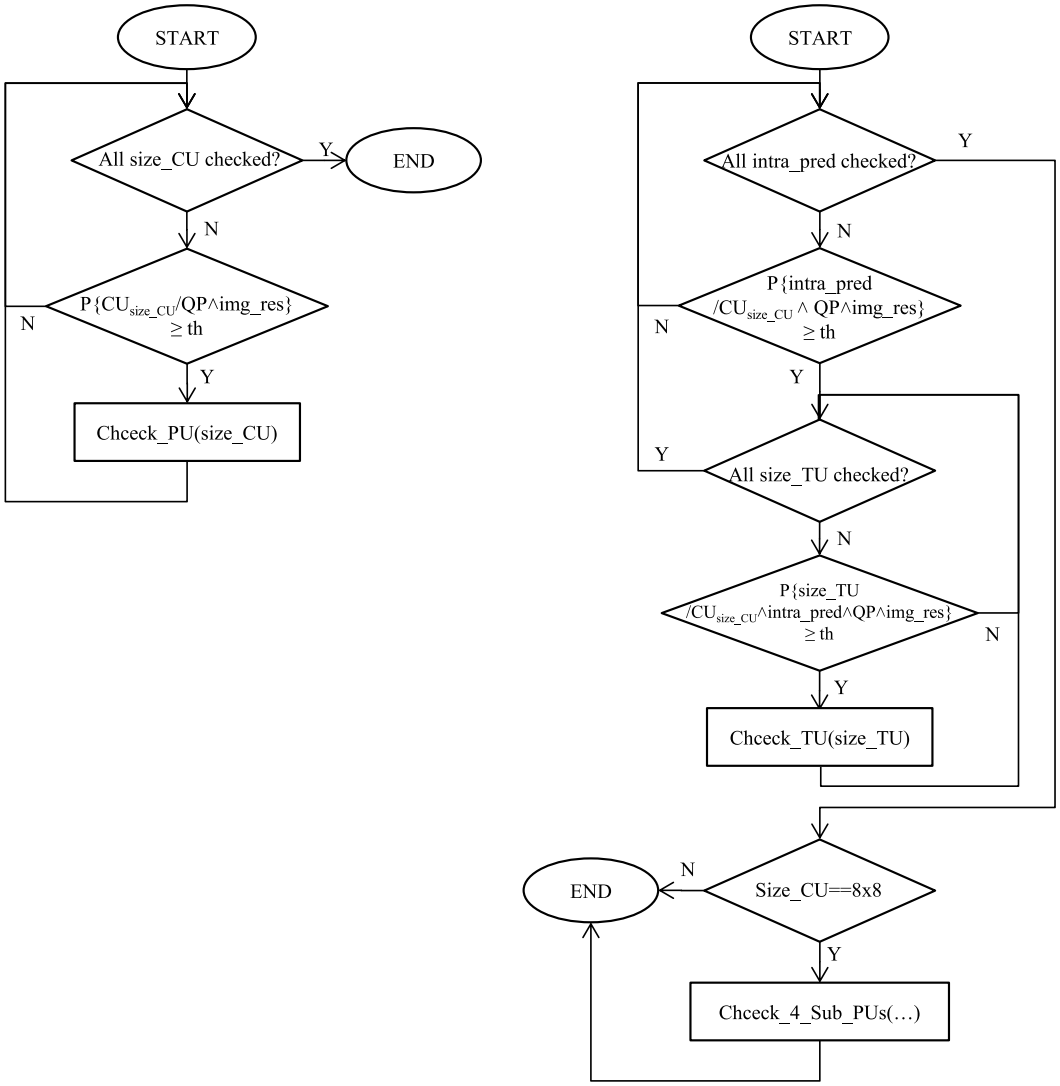


Fig. 5 Procedure of testing the coding modes in the proposed solution. Sizes of CU/PU and TU as well as the intrapredictors whose probability of using is less that the "th" thresholds are skipped by the encoder.

encoding time, the size of the encoded data stream, and the peak signal-to-noise ratio image quality measure were tested.

The partial results obtained for the proposed modified encoder and the original encoder were compared. As a result, the following attributes of the modified encoder have been determined:

- Reduction of the video encoding time (denoted as "encoding time reduction") when compared with the encoding time of the original HM encoder. This is expressed as a percentage and calculated using the following formula:

$$\text{Encoding time reduction} = \left(1 - \frac{\text{encoding time}_{\text{modified encoder}}}{\text{encoding time}_{\text{original encoder}}}\right) \bullet 100\%.$$

On the basis of the partial results achieved for four different values of QP (QP = 22, 27, 32, and 37), the average encoding time reduction value was calculated for each of the test sequences separately.

- Expressed as the BD-rate,[37] the percentage by which the encoded video data increased the used bitrate, when compared with results of the original encoder (under the same objective quality of images in the two encoders). On the basis of the individual results obtained for QP = 22, 27, 32, and 37, the average BD-rate increase was calculated for each of the test sequences separately.

In Sec. 4.3, the paper presents more detailed results of the experiments.

4.3 *Experimental Results*

4.3.1 *Partial results obtained for individual test sequences*

The results presented in Figs. 6 and 7 reveal to what extent the use of detailed conditional statistics of the modes

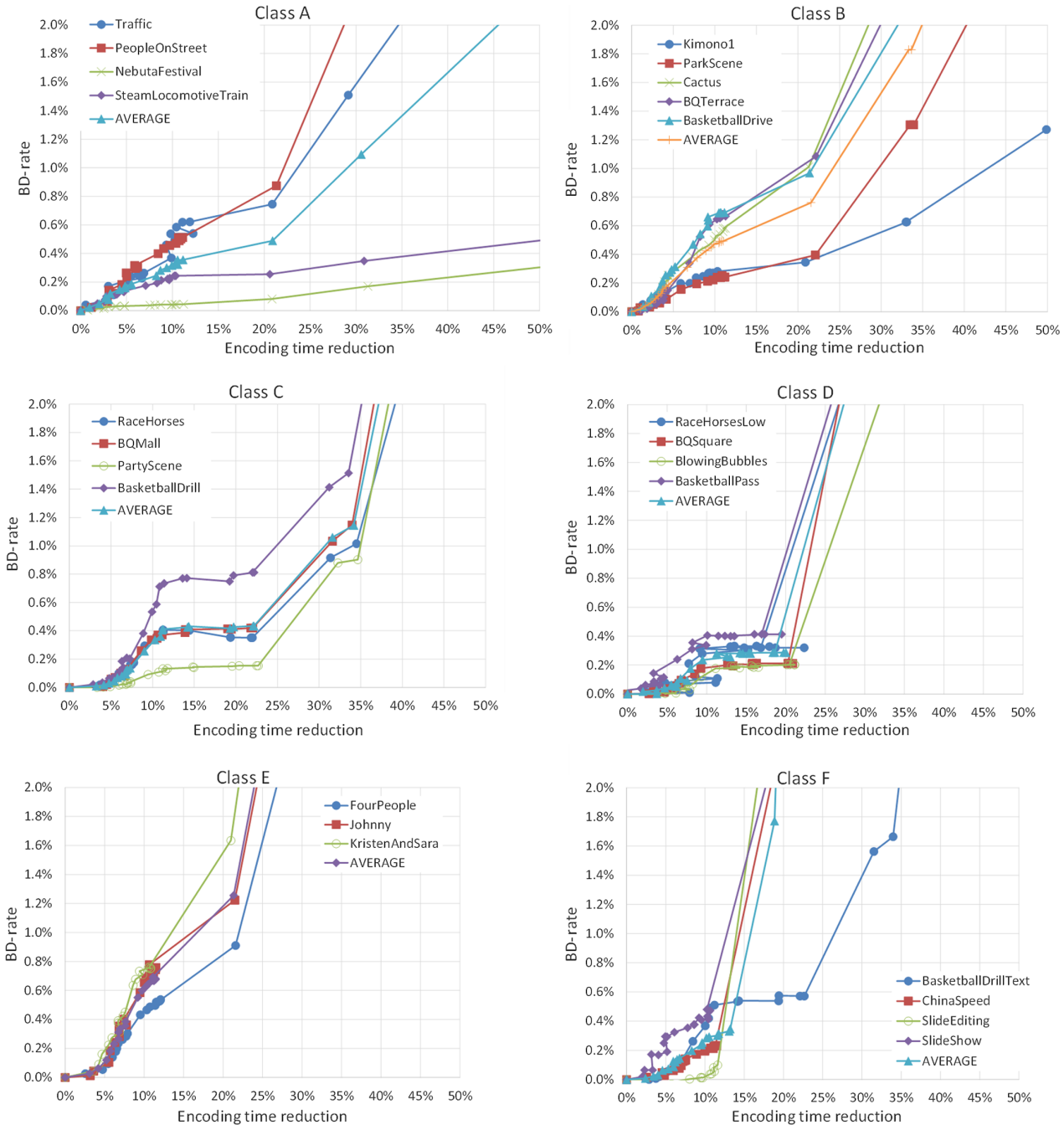


Fig. 6 Parameters of the proposed encoder in comparison with the original HM software. Results for sequences classes A, B, C, D, E, and F. The lines on the graphs connect the actual measuring points that were obtained during the experiments. Some of these points are not visible on the graphs (due to the presentation of the results for the range of encoding time reduction from 0% to 50%).

(in the procedure of modes selection) affects the complexity and performance of video encoding. To honestly judge the performance of the proposed encoder, the results obtained for each class of test sequences are independently presented. Although the relationship between the values of encoding time reduction and BD-rate depends on the content of the test sequence in general, the following main conclusions can be drawn.

Compared with the original HM software, the proposed encoder expedites the partitioning of the image into CU, PU, and TU. As a result, a large number of variants of CU/PU and TU are not considered in the encoder testing mode, significantly accelerating the video encoding. This conclusion has been confirmed by the authors in the course of a comparative analysis of a set of sizes of CU/PU and TU that were tested in the proposed and original encoders.

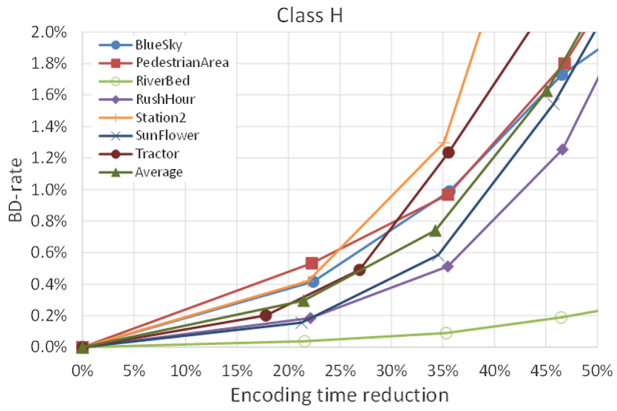


Fig. 7 Parameters of the proposed encoder in comparison with the original HM software for class H of test sequences. The lines on the graph connect the actual measuring points that were obtained during the experiments. Some of these points are not visible on the graph (due to the presentation of the results for the range of encoding time reduction from 0% to 50%).

However, for the considered PU blocks (of sizes 64×64 to 4×4), the sole information about statistics of the predictors' usage in the PU blocks would not allow the observable reduction in the number of the intrapredictors that are checked in the proposed encoder. This results from the fact that the probabilities of using individual predictors are relatively uniform (this is not the case for only 4 predictors of a set of 35); therefore, it is not feasible to exclude a high number of predictors from testing based solely on their statistical usage. This conclusion has been confirmed by the results presented in this work. Therefore, in the proposed method, the quick selection of the intrapredictor (one of 35) in a PU is supported by the RMD and MPM tools, which are present in the original software. In this way, the combined use of knowledge about the statistics of the predictors' usage and the RMD and MPM themselves enables one to test a limited small set of fewer than 10 preselected predictors.

When the values of encoding time reduction fell within the range of 0% to 20%, comparable results were obtained for individual classes of test sequences. The loss of compression efficiency (expressed as BD-rate) that occurs does not usually exceed 0.6%. However, the results obtained for individual classes differ quite significantly for the values of encoding time reduction higher than 30%.

As can be observed, the best results were obtained for classes A, B, C, and H of the test sequences. For these classes, with the BD-rate increase below 2%, the method was able to accelerate the encoder by 45%, 35%, 37%, and 47%, respectively.

Results of great importance were obtained for class H—the class with sequences that were not used during statistics gathering. The results for this class indicate whether the method is universally applicable to any sequence. The positive results obtained for this class confirm the hypothesis presented within this thesis.

4.3.2 *Influence of the "th" parameter on the results—discussion*

The value of the "th" parameter determines how efficient the previously described method can be. This value helps balance the trade-off between the encoding time reduction and the loss of encoding efficiency (usually expressed by a slight increase of the output bitrate). The larger the "th" value, the larger the time reduction, at the cost of increased bitrate and lower quality. By adjusting the "th" value, it is possible to set a desired operating point for our method. The time reduction necessary for the encoding process in this method inevitably reduces encoding efficiency; however, an acceptable value of "th" can usually be determined. This acceptable value is different for each class of sequences, depending on their properties (such as resolution, amount of motion in the sequence, or texture complexity). In our research, we tested different sequences and divided them into classes depending on their basic properties, as defined in Ref. 36.

For class A, we found that a reasonable "th" value is in the range of 1% to 1.5%, which reduces encoding time, on average, 30% to 50% at the expense of a BD-rate loss between 1.1% and 2.3%, respectively. For class B, a reasonable "th" value is in the range of 1.5% to 2.5%, which reduces encoding time, on average, 33% to 50% at the expense of a BD-rate loss between 1.8% and 4%, respectively. For class E, the "th" value in the range of 1.5% to 2% gives on average the encoding time reduction of 31% to 54% at the expense of a BD-rate loss between 4% and 8.9%, respectively. For class F, the "th" value in the range of 1.5% to 2% gives on average the encoding time reduction of 23% to 40% at the expense of a BD-rate loss between 9.4% and 25.2%, respectively. For this class, the highest losses are observed, especially for purely artificial sequences. Thus, it can be concluded that the statistical properties of the modes selected when encoding the artificial sequences does significantly differ from the statistics for natural scenes.

We observed that, for the remaining classes (C and D), we could increase the "th" value and still achieve similar encoding time reductions, as specified below. For class C, the "th" value in the range of 2.5% to 4% gives on average the encoding time reduction of 34% to 59% at the expense of a BD-rate loss between 1.1% and 8.4%, respectively. For class D, the "th" value in the range of 2.5% to 4% gives on average the encoding time reduction of 18% to 47% at the expense of a BD-rate loss between 0.3% and 5.9%, respectively.

The aforementioned values of parameter are not guaranteed to achieve the best results for every sequence since, during our experiments, we found single sequences in classes A and B that exhibited smaller quality losses for similar encoding time reductions. For those sequences, it is advised to use higher values of "th," within a similar range as for sequences from classes C and D.

The class F results were significantly less consistent. The class F results are very much varied, since this class includes artificial sequences that exhibit significantly different properties than the natural sequences with the addition of artificial content.

A thorough analysis of the results leads to the following additional conclusions. Better results were obtained for the sequences with a blurred content. For such sequences, the result of selecting a suboptimal coding mode has less influence on compression results than for sequences with a large amount of details and sharp edges.

For images with small details and edges, rare modes may be necessary to encode the content efficiently. As mentioned earlier, this is not the case for blurred sequences or those that contain large uniform areas. In addition, it can be assumed that the lower the resolution, the greater the losses—this is due to the image of a lower resolution having more details within one block.

Nonetheless, the presented method is able, for most of the sequences from all considered classes, to reduce the encoding time by approximately one third, at the expense of <2% of BD-rate loss (Table 1).

Table 1 Parameters of the proposed solution—detailed results for individual test sequences for selected values of "th" parameter.

	Sequence	BD-rate	Encoding time reduction	BD-rate	Encoding time reduction
		th = 1%		th = 1.5%	
Class A	Traffic	1.51%	29.15%	3.22%	48.22%
	PeopleOnStreet	2.35%	30.96%	5.09%	50.12%
	NebutaFestival	0.17%	31.29%	0.32%	51.86%
	Steam-LocomotiveTrain	0.35%	30.87%	0.49%	50.21%
	Average	1.09%	30.57%	2.28%	50.10%
		th = 1.5%		th = 2.5%	
Class B	Kimono1	0.63%	33.09%	1.27%	49.89%
	ParkScene	1.30%	33.49%	3.05%	49.78%
	Cactus	2.66%	33.24%	5.53%	49.93%
	BQTerrace	2.44%	33.54%	5.62%	50.50%
	BasketballDrive	2.12%	33.23%	4.67%	49.00%
	Average	1.83%	33.32%	4.03%	49.82%
		th = 2.5%		th = 4%	
Class C	RaceHorses	1.01%	34.49%	6.33%	59.47%
	BQMall	1.15%	33.96%	9.40%	59.24%
	PartyScene	0.90%	34.64%	8.46%	59.98%
	BasketballDrill	1.51%	33.54%	9.45%	58.95%
	Average	1.14%	34.16%	8.41%	59.41%
		th = 2.5%		th = 4%	
Class D	RaceHorsesLow	0.32%	16.84%	5.05%	44.52%
	BQSquare	0.21%	20.28%	8.07%	48.64%
	BlowingBubbles	0.20%	20.46%	4.71%	48.90%
	BasketballPass	0.41%	16.95%	5.71%	46.23%
	Average	0.29%	18.63%	5.89%	47.07%
		th = 1.5%		th = 2%	
Class E	FourPeople	2.98%	31.40%	6.98%	54.84%
	Johnny	3.91%	31.00%	8.61%	54.98%
	KristenAndSara	5.24%	30.53%	11.18%	54.19%
	Average	4.04%	30.98%	8.92%	54.67%
		th = 1.5%		th = 2%	
Class F	BasketballDrill-Text	0.57%	22.65%	1.56%	31.51%
	ChinaSpeed	0.23%	11.32%	2.58%	20.66%
	SlideEditing	22.07%	29.61%	60.40%	52.18%
	SlideShow	14.71%	30.27%	36.20%	55.15%
	Average	9.40%	23.46%	25.19%	39.87%

4.4 *Parameters of the Proposed Method with Respect to Other Methods*

To more thoroughly assess the potential of the proposed solution, its parameters were also compared with the performance of other solutions presented within the literature. The results of this comparison are shown in Table 2. Although for a specific degree of encoding time reduction, the loss of compression performance is higher for the proposed method (in comparison with other methods described in the literature), and the amount of computations performed in the proposed encoders is adjustable and can be controlled. In the case of high-resolution sequences (class B), it was possible to control the value of encoding time reduction for a wide range (up to 80%). The greater the degree of reduction of the encoding time, the greater the loss of compression efficiency. There is, therefore, the possibility to scale the computational complexity of the proposed encoders which, in the authors' opinion, is a highly desirable functionality. Thus, the method differentiates itself from other methods that have previously been proposed within the literature. The performance of the proposed solution was documented for a broad range of "th" threshold values and demonstrates a possible reduction in the complexity of the HEVC encoder against compression efficiency loss. However, it must be emphasized that this research considers only the details of the method that is used to reduce the number of considered coding modes for the purpose of speeding up a video encoder. Other aspects of the solution, such as how to determine the value of "th," are beyond the scope of this work and should thus be considered for future research.

5 Conclusions and Final Remarks

The idea of speeding up the video encoder by reducing the number of encoding modes evaluated during the encoding

Table 2 Performance of the proposed method and performance of other methods presented in the literature. Performance of the encoders evaluated in relation to the original HM. Letters next to numeric values show for which class of sequences the result was obtained.

Comparison of methods							
		Encoding time reduction	BD-rate			Encoding time reduction	BD-rate
Method in Ref. 6 (HM 10.0)				Method in Ref. 16 (HM 10.0)			
Classes: A, B, C, D, E, F	Min	54% D, E	0.51% D; 0.91% E	Classes: A, B, C, D, E, F	Min	44.25% D	0.34%
	Max	56% F	0.47%		Max	62.6% E	1.06%
	Average	55%	0.61%		Average	52.34%	0.82%
Method in Ref. 7 (HM 10.0)				Method in Ref. 18 (HM 14.0)			
Only selected sequences from different classes	Min	36.60%	1.60%	Classes: A, B, C, D, E	Min	52.15% B	0.88%
	Max	40.20%	1.70%		Max	55.07% C	2.34%
	Average	37.80%	1.70%		Average	53.52%	1.98%
Method in Ref. 8 (HM 4.0)				Method in Ref. 38 (HM 14.0)			
Classes: A, B, C, D, E	Min	19.44% C	0.74%	Classes: A, B, C, D, E	Min	44.9% D	0.63%
	Max	20.53% B	0.67%		Max	65.07% E	1.05%
	Average	19.99%	0.74%		Average	53.90%	0.70%
Method in Ref. 9 (HM 10.0)				Method in Ref. 39 (HM 15.0)			
Classes: A, B, C, D, E	Min	No data	No data	Incomplete Classes: A, B, C, D, E, F	Min	43.64% B	2.70%
	Max	No data	No data		Max	47.78% F	3.46%
	Average	54%	0.90%		Average	46.00%	3.35%
Method in Ref. 10 (HM 4.0)				Proposed method (HM 16.9)			
Only selected test sequences	Min	20.74%	0.45%	Classes: A, B, C, D, E, F + separate class H	th = 1.5%		
	Max	33.53%	0.23%		Min	20% D	0.28%
	Average	28.80%	0.47%		Max	50% A	2.28%
Method in Ref. 11 (HM 11.0)					Average	30%	3%
Classes: A, B, C, D, E	Min	49.68% D	1.50%		th = 2%		
	Max	58.47% E	2.06%		Min	19.73% D	0.29%
	Average	54%	2.38%		Max	54.6% E	8.90%
Method in Ref. 15 (HM 10.0)					Average	38.27%	6.59%
Classes: A, B, C, D, E, F	Min	34.25% C	0.60%		—	—	—
	Max	62.67% E	2.43%		—	—	—
	Average	47%	1.08%		—	—	—

has already been proven to be a viable method of lowering the computational complexity of video encoding in HEVC standard and has been adopted in the reference encoder.

This method of reducing the time needed to encode video relies on establishing a trade-off between simplicity and compression efficiency, which directly influences the quality of the encoded video. In this paper, we present a strategy for selecting omitted encoding modes, based on a statistical analysis of the encoding process. The method by which the encoding time is reduced is inherently different than other solutions already proposed within the literature.

The method restricts the set of intracoding modes that are tested in the video encoder. The choice of modes verified during encoding is based on detailed conditional statistics of the usage of individual coding modes. Detailed statistics of the sizes of CU/PU and TU blocks allow us to omit the need to check blocks of certain sizes that are rarely used and, consequently, increases the speed at which the encoded image is partitioned into CU/PU and TU. More importantly, the method presented in this paper, when additionally applied to the state-of-the-art acceleration methods, such as RMD and MPM tools, allows for a quick selection of the intrapredictor in each CU block in the image. The proposed method is able to complement the performance of other methods presented within the literature, allowing for an even greater reduction of encoder complexity, as proven by the presented results.

A reduction of ∼35% to 45% in encoding time for sequence classes A, B, C, and H, at the expense of a reduction in efficiency of ∼1.5% to 2%, can be obtained. This result is, however, better than those achieved by methods currently implemented in reference software. In general, better results were observed for high-resolution sequences as well as sequences with blurred content and large flat areas than for lower resolution sequences or ones that have a lot of small details in the image content.

Nevertheless, this method requires the knowledge of a rich set of conditional statistics of coding modes, as previously mentioned; these statistics must be stored in a memory of video encoder. For this reason, the proposed method may be more suitable for the software realizations of the encoder. Regarding the hardware realizations of the encoder, the statistics of the modes can be regarded as a starting point for further encoder optimization.

Acknowledgments

The research project was supported by the National Centre for Research and Development, Poland, under Grant No. LIDER/023/541/L-4/12/NCBR.

References

1. ISO/IEC and ITU-T, "High efficiency video coding (HEVC)," ISO/IEC 23008-2 (MPEG-H Part 2)/ITU-T Rec. H.265 (2013).
2. J.-R. Ohm et al., "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1669–1684 (2012).
3. ISO/IEC 14496-10, "Generic coding of audio-visual objects, part 10: advanced video coding," (2006).
4. J. W. Woods, *Multidimensional Signal, Image, and Video Processing and Coding*, Elsevier, Waltham, Massachusetts (2012).
5. G. J. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012).
6. D. Palomino et al., "Fast HEVC intra mode decision algorithm based on new evaluation order in the coding tree block," in *Picture Coding Symp.* (2013).
7. T. L. Silva, L. A. Silva Cruz, and L. V. Agostini, "HEVC intra mode decision acceleration based on tree depth levels relationship," in *Picture Coding Symp.* (2013).
8. W. Jiang, H. Ma, and Y. Chen, "Gradient based fast mode decision algorithm for intra prediction in HEVC," in *Proc. of 2nd Int. Conf. on Consumer Electronics, Communications and Networks (CECNet)*, pp. 1836–1840 (2012).
9. H. Zhang, Q. Liu, and Z. Ma, "Priority classification based fast intra mode decision for high efficiency video coding," in *Picture Coding Symp.* (2013).
10. G. Tian and S. Goto, "Content adaptive prediction unit size decision algorithm for HEVC intra coding," in *Proc. Picture Coding Symp. (PCS)*, pp. 405–408 (2012).
11. W. Zhao, T. Onoye, and T. Song, "Hierarchical structure based fast mode decision for H.265/HEVC," *IEEE Trans. Circuit Syst. Video Technol.* **25**, 1651–1664 (2015).
12. D. Xu, Q. Lin, and X. Dong, "Fast intracoding unit size decision algorithm for high-efficiency video coding," *J. Electron. Imaging* **23**(1), 013012 (2014).
13. L. Zhao et al., "Fast mode decision algorithm for intra prediction in HEVC," in *IEEE Visual Communications and Image Processing*, pp. 1–4 (2011).
14. Y. Piao, J. Min, and J. Chen, "Encoder improvement of unified intra prediction," JCTVC-C207 (2010).
15. L. Shen, Z. Zhang, and Z. Liu, "Effective CU size decision for HEVC intracoding," *IEEE Trans. Image Process.* **23**(10), 4232–4241 (2014).
16. B. Min and R. C. C. Cheung, "A fast CU size decision algorithm for the HEVC intra encoder," *IEEE Trans. Circuits Syst. Video Technol.* **25**(5), 892–896 (2015).
17. S. Jung and H. Park, "A fast mode decision method in HEVC using adaptive ordering of modes," *IEEE Trans. Circuits Syst. Video Technol.* **26**, 1 (2015).
18. K. Lim et al., "Fast PU skip and split termination algorithm for HEVC intra prediction," *IEEE Trans. Circuits Syst. Video Technol.* **25**(8), 1335–1346 (2015).
19. T. L. da Silva, L. V. Agostini, and L. A. da Silva Cruz, "Fast HEVC intra prediction mode decision based on EDGE directional information," in *Proc. of 20th European Signal Processing Conf. (EUSIPCO)*, pp. 1214–1218 (2012).
20. H. Sun, D. Zhou, and S. Goto, "A low-complexity HEVC intra prediction algorithm based on level and mode filtering," in *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 1085–1090 (2012).
21. L. Shen, Z. Zhang, and P. An, "Fast CU size decision and mode decision algorithm for HEVC intra coding," *IEEE Trans. Consum. Electron.* **59**(1), 207–213 (2013).
22. S. Cho and M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding," *IEEE Trans. Circuits Syst. Video Technol.* **23**(9), 1555–1564 (2013).
23. H.-M. Yoo and J.-W. Suh, "Fast coding unit decision algorithm based on inter and intra prediction unit termination for HEVC," *IEEE Int. Conf. on Consumer Electronics (ICCE)*, pp. 300–301 (2013).
24. S. Yang et al., "Fast inter mode decision process for HEVC encoder," in *IEEE 11th IVMSP Workshop 2013*, pp. 1–4 (2013).
25. J. Xiong et al., "A fast HEVC inter CU selection method based on pyramid motion divergence," *IEEE Trans. Multimedia* **16**(2), 559–564 (2014).
26. Y. Cao and S. Goto, "A mode filtering algorithm for accelerating HEVC FME," in *IEEE 15th Int. Workshop on Multimedia Signal Processing (MMSP)*, pp. 218–223 (2013).
27. C. E. Rhee et al., "A survey of fast mode decision algorithms for inter-prediction and their applications to high efficiency video coding," *IEEE Trans. Consum. Electron.* **58**(4), 1375–1383 (2012).
28. A. Lee et al., "An efficient inter prediction mode decision method for fast motion estimation in HEVC," in *Int. Conf. on ICT Convergence (ICTC)*, pp. 502–505 (2013).
29. M. Kim, H.-J. Lee, and N. Ling, "Fast merge mode decision for diamond search in high efficiency video coding," in *Visual Communications and Image Processing (VCIP)*, pp. 1–6 (2013).
30. J. Xiong et al., "MRF-based fast HEVC inter CU decision with the variance of absolute differences," *IEEE Trans. Multimedia* **16**(8), 2141–2153 (2014).
31. J. Xiong et al., "Fast HEVC inter CU decision based on latent SAD estimation," *IEEE Trans. Multimedia* **17**(12), 2147–2159 (2015).
32. J. Lainema et al., "Intra coding of the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1792–1801 (2010).
33. "HEVC test model reference software," https://hevc.hhi.fraunhofer.de.
34. O. Bougacha et al., "Statistical analysis of intra prediction in HEVC video encoder," *Int. Image Processing, Applications and Systems (IPAS)*, Hammamet, pp. 1–5 (2016).
35. F. Belghith et al., "Statistical analysis and parametrization of HEVC encoded videos," in *World Congress on Information Technology and Computer Applications (WCITCA)*, Hammamet, pp. 1–5 (2015).
36. F. Bossen, "Common test conditions and software reference configurations," in *Joint Collaborative Team on Video Coding (JCT-VC) of ITUT SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, Stockholm, Sweden Doc. JCTVC-J1100 (2012).

37. G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," in *ITU-T SG16/Q6, 13th VCEG Meeting*, Austin, Texas, Doc. VCEG-M33 (2001).
38. T. Zhang et al., "Fast intra-mode and CU size decision for HEVC," *IEEE Trans. Circuits Syst. Video Technol.* **27**(8), 1714–1726 (2017).
39. J. Tariq and S. Kwong, "Adaptive stopping strategies for fast intra mode decision in HEVC," *J. Visual Commun. Image Represent.* **51**, 1–13 (2018).

Krzysztof Wegner is a coauthor of papers and documents on video compression, free view television, depth estimation, view synthesis, and 14 patent applications. He is involved in ISO standardization activities where he contributes to the development of future three-dimensional (3-D) video coding standards. Currently, he is a cochair of MPEG AHG FTV and coeditor of ISO/IEC and ITU-T (JCT-3V) 3D-HEVC Standard and Test Model.

Damian Karwowski is an assistant professor at Poznan University of Technology, a member of the Chair of Multimedia Telecommunications and Microelectronics. He is an author and co-author of more than 50 papers on video compression. He has been taking part in several research as well as industry-oriented projects that encompasses video and audio compression. He was technical program chair of international conferences PCS 2012 and IWSSIP 2017.

Jakub Stankowski received his MSc degree from Poznań University of Technology in 2008. Currently, he is a researcher at the Chair of Multimedia Telecommunications and Microelectronics. He is an author and coauthor of several papers on digital video compression.

Tomasz Grajek received his MSc and PhD degrees from Poznań University of Technology in 2004 and 2010, respectively. He works as an assistant at the Chair of Multimedia Telecommunications and Microelectronics. He is a coauthor of over fifty papers on digital video processing and compression and holds ten patents (EPO and US). He has been leading several projects for industrial research and development.

Krzysztof Klimaszewski received his MSc degree and PhD from Poznan University of Technology in 2005 and 2012, respectively. The main areas of his professional activities are video compression, free viewpoint video, and electronic design. He is a coauthor of several scientific papers and patents.

Olgierd Stankiewicz is currently an assistant professor at Poznan University of Technology. Since 2008, actively involved in development of the 3-D video coding technologies and standards in activities like JPEG-PLENO, MPEG free viewpoint TV and MPEG-I. In years 2011–2014, a coordinator of MPEG reference software for 3D-video coding standards based on AVC. Published over 200 standardization documents and papers on: depth estimation, view synthesis, video compression, and hardware implementations.

RATE-DISTORTION OPTIMIZED QUANTIZATION IN HEVC: PERFORMANCE LIMITATIONS

Jakub Stankowski, Cezary Korzeniewski, Marek Domański, Tomasz Grajek
Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology
60-965 Poznań, Polanka 3 Street, POLAND
{jstankowski, tgrajek}@multimedia.edu.pl

***Abstract*—Rate-Distortion Optimized Quantization (RDOQ) is an encoding optimization technique that may be modified with no implications to the bitstream compliance with the standard. In the paper, we propose an RDOQ variant with exact cost estimation resulting in higher compression ratios as compared to HEVC Test Model. This improvement is obtained at the cost of the complexity increase that may be scaled by replacing the individual steps of exact estimation by the simplified ones from HEVC Test Model. In that way, a family of RDOQ variants is created. The performance of the exact variant from this paper defines a kind of the performance limit for RDOQ.**

I. INTRODUCTION

The hybrid video coding scheme is the most common and the widest adopted approach to compressed video transmission. During the last two decades three generations of hybrid video coding technologies have been developed (Fig. 1) [1].

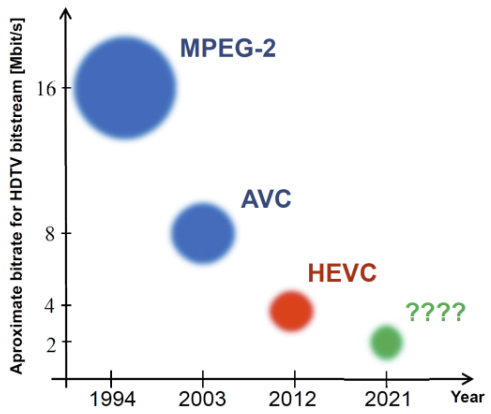


Figure 1. Generations of video coding technology. The reqired bitrate for every new generation is halved (with preserving the same quality).

In the most widespread video compression techniques like MPEG-2 Video/H.262 [2], AVC (MPEG-4 part 10 and H.264) [3,4] and novel HEVC (MPEG-H part 2 and H.265) [5,6] only bitstream syntax, bitstream semantics and decoder operations are standardized. Therefore, there is a wide margin of flexibility in encoder operation leading to the development of a number of encoder control algorithms. The development of the new control algorithms result in further improvement in coding efficiency within single standard (Fig. 2).

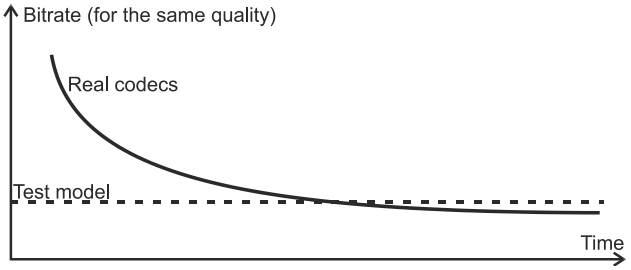


Figure 2. Bitrate for the same quality over time for singe compression standard.

In the paper the authors focus on calculation of quantized transform coefficients due to the fact that those coefficients constitute about 50-70% of compressed video bitstream [7]. Therefore, efficient calculations of quantized transform coefficients have significant influence on coding efficiency. However, the selection of quantization parameter (QP) value is out of the scope of this work.

II. QUANTIZED COEFFICIENTS CALCULATION

In the video encoder operation, the quantization stage is the only one responsible for lossy data reduction by both quantizing and reducing the number of transform coefficients. Therefore, the calculation of the quantized transform coefficients has a significant impact on the compression efficiency. The compression efficiency (in term of rate-distortion) offered by the encoder utilizing scalar quantization can be significantly improved by choosing the more sophisticated way of calculation of the quantized transform coefficients (it is not standardized and may be performed in any manner).

During the evolution of video compression techniques a number of approaches targeting the improvement of quantized coefficient calculation has been developed. For example:

- Dead zone quantizer aiming at the reduction of low magnitude coefficients (in particular magnitude equal to 1) by introducing the dead zone step in uniform quantizer [8,9].
- Coefficient thresholding "outliers" in AC coefficients of magnitude equal to 1 after quantization are discarded due to too high cost of transmission [10].
- Adaptive Rounding based on adaptive adjustment of a rounding offset used in coefficient quantization stage [11].
- Adaptive Quantization Matrix Selection (AQMS) allows for adjustment of quantization matrices to encoded signal also on the decoder side [12].
- Rate Distortion Optimized Quantization based on rough estimating the RD cost of modification or removal of selected transform coefficient or transform coefficients group [13].

Due to its efficiency the RDOQ has been selected as one of two possible quantization approaches in the HEVC test model development [14]. Therefore, in this paper the RDOQ has been selected as a most promising technique for future improvements. The performance limitations for improved RDOQ have been evaluated.

III. RATE-DISTORTION OPTIMIZED QUANTIZATION (RDOQ)

The improved quantizer can take into account both the quantization error and the number of bits required to transmit the transform coefficient and determine the optimal set of quantized transform coefficients. The optimization can be performed for every block of transform coefficients (i.e. Transform Unit (TU) in HEVC) and the optimal cost is calculated by minimizing the Lagrangian function [15].

The above introduced approach has been proposed in [16] for JPEG and MPEG, and then in [17,18] for MPEG-4 AVC and in [13] for KTA. The general idea of presented method is to find an optimal set of quantized transform coefficients which corresponds to the lowest RD cost. Theoretically the determination of optimal set of quantized coefficients requires exhaustive search by evaluating all possible combinations. The exhaustive approach is impractical due to extreme computational complexity of the encoder. Therefore, fast suboptimal approaches have been introduced. Additionally, in [13] the name "Rate-Distortion Optimized Quantization (RDOQ)" has been proposed, which is widely adopted

The purpose of the RDOQ is to find the optimal or suboptimal set of quantized transform coefficients representing a residual data in a encoded block. The RDOQ calculates the image distortion (introduced by quantization of transform coefficients) in encoded block and a number of bits needed to encode the corresponding quantized transform coefficient. Based on these two values, the encoder chooses better coefficient value, by calculating RD cost.

IV. THE RDOQ IN HEVC

The RDOQ has been included in the HEVC reference software (HM) and intensively used during HEVC development and performance. This section describes the RDOQ algorithm adapted to HEVC.

The adaptation of the RDOQ in HM is closely related to HEVC residual coding techniques. In the HEVC the Transform Unit size may vary from 4x4 to 32x32 pixels and only square units are allowed. After transformation and scanning, coefficients are divided into coefficient groups (CG) containing 16 transform coefficients (Fig. 3). The detailed description of HEVC transform coefficient coding could be found in [19].

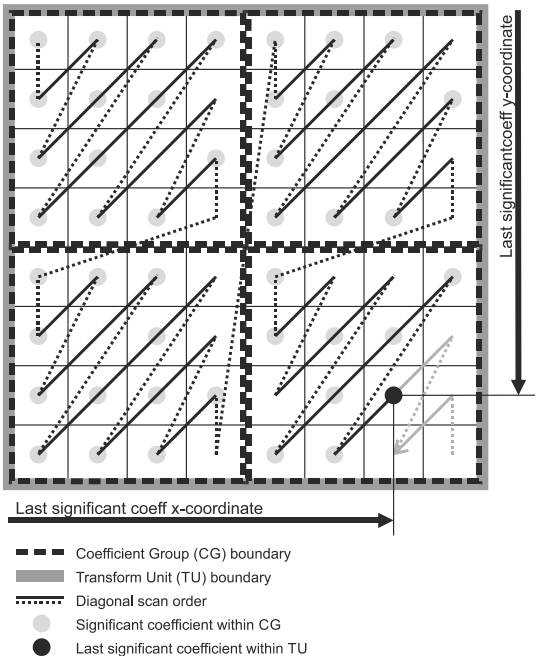


Figure 3. The exemplary 8x8 Transform Unit (TU) with division into four Coefficient Groups (CGs), diagonal scan order and last significant coeff marked.

The RDOQ operation in encoder can be divided into three stages: quantization of transform coefficients, elimination of coefficient groups (CG) and selection of the last non-zero coefficient.

A. *Quantization of transform coefficients*

In this stage the encoder performs calculation for each of transform coefficients separately. In the first step, the encoder calculates the value *Level* by quantizing the magnitude of transform coefficient by using the uniform quantizer without dead zone. In the next step, the encoder considers two additional magnitudes of the analyzed quantized coefficient: *Level-1* and *0*. For every of the mentioned coefficient magnitudes, the encoder calculates the RD cost of encoding the coefficient with the selected magnitude and chooses the one with the lowest RD cost.

It is worth to mention that, when compared to coefficient magnitude value equal to *Level* setting the magnitude to *0* value

allows for more significant bitrate reduction than setting the lower magnitude value (*Level-1*). However, the elimination of selected transform coefficient by setting the magnitude to 0 may cause significant distortion.

B. Elimination of coefficient groups

In this stage the encoder performs the calculation for each of transform coefficient group (CG). The encoder calculates the RD cost of eliminating the whole CG. The elimination of the whole CG is performed by quantization of all coefficient in CG to magnitude zero. The encoder calculates the RD cost of elimination of the analyzed CG and, if the elimination allows for cost reduction, the selected CG is eliminated.

The elimination of the entire CG can lead to substantial bitrate reduction (there is no need to transmit *sig_coeff_flag* for every coefficient within CG) while introducing substantial distortion to the reconstructed image.

C. Selection of the last non-zero coefficient

The last stage of RDOQ is performed after steps A and B for all remaining CGs in TU. RDOQ algorithm analyzes coefficients to find the best (in term RD cost) last non-zero coefficient position. This step is included because the encoder has to encode the (x, y) coordinates of the last non-zero coefficient in the bitstream.

D. Calculation of RD Cost

During RDOQ operation the encoder has to calculate the cost of every considered set of transform coefficients or coefficient groups. This cost (*RD_cost*) may be calculated by taking into account the number of bits (B) required to encode selected coefficient, CG or TU, the introduced distortion (D) and weighting both values by Lagrange multiplier (λ) [13]:

$$RD\_cost = D + \lambda \cdot B \tag{1}$$

In different implementations, encoders may use the exact or estimated values of introduced distortion and a number of bits required to encode selected transform coefficient, coefficient group or transform unit (see Table I).

TABLE I. EXEMPLARY PERFORMANCE/COMPLEXITY TRADEOFS.

	Estimated number of bits calculation	**Exact number of bits calculation**
Estimated distortion calculation	RDOQ implementation in HEVC	Posible performance/complexity tradeof
Exact distortion calculation	Posible performance/complexity tradeof	RDOQ implementation evaluated in this paper

The usage of estimated values leads to some mistakes in the best coefficient set selection and causes some compression performance degradation. However the estimation of rate and distortion can speed up the encoder operation.

V. THE SIMPLIFIED RDOQ IN HEVC REFERENCE SOFTWARE

In the RDOQ implemented HEVC test model (HM16) [20] the encoder uses only estimated values of introduced distortion (represented by square quantization error) and a number of bits required to encode selected transform coefficient, coefficient group or transform unit.

For example, for every of the examined coefficient magnitude the encoder calculates the cost RD_cost*(L, c)* of encoding the coefficient c with the magnitude L according to (2) and chooses the case with the lowest RD cost.

$$RD\_cost\,(L, c) = est\_D(L, c) + \lambda \cdot est\_B(L, c), \tag{2}$$

where:
c – transform coefficient identifier,
L – value of quantized transform coefficient c,
RD_cost(L, c) – cost of quantization coefficient c to value L,
est_D(L, c) – square quantization error,
est_B(L, c) – estimated number of bits needed do encode coefficient c quantized to value L,
λ – Lagrange multiplier.

The detailed description of RDOQ implementation in HEVC can be found in [21].

VI. THE RDOQ WITH EXACT COST CALCULATION

The RDOQ implementation evaluated in this paper is called "Exact cost RDOQ", because its main feature is the exact calculation of introduced distortion and a number of bits required to encode selected transform coefficient, coefficient group or transform unit. Moreover, the higher number of possible modifications of coefficient magnitude is considered – besides *L' = Level*, *L'=Level-1* and *L'=0*, the *L'=Level+1* is also examined.

The exact calculation of introduced distortion is performed in TU granularity by scaling the quantized transform coefficients, inverse transformation of scaled coefficients, calculation of reconstructed image block (corresponding to currently processed TU) and calculation of Sum of Square Differences (SSD) between original and reconstructed blocks. The exact calculation of a number of bits required is also performed for entire TU. The encoder uses CABAC engine to calculate a number of bits. The CABAC internal state is stored, the examined TU is encoded in order to calculate a number of required bits and the state of CABAC is restored (Fig. 4).

Both abovementioned operations are performed at TU level, therefore the examination of each modification of single transform coefficient requires reconstruction (SSD calculation) and encoding (number of bits calculation) of entire TU containing examined coefficient. For each of examined coefficient value the number of bits needed to encode TU and image distortion are calculated:

$$RD\_cost\,(L, c) = SSD(L, c) + \lambda \cdot B(L), \quad (3)$$

where:
c – transform coefficient identifier,
L – value of quantized transform coefficient c,
$RD\_cost(L, c)$ – cost of quantization coefficient c to value L,
$SSD(L, c)$ – sum of square differences between original and reconstructed image blocks corresponding to encoded TU,
$B(L)$ – exact number of bits needed do encode TU with coefficient c quantized to L value,
λ – Lagrange multiplier.

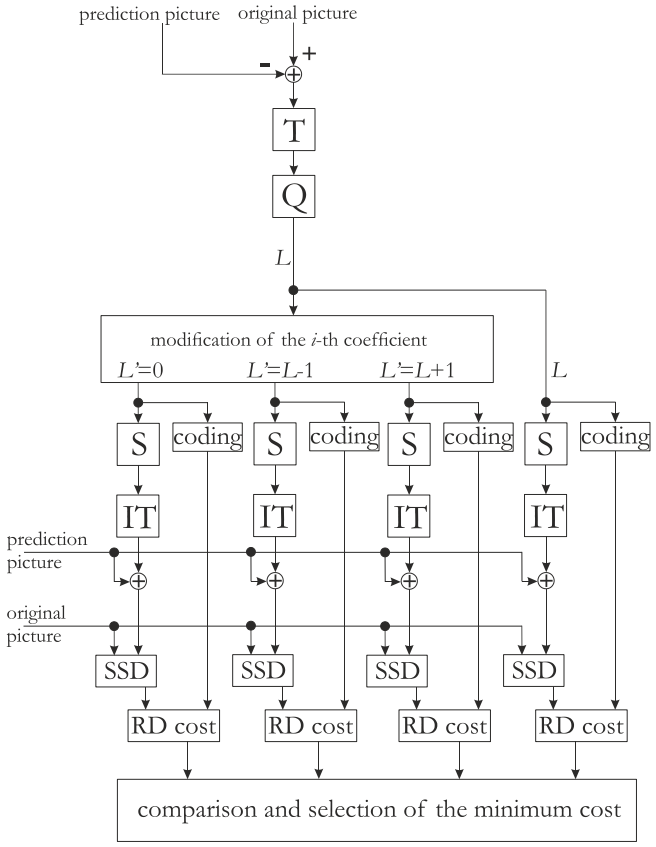


Figure 4. Flowchart proposed RDOQ improvement
(T – transform, Q – quantization, S – scaling, IT – inverse transform, SSD – Sum of squared differences)

VII. Experimental evaluation

The experimental evaluation has been performed under the conditions described in the Common Test Conditions and Software Reference Configurations [14]. The test conditions described in [14] and previous versions of this document have been used by JCT-VC (Joint Collaborative Team on Video Coding of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11) to evaluate proposals during the development of HEVC technology.

The set of test sequences described in [14] consists of 6 classes differing in resolution and content, where: class A – natural content, 4k (cropped to 1600p) resolution, class B – natural content 1080p resolution, class C – natural content WVGA resolution, class D – natural content WQVGA resolution,. class E – teleconference 720p resolution and class F – synthetic and mixed content, different resolutions.

The [14] defines four coding scenarios depending on the used picture types and GOP structure: "All-Intra" – with intra coded pictures only, "Random access" – with hierarchical GOP, "Low-delay P" and "Low-delay B" – only with forward temporal prediction. In this paper only "All-Intra" and "Random access" scenarios have been evaluated.

VIII. Experimental Results

The experimental results of "All-Intra" test scenario have been shown in Table I. and results of "Random access" test scenario have been shown in Table II. The results are presented as BD-delta (Bjontegaard delta) [22]. However, due to interpolation issues and their influence on BD-delta results, the 3rd order polynomial interpolation has been replaced by Piece-wise Cubic Interpolation [23].

TABLE I. Experimental results for "All-Intra" test scenario.

	HM RDOQ performance [%]			**Exact cost RDOQ implementation performance [%]**			**HM RDOQ vs Exact cost RDOQ [%] (performance limit)**		
Class	***Y***	***Cb***	***Cr***	***Y***	***Cb***	***Cr***	***Y***	***Cb***	***Cr***
A	-4.1	1.0	-0.7	-5.1	0.7	-1.2	-1.0	0.0	-0.4
B	-5.4	0.1	-1.4	-6.5	-0.9	-1.1	-1.2	0.0	-0.2
C	-3.5	1.0	-0.3	-4.4	1.1	-0.2	-1.0	0.4	0.3
D	-3.3	1.2	0.0	-4.2	1.3	0.2	-1.0	0.3	0.4
E	-3.7	-0.9	-2.9	-4.5	-1.0	-3.2	-0.9	0.1	-0.2
F	-2.6	0.0	-0.9	-3.8	-0.3	-1.3	-1.2	-0.1	-0.2
All	**-3.8**	**0.4**	**-1.0**	**-4.8**	**0.1**	**-1.0**	**-1.1**	**0.1**	**0.0**

TABLE II. Experimental results for "Random access" test scenario.

	HM RDOQ performance [%]			**Exact cost RDOQ implementation performance [%]**			**HM RDOQ vs Exact cost RDOQ [%] (performance limit)**		
Class	***Y***	***Cb***	***Cr***	***Y***	***Cb***	***Cr***	***Y***	***Cb***	***Cr***
A	-3.3	-4.9	-6.4	-4.0	-3.0	-4.7	-0.9	-0.2	-0.3
B	-5.1	-5.9	-6.3	-5.6	-6.7	-6.4	-0.9	0.0	-0.4
C	-4.0	-5.9	-6.4	-4.8	-5.2	-5.6	-0.9	0.3	0.0
D	-3.9	-5.0	-6.1	-4.8	-4.9	-5.6	-0.9	0.1	0.4
E	-3.2	-2.6	-4.3	-4.0	-2.8	-4.9	-0.9	-0.3	-0.7
F	-3.7	-4.6	-5.1	-4.9	-4.4	-5.0	-1.3	-0.2	-0.2
All	**-4.0**	**-5.0**	**-5.8**	**-4.7**	**-4.5**	**-5.4**	**-1.0**	**-0.1**	**-0.1**

As shown in the Table I for All-Intra and Table II for Random access test scenario the HM RDOQ allows to achieve 3.8% and 4.0% average bitstream reduction for the same image quality respectively. The achieved performance improvement limit (for exact cost calculation algorithm) is 4.8% and 4.7% respectively. Therefore, in the extreme case the HM RDOQ performance could be improved at least by 1% in term of bitstream reduction.

The average encoding time for RDOQ with precise calculations has been 3 to 4 times longer than for HM RDOQ. Therefore the fast HM RDOQ implementation offers 1% lower performance but with significantly lower computational complexity.

IX. Conclusions

In this paper the performance limitation of RDOQ in HEVC has been evaluated. The increase of compression efficiency provided by the fast RDOQ implementation has been measured. Moreover the RDOQ implementation with exact calculation of distortion and number of bits has been introduced. The described approach to RDOQ algorithm allows to find the performance limit which could be reached by the exact calculation of bits number and distortion.

In this paper the upper limit of RDOQ introduced performance improvement has been determined. However it is possible to achieve another performance-encoding complexity tradeoff i.e. by using the exact calculation of distortion in combination with the estimation of the number of bits.

Acknowledgement

Research project was supported by National Science Centre, Poland, according to the decision DEC-2012/05/B/ST7/01279.

References

[1] M. Domański, T. Grajek, D. Karwowski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, J. Stankowski, K. Wegner, "Coding of multiple video+depth using HEVC technology and reduced representations of side views and depth maps", 29th Picture Coding Symposium, PCS 2012, Kraków, Poland, May 2012.

[2] ISO/IEC 13818-2:2013, Information technology - Generic coding of moving pictures and associated audio information - Part 2: Video / Recommendation H.262 : Information technology - Generic coding of moving pictures and associated audio information: Video, International

[3] ISO/IEC 14496-10:2014, Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding / Recommendation H.264 : Advanced video coding for generic audiovisual services, International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), Feb. 2014.

[4] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[5] ISO/IEC 23008-2:2013, Information technology – Coding of audio-visual objects – Part 2: High efficiency video coding / Recommendation ITU-T H.265, High efficiency video coding for generic audiovisual services, 2013.

[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard,", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1648–1667, Dec. 2012.

[7] J. Stankowski, D. Karwowski, T. Grajek, K. Wegner, J. Siast, K. Klimaszewski, O. Stankiewicz, M. Domański, "Bitrate Distribution of Syntax Elements in the HEVC Encoded Video", International Conference on Signals and Electronic Systems, ICSES 2014, Poznań, Poland, Sep. 2014.

[8] G.J. Sullivan, "On embedded scalar quantization", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2004, vol.4, pp. 605-608, Montreal, Canada, May 2004.

[9] T. Wedi, S Wittmann, "Quantization offsets for video coding", IEEE International Symposium on Circuits and Systems, ISCAS 2005, pp. 324-327, Kobe, Japan, May 2005.

[10] Y.-K. Tu, J.-F. Yang, M.-T. Sun, "Rate-Distortion Modeling for Efficient H.264/AVC Encoding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 5, pp. 530-543, May 2007.

[11] G.J. Sullivan, "Adaptive quantization encoding technique using an equal expected-value rule", JVT meeting contribution JVT-N011, Hong Kong, Jan. 2005.

[12] A. Tanizawa, T. Chujoh, "Adaptive Quantization Matrix Selection", ITU-T Q.6/SG16 VCEG, document: D.266, 2006.

[13] M. Karczewicz, Y. Ye, I. Chong, "Rate Distortion Optimized Quantization", ITU-T SG 16/Q 6 VCEG, document: VCEG-AH21, Jan. 2008.

[14] F. Bossen, "Common test conditions and software reference configurations", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document: JCTVC-L1100, Geneva, Switzerland, Jan. 2013.

[15] T. Wiegand, M. Lightstone, D. Mukherjee, T. Campbell, S. Mitra "Rate-distortion optimized mode selection for very low bit-rate video coding and the emerging H.263 standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 182 -190, Apr. 1996.

[16] K. Ramchandran, M. Vetterli, "Rate-distortion optimal fast thresh-olding with complete JPEG/MPEG decoder compatibility," IEEE Transactions on Image Processing , vol. 3, no. 5, pp. 700–704, Sep. 1994.

[17] E.-H. Yang, X. Yang, "Rate Distortion Optimization of H.264 with Main Profile Compatibility", IEEE International Symposium on Information Theory, ISIT 2006, pp. 282 – 286, Seatlle, USA, July 2006

[18] E.-H. Yang, X. Yu, "Rate distortion optimization for H.264 inter-frame coding: A general framework and algorithms," IEEE Transactions on Image Processing , vol. 16, no. 7, pp. 1774–1784, Jul. 2007.

[19] J. Sole, R. Joshi, N. Ngyuen, T. Ji, M. Karczewicz, G. Clare, F. Henry, A. Duenas, "Transform Coefficient Coding in HEVC", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1765-1777, Dec. 2012.

[20] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman, G. J. Sullivan (editors), "High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document: JCTVC-R1002, Sapporo, Japan, Jun. 2014.

[21] J.-R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan; T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1669-1684, Dec. 2012.

[22] G. Bjontegaard, "Improvements of the BD-PSNR model," ITU-T SG16 Q.6 Document, VCEG-AI11, Berlin, Germany, Jul. 2008.

[23] J. Wang, X. Yu, D. He, "On BD-rate calculation", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document: JCTVC-F270, Torino, Italy, Jul. 2011.

Analysis of the complexity of the HEVC motion estimation

Jakub Stankowski [1], Damian Karwowski, Krzysztof Klimaszewski,
Krzysztof Wegner, Olgierd Stankiewicz, Tomasz Grajek
Poznan University of Technology, Polanka 3, 60-965 Poznań, POLAND
[1] *jstankowski@multimedia.edu.pl*

***Abstract* – The paper presents analysis and comparison of complexity of selected algorithms of motion estimation used in video compression. With the use of own, highly optimized software implementation the authors explored complexity of the methods applied in the framework of the new High Efficiency Video Compression (HEVC) technology. The influence of many different factors on motion estimation complexity has been deeply studied, including the implementation technique, type of the algorithm, number of processor threads used, kind of metric of blocks similarity. The results obtained allowed to formulate guidelines and conclusions that may be useful for future implementation of motion estimation algorithms in context of HEVC encoders.**

Keywords – **Video encoder, HEVC, motion estimation, complexity**

I. INTRODUCTION

The technique of hybrid video compression has revolutionized the way in which we store and transmit digital images. In this technique, the blocks of the currently encoded image are effectively predicted and coded on the basis of other data (derived from the same image - i.e. intra-frame predictive coding, or derived from other neighbouring images – i.e. inter-frame predictive coding) resulting in a very high efficient video compression. Just for this reason, the technique in recent years has been the subject of international standardization and gained a great interest from the market.

The highest achievement in the field of hybrid video compression is recently standardized High Efficiency Video Coding (HEVC) technology[2]. HEVC high efficiency of compression comes for the price of high complexity of a video encoder. It is commonly known that high complexity of an encoder is mainly caused by motion compensated inter-frame prediction, which requires computationally intensive motion estimation process. For every block size of every picture, an encoder has to find optimal motion vector against the number of previously encoded frames in order to choose the best compression mode for a given fragment of an image. Unfortunately, this highly increases the complexity of an encoder[3]. Therefore, necessary works on complexity reduction of an encoder must take place before any market adaptation (of a new coding technology).

Although a number of different algorithms of motion estimation have been developed and tested since the beginning of hybrid video compression development (e.g. Two Dimensional Logarithmic Search[7], Hexagon-based Search[8], Diamond Search[9], Uneven Multi-Hexagon Search[10], Enhanced Predictive Zonal Search[11] or TZ Search[12]) there are very little details about real complexity of those methods applied in the employed to the market encoders. It is especially true when consider those methods in the new HEVC framework.

First of all, relatively poor implementations of the known methods were used in the previously conducted studies i.e. [11][12] and the improvements made to those methods i.e. [13][14]. The implementation (the source code) of the considered methods was usually non-optimized. It did not use the available for over the decade fast SIMD (Single Instruction Multiple Data) vector instructions such as x86 SSE and AVX. Moreover, very often an implementation did not exploit the computational power of modern computers as it has been single-threaded only. Besides, the implementations flaws of methods studied in the literature, there is very little about the impact of the other techniques that accelerate motion estimation process, such as image pre-interpolation technique for calculating motion vectors performed with ½-pel and ¼-pel accuracy.

For the reasons mentioned above, there is a strong need to came back to complexity analysis of the known motion estimation methods. This paper tries to analyse the cause of high complexly of motion estimation algorithms currently known from the literature.

The goal of this paper is to measure and compare the complexity of selected motion estimation algorithms. Because the authors are interested in showing the results from the practical (production) point of view, the selected algorithms were first implemented from scratch with the use of advanced programming techniques such as SIMD or multithread processing, and then compared against each other and optimized version known from reference encoders. All of the work have been done in the context of HEVC.

II. EFFICIENT IMPLEMENTATION OF THE SELECTED METHODS

In order to properly meet the assumed objectives of the works, three different algorithms of motion estimation have been fully implemented from scratch. These are:

- Brute force ‘Full Search’ algorithm – the method which is a reference to other algorithms. This algorithm is particularly easy to optimize.
- ‘Two Dimensional Logarithmic Search’ (hereafter referred to as ‘Log2 search’) – as example of basic “fast” motion estimation algorithm.
- ‘TZ Search’ – as example of commonly recognized state-of-the-art algorithm. The ‘TZ Search’ is used as default algorithm in reference implementations of many video coding

technologies like Scalable and Multiview AVC reference software (JSVM) [4], HEVC reference software (HM) [5] and Joint Exploration Test Model Software for beyond HEVC development (JEM) [6].

Moreover, each of the selected algorithms has been implemented in three different ways:

- First: using plain C++ programming language. The goal was to prepare the algorithmically optimized implementation including efficient memory and resources usage. This implementation makes the starting point for the two next.
- Second: using SSE instructions. In the C++ program code all critical functions (distortion metric calculation, reference picture interpolation, prediction signal calculation, etc.) has been implemented using SSE intrinsic. In this way data level parallel processing (SIMD) offered by most x86 CPUs is exploited.
- Third: using AVX instructions. All SSE instructions that were introduced in the second way of implementation have been substituted by appropriated AVX instructions, which are available in modern x86 CPUs.

In the case of implementations in which SSE and AVX vector instructions were used, some level of acceleration was achieved due to processing of vectors of data within one processor instruction. Additionally, all three versions have been implemented with multithreaded parallel processing in order to examine the impact of the number of active processor cores on duration of motion estimation process. In total, six versions of each of the algorithm have been prepared.

III. Methodology of Experiments

All the implementations that were presented in Section II have been prepared as standalone modules that can be easily put in and out from any encoder implementation. In order to test complexity of the algorithms from the point of view of their implementation in the HEVC video encoder, the software has been specially adjusted in order to take into account the actual conditions in which the motion estimation process is carried out in the encoder. In particular, the following conditions have been included:

- Calculations are carried out for a numerous of block sizes that were defined in the HEVC encoder. These are 64x64, 32x32, 16x16, 8x8, 4x4.
- Starting point for estimation process is calculated on the basis of prediction of a motion vector. Prediction of a motion vector is performed as defined in the HEVC standard.
- Full pel, half-pel and quarter-pel accuracy of motion estimation is considered. In the case of half- and quarter-pel accuracy interpolation of reference pictures is implemented using HEVC interpolation algorithm.
- Both uni- and bi-directional inter-frame prediction is considered.
- In the multithreaded parallel implementation there have been retained limitations of parallel processing of blocks, known from the HEVC ‘wavefront’ scanning scheme.

The study of the complexity have been made on the basis of execution time of the authors implementation on the PC platform equipped with Intel core i7 – 5820K (3.6 GHz – 1 core, 3,4 GHZ – 2 cores, 3.3 GHz – more than 2 cores) and 64 GB of RAM (68 GB/s memory bandwidth).

Base on the collected data, the following research works were done:

- Comparison of complexity of selected motion estimation algorithms (‘Full Search’, ‘Log2 Search’, and ‘TZ Search’).
- Analysis of the impact of implementation technique on complexity of each selected motion estimation algorithms (plain C++, SSE, or AVX).
- Analysis of the impact of the number of processor threads.
- Comparison of complexity of the algorithm when operating with various block similarity metrics (like SSD, SAD).
- Motion estimation complexity reduction with help of image pre-interpolation.

The execution times of the algorithms have been obtained by performing motion estimation for a set of 16 natural video sequences recorded in full HD spatial resolution with 25, 30 and 50 fps. The sequences are characterized by diverse motion with both static and moving background. The partial motion estimation time results received during the experiments were averaged over entire set of sequences.

IV. Complexity of Motion Estimation in the HEVC- Experimental Results and Analysis

A. *Impact of implementation technique on the complexity of motion estimation.*

The experiments were performed in order to evaluate implementation type influence. During the experiments, the TZ search algorithm has been used, and SSD has been set as a distortion metric. In the presented work we focus on x86 processors, therefore, set of 3 optimized implementations has been investigated: plain C++, SSE based and AVX based, with 6 threads working in parallel. The optimized C++ is approximately 10% faster when compared to HM [5] reference software implementation. The results of experiment are shown on Figure 1.

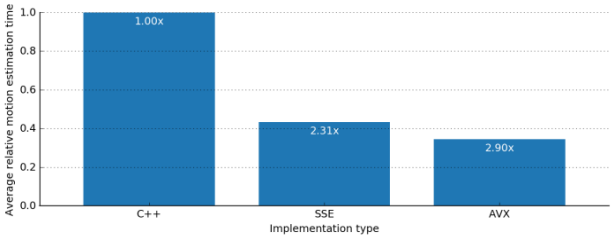


Figure 1. Comparison of the performance of different implementation types (plain C++, SSE and AVX) with speedup over optimized plain C++ implementation. All results gathered for TZ search algorithm and 6 working threads.

The comparison between different implementations clearly shows that the effort spent on optimization of calculations to exploit SIMD instructions is worthwhile. The prepared implementations perform significantly better than the optimized C++ implementation that does not use SIMD instructions. The time reduction is about 57% and 66% (2.31x and 2.90x speedup) for SSE and AVX implementation, respectively.

B. *Motion estimation algorithm and its complexity*

It is obvious that Full Search (FS) algorithm has the highest computational complexity measured as a time of motion estimation. The next experiment was to check the improvement in time performance offered by fast algorithms (Log2 and TZ search) for a fully optimized production quality encoder. The performance

ratios between Full Search and fast algorithms may be influenced by using the best performing (according to the previously described experiment) AVX instruction set. The test scenario assumes the use of multithreaded, AVX optimized implementation and 6 working threads. The results of experiment are shown in Figure 2.

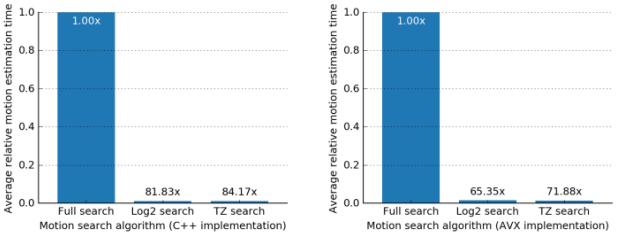


Figure 2. Comparison of search algorithms complexity (measured as relative motion estimation time) with speedup over Full search. All results gathered for SSD distortion metric and 6 working threads.

It is obvious from the results that even for the SIMD optimized methods there is a significant difference in performance between full search and fast search algorithms, although the difference is smaller than for a C++ implementation without SIMD instructions (TZ search is 84 times faster than full search when not using SIMD and 72 times faster than full search when both are using SIMD implementation).

C. *The number of processor threads and its impact on the complexity of motion estimation.*

The next experiment concerned the influence of the number of threads onto the performance of motion estimation. Similarly as before, the best preforming method, namely, an AVX optimized TZ search motion estimation, was chosen for the test. The number of parallel working threads was adjusted in the range from 1 to 12 threads. The results are shown on Figure 3. It is important to note, that all the tests were performed on a computer with a 6 core (included HyperThreading mechanism – 12 threads) processor. The results clearly show that, as long as the number of threads does not exceed the number of cores multiplied by 2, the performance increase stays approximately proportional to the number of threads used (4 threads perform almost 4 times better than a single thread). The increase is not perfectly proportional, most probably due to a reduced CPU frequency when using more than 1 core.

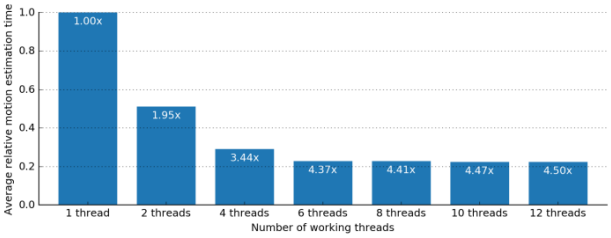


Figure 3. Comparison of different number of working threads on 6 core (with HyperThreading) CPU with speedup over singlethreaded mode. All results gatheres for AVX optimized implementation, TZ search algorithm and SSD distortion metric.

Increasing the number of working threads over the number of actual cores in the processor does not improve the performance. In fact, the performance for threads number exceeding the number of cores leads to a slight (but, nonetheless, noticeable) decrease of the performance of the coder. Also – the increased number of threads increases the required transfer from memory, and this one stays the same, independently of the number of cores used. This forms the most important bottleneck for a multithreaded coder.

It is therefore advised to set the number of threads to match twice the number of cores used in the actual machine that performs the compression.

D. *Metric of similarity of image blocks and its influence on complexity of motion estimation.*

Another important factor influencing the complexity of an encoder is the type of the metric used during motion estimation. The metric is used to express the similarity between two blocks of images – from a current and from a reference image. The most commonly used is SAD – Sum of Absolute Differences, since the SSD – the Sum of Squared Differences can be, due to the presence of the squaring phase, perceived as more computationally expensive. For more precise estimation, like ½- and ¼-pel accuracy motion estimation, the SATD (Sum of Absolute Transform Differences) is used [4][5][6]. This is much more complex, since it involves computation of the transform of a block. The obtained results are shown in Figure 4. The numbers represent the performance ratio when compared to the SAD full pel + SAD fractional pel metric scenario. The lower the bars, the better the performance. It is clearly shown, that with the use of AVX instructions, the SAD and SSD metrics are identical in terms of computational complexity. The same is not true for a simple C++ implementation without SIMD, where the SSD requires approximately 53% more time to calculate. What is more, the use of AVX instructions flattens the performance graph – the most complex combination of metrics, the SSTD full pel + SSTD fractional pel, needs 345% of time needed for SAD + SAD, while for simple C++ implementation it requires 435% of time. Take into account the fact that AVX performs overall approximately 3 times better for SAD itself (as shown in part IV A) and the differences become even more appreciable.

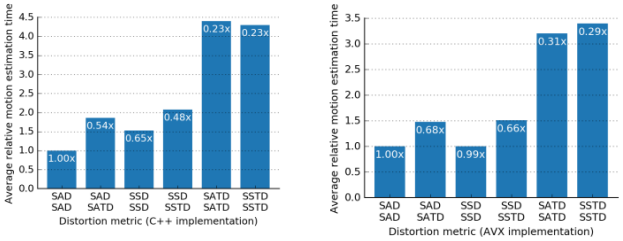


Figure 4. Comparison of different distortion metrics for Full pel / Half- and Quarter-pel block position. Speddup calculated with SAD/SAD as reference. All results gathered for TZ search algorithm and and 6 working threads.

E. *Pre-interpolation of reference pictures and influence of the technique on complexity of motion estimation.*

The motion estimation with ½- and ¼-pel accuracy requires additional effort caused by performing the reference picture interpolation. When compared to AVC [1], HEVC interpolation filters are much more complex and almost double the number of required operations (memory bandwidth, additions and multiplications) to calculate interpolated value [15]. Therefore, the reference picture pre-interpolation and buffering technique described in [16] has been evaluated. In case of AVC the best

tradeoff between complexity and memory usage was to store ½-pel interpolated images and perform ¼-pel interpolation (simple bilinear filter, based on full-pel and ½-pel samples) on-the-fly. In case of HEVC both ½- and ¼-pel blocks are interpolated based on full-pel samples. Moreover, 9 of 15 possible sample positions require two interpolation passes – first horizontal pass and second vertical pass. Due to the increased complexity of interpolation stage, the decision was made to evaluate 4 different scenarios: no pre-interpolation (all calculations are done on-the-fly), pre-interpolation for ½-pel samples, pre-interpolation for ¼-pel samples and pre-interpolation for horizontal samples (to avoid the need of 2 interpolation passes). Only the best performing AVX scenario is considered here. All result have been shown in Figure 5.

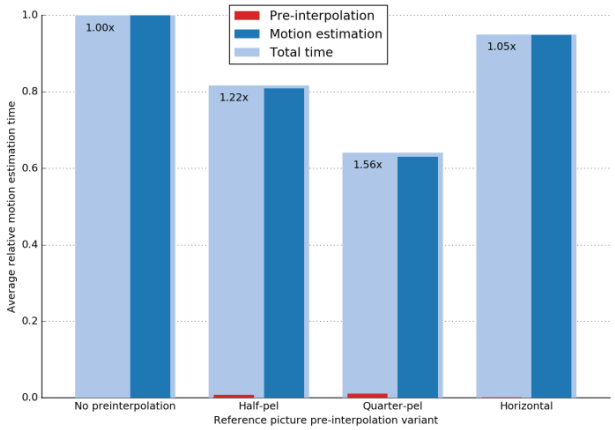


Figure 5. Comparison of pre-interpolation approaches with speedup calculated over on-the-tly (no pre-interpolation) sheme. All results gathered for AVX optimized implementation, SSD distortion metric, TZ search algorithm and and 6 working threads.

The experiments clearly show that buffering of pre-interpolated reference images allows a substantial speedup in the HEVC motion estimation stage. The pre-interpolation for all positions (Quarter-pel) offers best performance (1.56 times faster than on-the-fly) but requires to buffer additional 15 images (thus, requiring 16 time more memory). The ½-pel only and horizontal only pre-interpolation cases require to store only 3 additional pictures but offer lower speedup. Especially the horizontal pre-interpolation case is not worth of use when compared to ½-pel pre-interpolation.

V. Conclusions and Final Remarks

The results shown in this paper support clearly the following conclusions. Firstly, the performance of the encoder can be significantly increased by speeding up the motion estimation phase (the most time consuming one) with the use of SIMD (Single Instruction Multiple Data) instructions. The best performing implementation obtained in the research was making use of the AVX instruction set. The expected increase of the performance for the motion estimation phase for the most frequently used scenario is 3 times. More importantly, the AVX implementation equalizes the performance of SSD and SAD metrics used in motion compensation, so any of them can be selected, without any performance penalty.

Acknowledgment

Research project was supported by The National Centre for Research and Development, POLAND, Grant no. LIDER/023/541/L-4/12/NCBR/2013.

References

[1] ISO/IEC 14496-10, Int. Standard “Generic coding of audio-visual objects – Part 10: Advanced Video Coding” 7th Ed., 2012, also: ITU-T Rec. H.264, Edition 7.0, 2012.

[2] ISO/IEC 23008-2 (MPEG-H Part 2) / ITU-T Rec. H.265: High Eciency Video Coding (HEVC), Apr. 2013.

[3] F. Bossen, B. Bross, K. Suhring, D. Flyn, HEVC Complexity and Implementation Analysis, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, December 2012.

[4] H. Schwarz , M. Wien and J. Vieron, JSVM software manual, Doc. JVT-S070, 2006

[5] F. Bossen, D. Flynn, K. Sharman, K. Sühring, “HM Software Manual”, Joint Collaborative Team on Video Coding of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document: JCTVC-Software Manual

[6] Joint Exploration Test Model Software, Joint Video Exploration Team, of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, https://vceg.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/tags/HM-16.6-JEM-1.0/

[7] J. Jain and A. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," in IEEE Transactions on Communications, vol. 29, no. 12, pp. 1799-1808, Dec. 1981.

[8] Ce Zhu, Xiao Lin, and Lap-Pui Chau, Hexagon-based search patten for fast block motion estimation, IEEE Trans. Circuits Syst. Video Technol., 12(5) (2002) 349–355.

[9] S. Zhu, K.-K. Ma, A new diamond search algorithm for fast block matching motion estimation, in: Proceedings of the International Conference on Information and Communication and Signal Processing, Singapore, Sept., 1997, pp. 292–296.

[10] Peng Yin, Hye-Yeon Cheong Tourapis, Alexis Michael Tourapis, Jill Boyce, Fast mode decision and motion estimation for JVT/H.264, in: Proceedings of the IEEE International Conference on Image Processing 2003, September 2003, vol. 2, pp. 853–856.

[11] A.M. Tourapis, Enhanced predictive zonal search for single and multiple frame motion estimation, in: SPIE Proceedings of the Visual Comm. Image Proc., 2002.

[12] Xiu-li T., Univ H., Sheng-kui D., An analysis of TZSearch algorithm in JMVC, International Conference on Green Circuits and Systems (ICGCS), Shanghai, 21-23 June 2010, s. 516 – 520.

[13] Xufeng Li, Ronggang Wang, Wenmin Wang, Zhenyu Wang and Shengfu Dong, "Fast motion estimation methods for HEVC," Broadband Multimedia Systems and Broadcasting (BMSB), 2014 IEEE International Symposium on, Beijing, 2014, pp. 1-4.

[14] N. Purnachand, L. N. Alves and A. Navarro, "Fast Motion Estimation Algorithm for HEVC," Consumer Electronics - Berlin (ICCE-Berlin), 2012 IEEE International Conference on, Berlin, 2012, pp. 34-37.

[15] Hao Lv, Ronggang Wang, Xiaodong Xie, Huizhu Jia and Wen Gao, "A comparison of fractional-pel interpolation filters in HEVC and H.264/AVC," *Visual Communications and Image Processing (VCIP), 2012 IEEE*, San Diego, CA, 2012, pp. 1-6.

[16] T. Grajek, D. Karwowski, A. Łuczak, S. Maćkowiak, M. Domański: „Architecture of Algorithmically Optimized MPEG-4 AVC/H.264 Video Encoder”, LNCS 7594, pp. 79-86, Springer, Heidelberg (2012)

Computational complexity tradeoffs in HEVC motion estimation

Jakub Stankowski, Tomasz Grajek, Damian Karwowski, Krzysztof Klimaszewski, Olgierd Stankiewicz, Krzysztof Wegner
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznan, Poland
jstankowski@multimedia.edu.pl

***Abstract*—The paper reports the results of the experiments conducted on the HEVC encoder, regarding the complexity and performance of the motion estimation algorithm. The research focuses on the search range of motion vectors and its influence on the encoding time of the video with the use of a HEVC encoder, as well as on the method for choosing the starting point of the motion estimation algorithm. Performance is measured as the time reduction and Bjontegaard metric bitrate reduction.**

Keywords—HEVC compression, motion estimation, motion vectors

I. INTRODUCTION

The most recent development in video coding, standardized as High Efficiency Video Coder (HEVC) by Moving Pictures Experts Group (MPEG) and as H.265 standard by International Standardization Organization (ISO), gives a significant advantage in compression efficiency compared to previously developed standards, like AVC. With about 50% reduction of bitrate compared to AVC with the same quality of reconstructed video [1], HEVC is sure to replace older standards in the coming years.

The high performance of the new codec, however, comes at a cost of greatly increased computational complexity of the encoding process. Entirely revised concept of macroblocks, replaced by a much larger tree structure called Coding Tree Unit (CTU) and introduction of complex prediction algorithms poses a great challenge for hardware and software that encodes the video using the new standard.

In order to make it easier to use the HEVC, many different ways of simplifying the process of encoding are being proposed and evaluated.

One of the most time consuming tasks during the encoding process is motion estimation. The HEVC performs motion estimation using a sophisticated iterative method called TZsearch [2]. The time that is required for this method to find the motion vector depends on the contents of the currently estimated block and on the accuracy of the motion vector prediction. The time required for the motion estimation depends also on the search area.

Research project was supported by The National Centre for Research and Development, POLAND, Grant no. LIDER/023/541/L-4/12/NCBR/2013.

The works reported in this paper investigate the complexity of the motion estimation process and the potential speedup of the encoding based on the modifications in motion vector estimation process. The issue of motion estimation has already been the subject of the authors' research that were described in the work [4]. This work is a continuation of those studies, focused on different aspects of the process, and deeply explores to what extent the motion vector search range and the motion estimation starting point affect the performance of HEVC motion estimation process.

II. MOTION ESTIMATION IN HEVC AND EXPERIMENTS METHODOLOGY

The motion estimation in HEVC works on the tree like structure. In this structure, every CTU can be further divided into smaller coding units (CU). Each CU is, in turn, encoded using prediction performed in prediction units (PU). Each CU can be coded using a single square PU, two square or rectangular PUs or 4 square PUs of half the size of CU. For each PU, the motion vector or plurality of motion vectors need to be estimated.

In order to study the issues connected with time complexity of the motion estimation process as used in HEVC, a set of tests was performed. Two phases of the experiment can be clearly distinguished in the research. In the first phase, the influence of the search range and position of the starting point on performance (complexity and coding efficiency) of HEVC motion estimation were deeply studied. Experience taken from this stage of the research was used in the second phase to elaborate modifications of HEVC motion estimation process. The efficiency of the modified encoder was thoroughly investigated with the use of the framework of the HM 16.9 reference software.

All the tests were performed using the widely accepted dataset, also used by MPEG in its works on video compression standardization and development [3]. The set includes 24 sequences with different properties and different types of motion. The tests were conducted in accordance to the "common test conditions", as defined by MPEG for the works on the video compression technologies [3].

For the purposes of the first stage of the research, a standalone software was prepared from scratch, which enabled an accurate estimation of the time that is spent in the encoder on the process of motion estimation. In this place, special attention was paid to take into account all nuances of the motion estimation used in the HEVC video encoder. Prepared software gave the possibility of testing the complexity of the algorithm with many different settings outlined below.

As already highlighted above, some parts of the HEVC motion estimation algorithm have also been changed in the HM 16.9 software in order to test the possibility of improving the algorithm used in the HM software.

All of the results were obtained on computers of the following configuration: Intel core i7 – 5820K (3.6 GHz – 1 core, 3,4 GHZ – 2 cores, 3.3 GHz – more than 2 cores) and 64 GB of RAM (68 GB/s memory bandwidth).

III. Performance Analysis of Motion Estimation – Results and Discussion

A. *Encoding time versus search range*

The first experiment was focused on assessing the time required for motion estimation process when the search range is different. The results of this experiment are shown on Fig. 1. The reference time is the time used by the motion estimation with search range (SR) set to ±64 pixels. The starting points (SP) for the motion estimation considered during the search are the ones coming from the prediction mechanism used in HEVC as well as zero prediction vector (in this case the starting point of the search is situated in the position of the block itself, i.e. motion of this block is predicted to be zero). The results are averaged over all of the test sequences.

The first observation is that reduction of the search range does influence the estimation time, but the time is not proportional to the area of the search. The time used as a reference of motion estimation process, for the area of ±64 x ±64 pixels is only five times smaller than the time for search range of ±256 x ±256 pixels, which range has 16 times greater area. Similarly, for smaller search ranges, the decrease of the time for search range ±4 x ±4 pixels is only by a factor of little less than 3, while the search area is 256 times smaller. The justification of the results lies in the nature of the TZsearch algorithm. It operates partially as a logarithmic search, therefore limiting the influence of the search area.

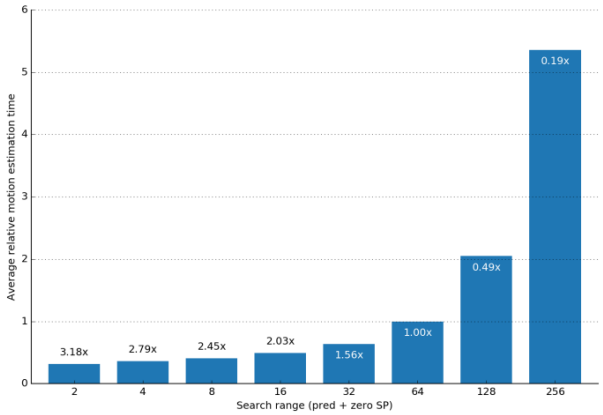


Fig. 1. Dependency of the time consumed by motion estimation algorithm on the search range.

According to those results, while not drastic, the noticeable reduction of the encoding time can be achieved by reducing the search range for motion vectors. Presented results show precisely what is the factor of the time reduction.

However, the reduction of the search range inevitably leads to the loss of the quality of encoded data, especially so for the sequences with complex and fast motion. This kind of motion does not lend itself to an accurate prediction, and reduction of the search range for the motion vectors causes the encoder to select the motion vector that can be far from being optimal.

To verify the quality impact of the reduced search range, another set of data was also gathered during the experiment. This time, the quality of the prediction signal is calculated for different block sizes (only one size of the block is allowed for each of the block size set of results). This quality is measured as a PSNR metric calculated with the original frame undergoing prediction used as reference. The results can be seen on Fig. 2. Additionally to the search range, also the block size is recorded. As previously, the results show an average over all of the sequences used in the test.

As can be seen, the highest PSNR of the prediction signal is achieved when using small blocks of the size 4x4 pixels (size not allowed for inter coded units in HEVC). In this case, the difference of quality between the search range equal to ±256 pixels and ± 2 pixels is 2dB. This difference decreases with the growing size of the search range, and is equal to 1.4dB for the largest possible blocks (64x64 pixels). For the larger block the quality of prediction signal significantly decreases, this is, however, not a surprising fact, since smaller blocks can better adjust to the motion field of the sequence. It must be noted however, that in the case of the use of smaller bocks the total bit cost of sending motion vectors increases (since there is more of them), which quantitatively has not been presented in this work. However, it may be an interesting subject of future research.

B. *Encoding time versus starting point*

As already stated, the performance of the search algorithm also depends on the point where the search starts. Choosing the starting point close to the final motion vector coordinates greatly reduces the time spent on the process of finding the motion vector. In the next experiment, the influence of the

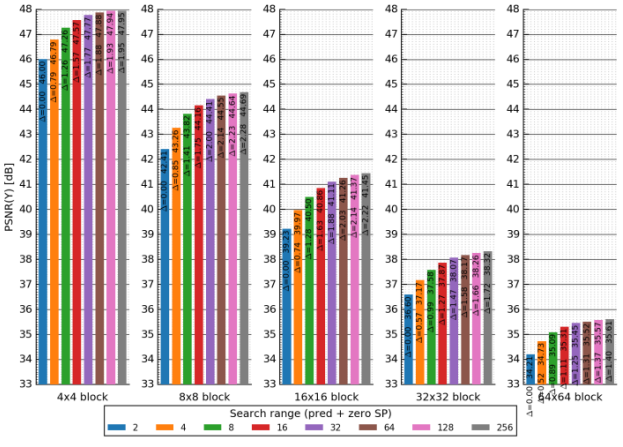


Fig. 2. Dependency of the prediction signal PSNR on the block size and the search range.

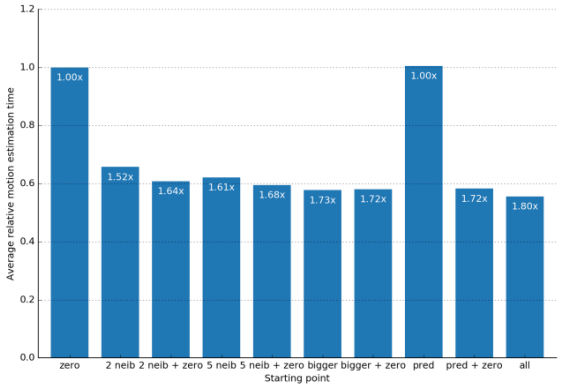


Fig. 3. Dependency of the time consumed by motion estimation algorithm on the starting point.

starting point on the time of the process of motion estimation was evaluated. The results are shown on Fig. 3. The following starting points were investigated:

- *zero*: the motion estimation algorithm starts from the zero position (i.e. the corner of the collocated block),
- *2 neib*: two spatially neighboring blocks motion vectors are checked, and the one with smaller difference to the currently encoded block is used as starting point,
- *2 neib + zero*: a combination of two above cases,
- *5 neib*: five spatially neighboring blocks motion vectors are checked, and the one with the smallest difference to the currently encoded block is used as starting point,
- *bigger*: the motion vector of the parent block (i.e. the bigger one that is one level higher in the CTU division tree) is used as the starting point,
- *pred*: the method used natively in HEVC, selected from the set of candidates as described by HEVC,
- *all*: all of the starting points mentioned above are verified, and the one with the smallest difference to the currently encoded block is used as a starting point.

It can be seen that the process of the motion estimation greatly depends on the starting point, since the closer the starting point is to the ultimately chosen motion vector, the faster the TZsearch algorithm is able to terminate. A reduction of about 40% of the motion estimation time can be expected when all of the mentioned starting points are considered, comparing to the case where only zero starting point is used or the points taken from the algorithm used in HEVC. This shows that the choice of the potential starting points in HEVC is not effective and can be improved by selecting different prediction candidates.

The choice of the starting point for the motion estimation also does influence the prediction signal PSNR, measured with the original encoding frame taken as reference. The corresponding results are presented on the Fig. 4. It can be seen that the significant influence on the PSNR of the prediction signal is visible only for block sizes 4x4 pixels, 8x8 pixels and 16x16 pixels. For large blocks of 64x64 pixels there is practically no difference between different starting points of motion estimation algorithm. For large blocks the quality is already so low, that it seems it cannot be degraded any further by a bad choice of the starting point for motion estimation. This comes from the fact that large blocks do not reflect the motion of objects in the scene with enough precision. What is very prominent is the fact that the selection of the starting point using the method from HEVC performs bad for the small block sizes.

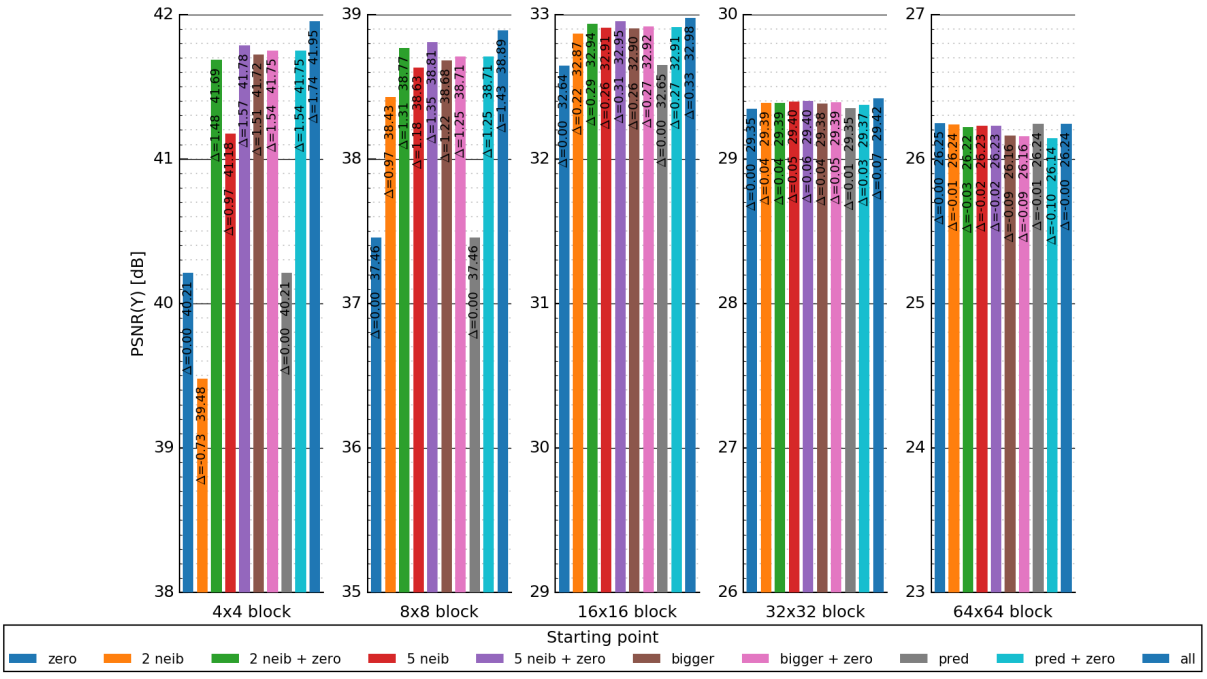


Fig. 4. Dependency of the time consumed by motion estimation algorithm on the starting point.

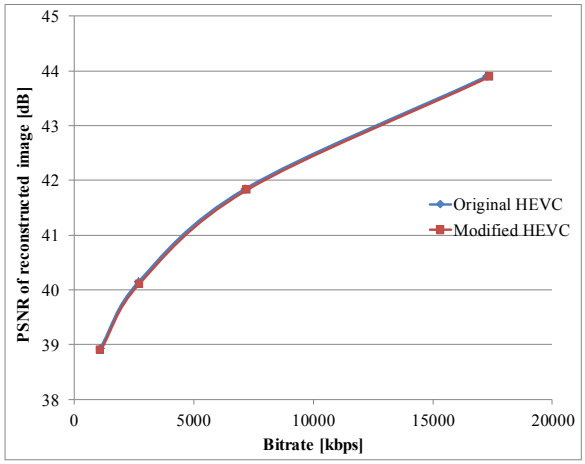


Fig. 5. The quality versus bitrate for scenario A1.

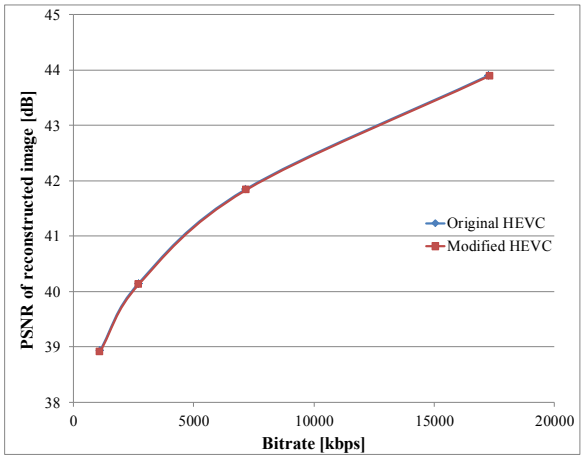


Fig. 6. The quality versus bitrate for scenario A2.

IV. Modifications of Motion Estimation in the HEVC – Results and Disscussion

In order to exploit the described dependency of the motion estimation time on the search range, the following method in two variants has been developed and implemented in HEVC reference codec version HM 16.9. It was decided that the smaller the PU block, the smaller the search range should be, since the search for their motion vectors consumes the most time for the unmodified HEVC encoder In the test, the search range was therefore set to correspond with the size of the PU.

A. Algorithm description

For each prediction unit (PU), for which the motion estimation is performed during the encoding process, the search range is selected in the following way.

1. In variant A1, the search range was set to be equal to the maximum from the dimensions of the PU. This means that the PU of the size of 16x16 pixels had search range set to ±16 points, and the PU of the size of the 32x16 pixels had the search range set to ±32 points.
2. In variant A2, the search range was set to be equal to the minimum from the dimensions of the PU. This means that the PU of the size of 16x16 pixels had search range set to ±16 points, and the PU of the size of the 32x16 pixels had the search range set to ±16 points.

B. Results

The results obtained using those two strategies are presented on Fig. 5 and 6, averaged over all of the tested sequences. Time reduction factor is presented in Tables 1 and 2. For the Fig. 5 and 6, four encoding runs were performed for the QP parameter equal to 22, 27, 32, 37.

The aggregated results are presented on Fig. 7. Those results are expressed in terms of the bitrate increase in the Bjontegaard metric, as described in [5], versus the encoding time change. For comparison purposes, also the results that were obtained using a constant search range smaller than the default ±64 x ±64 are presented. In this case, every block, regardless of its size, had the motion vector estimated within the same constant search range. There are a few interesting observations concerning those results.

First of all, the time for search range reduced to ±32 x ±32 leads to the increase of encoding time. This may be caused by less accurate motion estimation that causes the algorithms to predict and choose worse starting points and to extend the time spent on motion estimation. This increase of time is not compensated by the decrease of the time spent during motion compensation on comparisons for each block and each tested vector value. Another explanation would be that more time is spent on other stages when the estimated motion vector is worse than optimal one obtained for reference case (search range ±64 x ±64).

The same observation holds true for the case with search range set to ±4 x ±4 pixels. For this case a massive 12% bitrate increase (in terms of Bjontegaard metric) can be observed with only a slight decrease of compression time (by about 0,7%).

Another observation is that both methods, A1 and A2, outperform other tested scenarios either in terms of encoding time reduction or bitrate increase.

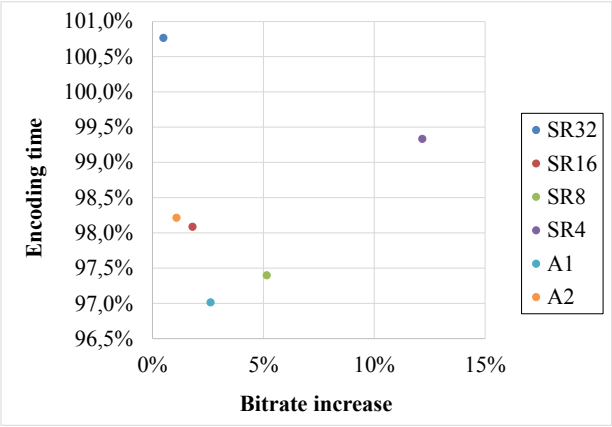


Fig. 7. Encoding time (relative to default HEVC value of ±64 x ±64) and bitrate increse (relative to ±64 x ±64) for different search range establishing algorithm.

TABLE I. TIME REDUCTION FOR SCENARIO A1

QP	Compression time relative to original HEVC
22	94.8%
27	97.2%
32	97.4%
37	96.9%
Average:	96.6%

TABLE II. TIME REDUCTION FOR SCENARIO A2

QP	Compression time relative to original HEVC
22	94,0%
27	96,2%
32	100,2%
37	99,3%
Average:	97.4%

V. CONCLUSIONS

The results presented in the paper show, quite surprisingly, that the reduction of time used for motion estimation does not significantly reduce the overall compression time. This may be caused by the fact that choosing suboptimal motion vectors has a side effect of decreasing the performance of the further stages of video encoding. The stages like choosing the transform block size and estimating the number of bits may perform worse, since they must process more data (since the prediction signal is worse, and, therefore, the error signal undergoing the coding process, is larger and more complex).

Another important conclusion is that the starting point selection for motion estimation in HEVC might be further improved in order to provide faster compression without a significant loss in compression efficiency and reconstructed image quality.

REFERENCES

[1] G. Correa, P. Assuncao, L. Agostini, L. Cruz, Performance and Computational Complexity Assessment of High-Efficiency Video Encoders, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, December 2012.

[2] Xiu-li T., Univ H., Sheng-kui D., An analysis of TZSearch algorithm in JMVC, International Conference on Green Circuits and Systems (ICGCS), Shanghai, 21-23 June 2010, s. 516 – 520.

[3] Bossen F.. 2012. „Common test conditions and software reference configurations". Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-J1100.

[4] J. Stankowski, D. Karwowski, K. Klimaszewski, K. Wegner, O. Stankiewicz, T. Grajek, Analysis of the Complexity of the HEVC Motion Estimation, 23rd International Conference on Systems, Signals and Image Processing (IWSSIP), Bratislava, 23-25 May, 2016.

[5] Bjøntegaard G.. 2001. „Calculation of average PSNR differences between RD curves". ITU-T SG16/Q6, Doc. VCEG-M33.

3D-HEVC EXTENSION FOR CIRCULAR CAMERA ARRANGEMENTS

Jakub Stankowski, Łukasz Kowalski, Jarosław Samelak,
Marek Domański, Tomasz Grajek, Krzysztof Wegner

Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics

ABSTRACT

The paper presents an extension of 3D-HEVC for the circular camera arrangements. It generalizes the derivation of the disparity vectors from depth data for sequences captured using cameras located on an arc. The general equation for disparity calculations has been implemented instead of the simplified equation used in 3D-HEVC. Experiments have been performed on widely recognized multiview test sequences with both circular and linear camera arrangements. In the case of sequences obtained using cameras located on an arc, the proposed extension provides about 6% bitrate reduction when compared to 3D HEVC, and does not influence the 3D-HEVC performance for linear view arranged sequences.

Index Terms **—** 3D-HEVC, FTV, multiview compression, non-linear camera arrangements, circular, arc camera setup

1. INTRODUCTION

Currently 3D-HEVC [1] is the state-of-the-art compression technology for 3D video in "multiview video and depth" format (MVD) [2]. This technology has been developed by Collaborative Team on 3D Video Coding Extensions Development (JCT-3V) formed between ISO/IEC and ITU-T and is foreseen to be incorporated into the HEVC standard (High Efficiency Video Coding) as Annex I of ISO/IEC MPEG-H Part 2 and ITU-T Recommendation H.265.

The 3D-HEVC is built on the top of the MV-HEVC [1] codec. The MV-HEVC utilizes inter-view prediction between the views and no information about scene structure (depth) is used. The MV-HEVC simply uses side views as another source for inter-prediction (in the same way as previous fames are used for motion compensated inter-prediction).

The goal of the 3D-HEVC was to exploit the information about 3D scene structure (in form of depth maps) to increase coding efficiency of 3D video. During the development of the 3D-HEVC technology, explicit 1D parallel views arrangement has been assumed. Therefore, the 3D-HEVC coding efficiency for non-linear camera arrangements is far from optimal.

Many modern Super Multiview (SMV) displays require circular (arc) view arrangements for better user experience. Therefore, efficient compression technology for non-linear (e.g. arc) camera arrangements is of great interest.

In this paper we present extension of 3D-HEVC towards arbitrary view setup, in particular the arc view arrangement.

2. GENERALIZATION OF 3D-HEVC TOOLS

Inter-view prediction which is the core tool of MV-HEVC, in general, does not assume any particular view arrangement. But, in order to increase the coding efficiency of 3D-HEVC, a number of specific tools was introduced and others were simplified with explicit 1D parallel view arrangement assumption.

The most significant simplification states that the views are vertically aligned. Therefore, disparity vectors are restricted only to horizontal direction, and they can be derived from depth data through simple linear equation.

$$\begin{aligned} d_h &= a \cdot v + b \\ d_v &= 0 \end{aligned} \tag{1}$$

where the horizontal component d_h of the disparity vector is calculated from depth sample value v and scaled to disparity by using scale factor a and offset b transmitted for each view, and vertical component of disparity vector d_v is always set equal to zero.

In the proposed 3D-HEVC extension, we have removed restriction only for horizontal direction, and applied general derivation of disparity vector based on depth data.

Based on the point position x, y in the picture and associated depth sample value v, position of the corresponding point x_r, y_r in the reference view can be calculated. Then, derived disparity vector is simply the difference in the position of considered point and the corresponding point in the reference view

$$\begin{aligned} d_h &= x_r - x \\ d_v &= y_r - y \end{aligned} \tag{2}$$

The position of the corresponding point in the reference view can be derived through depth based projection of the point according to the equation (3)

$$\begin{bmatrix} z_r \cdot x_r \\ z_r \cdot y_r \\ z_r \\ 1 \end{bmatrix} = P_r \cdot P^{-1} \cdot \begin{bmatrix} z \cdot x \\ z \cdot y \\ z \\ 1 \end{bmatrix}, \tag{3}$$

where P_r, and P are projection matrices for both the reference view and the view being coded. We have defined projection matrices as a multiplication of intrinsic and extrinsic camera parameters that need to be transmitted for each view being coded in the bitstream.

$$\boldsymbol{P} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{R} \cdot \boldsymbol{T} \\ \boldsymbol{O}^T & 1 \end{bmatrix} \tag{4}$$

Necessary distance z of the considered point from camera of interest can be calculated based on the depth sample value v:

$$z = \left(\frac{v}{2^{DepthMapBitDepth}} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right)^{-1} \tag{5}$$

where Z_{near} and Z_{far} are depth maps normalization parameters transmitted in the bitstream, and $DepthMapBitDepth$ is bit depth of the depth sample value (typically 8 or 16 bits per sample is used).

Actual camera parameters for each view are a natural extension of the simplified camera parameters already used in 3D-HEVC. Beside already transmitted parameters: focal length

f_x, optical center c_x of the camera and position of the camera optical center along x axis T_x, we have to transmit additional intrinsic parameters: second focal length along vertical direction f_y, position of the optical center along second axis c_y, and extrinsic parameters: rotation matrix $\boldsymbol{R}$ and remaining coordinates of camera position T_y and T_z. T_y and T_z together with T_x create vector $\boldsymbol{T}$.

In the proposed 3D-HEVC extension we have modified disparity vector derivation process in such tools as Disparity Compensated Prediction (DCP), Neighboring Block Disparity Vector (NBDV), Depth oriented NBDV (DoNBDV), View Synthesis Prediction (VSP), Inter-view Motion Prediction (IvMP), Illumination Compensation (IC).

In the 3D-HEVC, contrary to the proposed extension, derived disparity vectors do not depend on the position in view being coded. For example, in the DoNBDV tool, the disparity for a given block is set to the value that corresponds with the maximum value of four corner depth samples value of virtual depth map block. In the proposed method, the disparity is calculated based on half of the maximum depth sample value and the position of selected corner of the block.

Considering circular (arc) view arrangements, a given depth sample value can represent a different distance plane for each view (which is not the case in 1D parallel view arrangement). In the proposed extension, corresponding depth sample value in the reference view can be derived simultaneously with the disparity vector through the equation (3). The equation (3) provides distance z_r of the correspondent point from the reference camera which can be normalized to depth sample value with the help of

$$v = \frac{\frac{1}{z} - \frac{1}{z_{far}}}{\frac{1}{z_{near}} - \frac{1}{z_{far}}} \cdot 2^{DepthMapBitDepth} \qquad (6)$$

The derivation of depth sample value has been applied to the synthesis of dependent views' depth maps.

3. EXPERIMENTS

Extensive experiments have been performed in order to evaluate the performance of proposed extension. Our improvements have been implemented in reference 3D-HEVC codec version 13 (HTM 13.0) [3]. For comparison purposes, unmodified 3D-HEVC codec and MV-HEVC codec have been used (also HTM 13.0).

All tests have been performed according to JCT-3V Common Test Conditions (CTC) [4], that describe default encoder configuration for evaluation of 3D codec performance used by International Organization for Standardization. Essential parameters of the encoders have been gathered in Table 1.

Table 1. Essential configuration parameters of the encoders used in the experiments.

Parameter	Value
Profile	Main
GOP size	8
Intra period	24
Hierarchical GOP	Yes
Number of referenced frames	4
Rate-distortion optimization	On
Search range for motion and disparity estimation	±64
Neighboring Block Disparity Vector (NBDV)	On
Depth oriented NBDV (DoNBDV)	On
View Synthesis Prediction (VSP)	On
Inter-view Motion Prediction (IvMP)	On
Illumination Compensation (IC)	On
View synthesis optimization for depth coding (VSO)	Off

As the 3D-HEVC is designed to support only linear camera arrangement, in the case of the encoding sequences with circular (arc) camera arrangement, positions of the cameras were set as distances from the base view. This way, a rough approximation of proper scaling of the disparity vector can be done by the encoder.

We have evaluated the proposed extension with three commonly recognized multiview test sequences with the arc arrangement of the cameras, namely: *Poznan Blocks* [5], *Ballet* and *Break Dancers* [6]. All sequences were provided with high quality depth data for all of the views.

For compatibility purposes, we have also tested our codec against sequences with linear arrangement of the cameras. For that purpose we have used six multiview sequences with depth maps that are recommended by the MPEG for 3D video codec evaluations (*Poznan Street*, *Poznan Hall 2* [7], *Dancer* [8], *Balloons*, *Kendo* [9], *Newspaper* [10]).

Exemplary images from used test sequences have been depicted in figure 1.

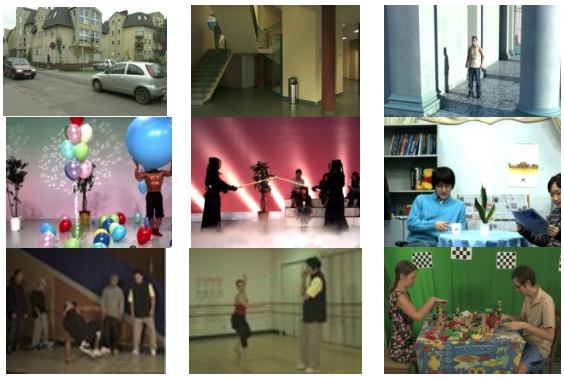


Figure 1. Exemplary pictures from multiview test sequences. From top-left: Poznan Street, Poznan Hall 2, Dancer, Baloons, Kendo, Newspaper, Breakdancer, Ballet, Poznan Blocks.

Three views and three corresponding depth maps of each sequence (see Table 2) have been encoded with proposed codec, original unmodified 3D-HEVC and MV-HEVC at 4 QP values (25, 30, 35, 40). During tests, the total bitrate of all views and all depth maps and average luminance PSNR of all views have been calculated.

Table 2. View positions of each sequences used in the experiment

Sequence	Coded views	View angle
Poznan Street	4, 5, 3	N/A (linear)
Poznan Hall 2	6, 7, 5	N/A (linear)
Dancer	5, 1, 9	N/A (linear)
Balloons	4, 3, 5	N/A (linear)
Kendo	3, 1, 5	N/A (linear)
Newspaper	5, 4, 6	N/A (linear)
Poznan Blocks	5, 4, 6	11 degree
Ballet	4, 3, 5	4 degree
Breakdancers	4, 3, 5	4 degree

4. RESULTS

The averaged luminance PSNR values and the corresponding total bitrates for all test sequences for three analyzed codecs have been gathered and presented in Table 3. RD curves for exemplary sequences with linear (Poznan Street) and arc (Poznan Blocks) camera arrangements have been shown in figures 1 and 2 respectively.

Table 3. Experimental results for all test sequences for analyzed codecs.

Sequence	QP	MV-HEVC		3D-HEVC		Proposed	
		Bitrate [kbit/s]	PSNR [dB]	Bitrate [kbit/s]	PSNR [dB]	Bitrate [kbit/s]	PSNR [dB]

Dancer	25	1389	41.88	1329	41.88	1329	41.88
	30	587	40.90	549	40.91	549	40.91
	35	307	39.48	281	39.52	281	39.52
	40	173	37.69	154	37.77	154	37.77
Poznan Hall2	25	3970	39.24	3849	39.25	3849	39.25
	30	1422	37.30	1357	37.33	1357	37.33
	35	653	35.33	620	35.37	620	35.37
	40	330	33.29	312	33.34	312	33.34
Balloons	25	6686	38.28	6458	38.30	6458	38.30
	30	2812	35.57	2671	35.60	2671	35.60
	35	1295	33.12	1205	33.16	1205	33.16
	40	617	30.85	558	30.91	558	30.91
Newspa-per	25	1587	42.84	1422	42.88	1422	42.88
	30	807	40.56	710	40.65	710	40.65
	35	452	38.11	394	38.22	394	38.22
	40	264	35.52	229	35.67	230	35.67
Poznan Street	25	1647	42.36	1466	42.42	1466	42.42
	30	855	40.11	754	40.22	754	40.22
	35	481	37.49	423	37.64	423	37.64
	40	277	34.70	246	34.88	246	34.88
Kendo	25	1710	40.49	1614	40.53	1614	40.52
	30	845	38.12	788	38.16	788	38.16
	35	457	35.64	421	35.68	420	35.68
	40	260	33.12	238	33.18	238	33.18
Poznan Blocks	25	2761	43.41	2736	43.34	2655	43.42
	30	1520	40.82	1509	40.74	1446	40.84
	35	868	38.05	859	37.94	813	38.05
	40	496	35.18	491	35.08	458	35.21
Ballet	25	936	41.62	892	41.58	871	41.63
	30	476	40.12	450	40.04	438	40.12
	35	267	38.18	250	38.07	242	38.18
	40	152	35.89	143	35.82	138	35.93
Break Dancers	25	2352	39.17	2216	39.13	2162	39.16
	30	964	37.71	890	37.67	860	37.72
	35	491	36.02	457	35.98	437	36.05
	40	261	34.10	246	34.08	233	34.15

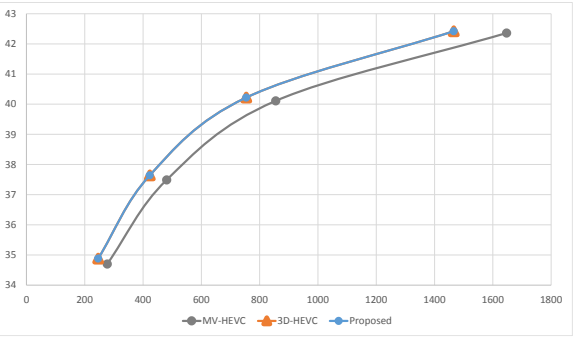


Figure 1. RD curves for Poznan Street sequence (linear camera arrangement) for MV-HEVC, 3D-HEVC and proposed codecs.

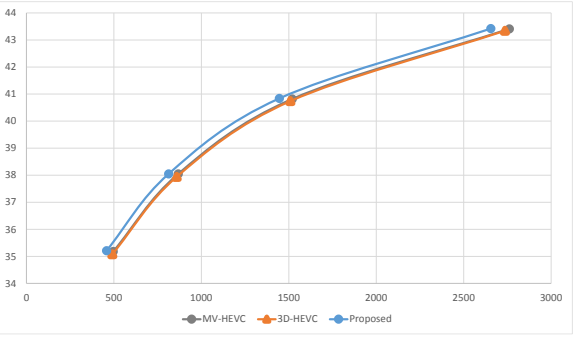


Figure 2. RD curves for Poznan Blocks sequence (arc camera arrangement) for MV-HEVC, 3D-HEVC and proposed codecs.

Table 4. Average bitrate changes calculated as Bjøntegaards rates for PSNR Y introduced by the proposed codec against 3D-HEVC and MV-HEVC for sequences with linear and arc view arrangements.

	Sequence	BD-Rate [%]

View arrange-ment		Proposed vs 3D-HEVC	Proposed vs MV-HEVC	3D-HEVC vs MV-HEVC
Linear	Poznan Street	0,05	-7.10	-7.11
	Poznan Hall 2	0,09	-9.14	-9.18
	Dancer	0,02	-14.34	-14.36
	Balloons	0,08	-8.21	-8.20
	Kendo	0,09	-6.05	-6.08
	Newspaper	0,04	-14.26	-14.28
	Average	**0,06**	**-9.85**	**-9.87**
Arc	Poznan Blocks	-6,17	-11.11	-5.28
	Ballet	-5,68	-8.82	-3.35
	Break Dancers	-6,89	-5.94	1.02
	Average	**-6,25**	**-8.62**	**-2.54**

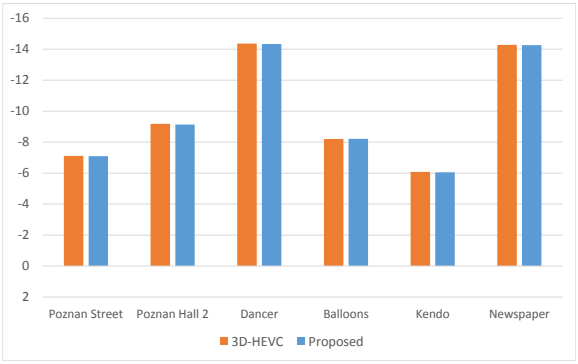


Figure 3. Bitrate reduction (Bjøntegaards rates for PSNR Y) against MV-HEVC for sequences with linear view arrangement

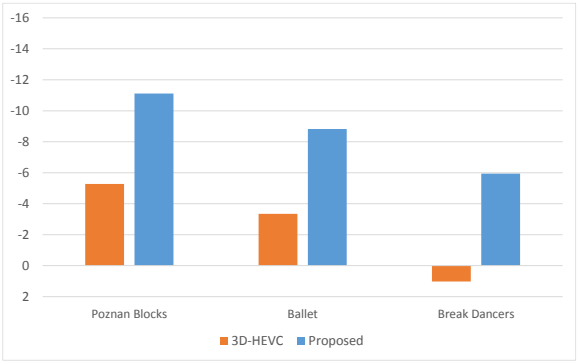


Figure 4. Bitrate reduction (Bjøntegaards rates for PSNR Y) against MV-HEVC for sequences with arc view arrangement

In order to assess the coding performance of the proposed codec, commonly used Bjøntegaard delta has been calculated [11] based on the data presented in Table 3. Achieved results have been gathered in Table 4 and depicted in figures 3 and 4. The sequences have been divided by the view arrangement (linear and arc).

For circular (arc) camera setup, our improved codec allows for about 6% bitrate reduction in comparison with 3D-HEVC and over 8% bitrate reduction in the case of MV-HEVC.

For linear (parallel) camera setup for which 3D-HEVC was designed, proposed modifications caused negligibly small bitrate increase while preserving the same quality of encoded video – on average 0.06%. Due to the fact that the proposed codec has been built on the top of the 3D-HEVC, the achieved bitstream reduction against MV-HEVC is almost the same as for 3D-HEVC against MV-HEVC (about 9.9%).

The small performance drop for sequences with linear view arrangement is caused mainly by the necessity of transmission of full camera parameters while in original 3D-HEVC only simplified version of camera parameters is used and transmitted.

Therefore, our modifications do not only deteriorate the quality of encoded video for linear camera setting, but also lead to a significant improvement in compression efficiency for sequences with arc arranged cameras.

5. CONCLUSIONS

In the paper, modifications of the 3D-HEVC codec, that allow to utilize knowledge of the arrangement of camera setup (in our case – arc camera setup), have been presented. The performance of the proposed new codec has been compared to the state-of-the-art codec 3D-HEVC (HTM version 13.0) and MV-HEVC. The performed experiments showed that the proposed modifications led on average to over 6% bitrate reduction for video sequences recorded with the arc arranged camera setup comparing to the unmodified 3D-HEVC codec. Whereas, for video sequences with linear camera setup, the observed bitrate change is negligibly small.

The obtained results show that there is still room for further improvement in compression efficiency of the 3D-HEVC standard, especially by relaxing explicit constraints on camera setup.

6. ACKNOWLEDGEMENT

Research project was supported by National Science Centre, Poland, according to the decision DEC 2012/05/B/ST7/01279.

7. REFERENCES

[1] G.J. Sullivan,, J.M. Boyce, Y. Chen, J.-R. Ohm, C.A. Segall, A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)", *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, Dec 2013.

[2] Müller K., Merkle P., and Wiegand T., "3D Video Representation Using Depth Maps", *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[3] 3D HEVC reference codec available online https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-13.0.

[4] K. Müller, A. Vetro, "Common Test Conditions of 3DV Core Experiments" *Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 7th Meeting: Doc. JCT3V-G1100*, San José, US, Jan. 2014.

[5] M. Domański, A. Dziembowski, A. Kuehn, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz, K. Wegner, „Poznan Blocks – a multiview video test sequence and camera parameters for Free Viewpoint Television" *ISO/IEC JTC1/SC29/WG11 Doc. M32243*, San Jose, USA, Jan. 2014.

[6] C.L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Transactions on Graphics, vol. 23, no. 3. pp. 600-608, Aug. 2004.

[7] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań multiview video test sequences and camera parameters," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M17050*, Xian, China, Oct. 2009.

[8] D. Rusanovskyy, P. Aflaki, M.M. Hannuksela, "Undo Dancer 3DV sequence for purposes of 3DV standardization," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M20028*, Geneva, Switzerland, Mar. 2011.

[9] M. Tanimoto, T. Fujii, N. Fukushima, "1D parallel test sequences for MPEG-FTV," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15378*, Archamps, France, Apr. 2008.

[10] Y.S. Ho, E.K. Lee, C. Lee, "Multiview video test sequence and camera parameters," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15419*, Archamps, France, Apr. 2008.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," *ITU-T SG16, Doc. VCEG-M33,* Austin, USA, Apr. 2001.

EXTENSIONS OF THE HEVC TECHNOLOGY FOR EFFICIENT MULTIVIEW VIDEO CODING

Jakub Stankowski, Marek Domański, Olgierd Stankiewicz, Jacek Konieczny, Jakub Siast, Krzysztof Wegner

Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poznań, Poland

ABSTRACT

The paper deals with multiview video coding using the new technology of High Efficiency Video Coding (HEVC). Implementation of multiview video coding in the framework of HEVC is described together with new specific tools proposed by the authors. Extensive experimental results are reported for compression performance comparison of MVC (ISO 14496-10), HEVC simulcast and two versions of proposed "multiview HEVC". For "multiview HEVC" the results indicate significant bitrate reduction of about 50%, as compared to the state-of-the-art MVC technology standardized as a part of AVC (MPEG-4, H.264).

Index Terms— HEVC, MVC, inter-view prediction, multiview coding

1. INTRODUCTION

Regarding 3D video, two basic multiview video formats are considered for transmission and compression:

a) multiview video with depth maps (MVD),
b) multiview video with no depth map.

Currently, the MPEG (ISO/IEC expert group) is working towards an international standard related to the first approach, i.e. coding of multiview video along with the respective depth maps. For the second format, i.e. multiview video with no additional depth information, there already exists a technique called Multiview Video Coding (MVC) that is incorporated into Advanced Video Coding (AVC) compression standard of ISO and ITU [1]. The MVC technique takes advantage of inter-view redundancy and provides bitrates reduced mostly by 15-25% as compared to independent compression of the views using AVC (called also as "simulcast AVC").

On the other hand, a new generic video compression technology called High Efficiency Video Coding (HEVC) is under standardization by ISO and ITU [2]. HEVC provides almost halved bitrates as compared to AVC. Therefore even independent compression of all views using HEVC (simulcast HEVC) appears to be significantly more efficient than the older but dedicated MVC technique [7]. In this paper, we are going to study this issue in details. Therefore, two main problems are addressed in this paper:

- Compression performance of "multiview HEVC" obtained by direct adoption of multiview coding tools into HEVC.
- Potential improvements of "multiview HEVC" by extensions of the prediction modes already used in HEVC.

For the sake of conciseness, in the considerations, the case of 3 views will be treated. Nevertheless the conclusions hold also for other numbers of view, with the reservation that the numbers referred to the compression gain against MVC may be somewhat different for other numbers of views.

2. MULTIVIEW HEVC (MV-HEVC)

The classic Multiview Video Coding (MVC) [1] uses inter-view predictions in addition to the standard inter-frame prediction as used in the generic Advanced Video Coding (AVC)[1]. In MVC, the pictures from other views but the same time instant may be taken also as the reference pictures. This is done by inserting additional reference pictures into the reference picture lists. This additional inter-view prediction improves the compression efficiency as compared to independent coding of the views using the generic AVC technique. In particular, when I-frames are replaced by P- or B-frames by the use of this additional prediction, the compression improvement is particularly high.

The HEVC codec structure is quite similar to that of AVC. Therefore, the idea of multi-view coding can be easily adopted to HEVC [3]. We call the new codec Multiview-HEVC (MV-HEVC). The adoption of multiview coding tools is nearly straightforward. For MV-HEVC, the difference is, that except of I-pictures in the base view, all other pictures may be B-pictures. In Fig. 1, the boxes marked in gray denote the B-pictures in MV-HEVC that would be P-pictures in MVC.

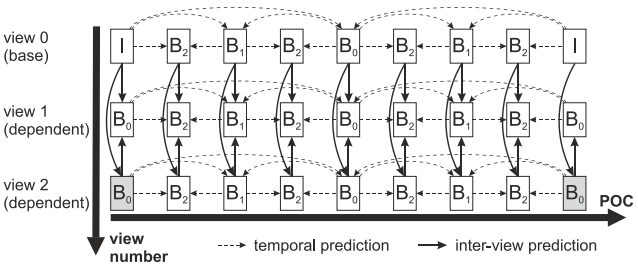


Figure 1. Temporal and inter-view prediction scheme.

3. IMPROVED PREDICTION MULTIVIEW-HEVC (IPMV-HEVC)

In the just proposed MV-HEVC, there still exists a potential for improvement of the compression performance. In the abovementioned proposal of MV-HEVC, inter-view predictions were used for texture, while now we are going to consider also modifications of motion data prediction. The MV-HEVC with such prediction tools will be called Improved Prediction Multiview HEVC (IPMV-HEVC).

In the generic HEVC, very efficient Advanced Motion Vector Prediction (AMVP) is used. The idea of AMVP is to use some predefined rules to select candidate motion vectors from the

neighboring prediction units in the same frame or from prediction units in the reference frame. These selected motion vectors and the related information like the identifier of the reference frame linked to the vector are the prediction candidates. The encoder selects one prediction candidate and transmits its identifier in the bitstream.

In this section, we describe the modification of prediction candidate derivation that exploits inter-view dependencies. The proposed modification is useful for both coding modes of AMVP: motion vector "competition" and motion data "merging".

3.1. Vector Scaling

The motion vector scaling is used for deriving AMVP candidates. In the original HEVC design, there is temporal scaling only. A scaled vector for target (t) prediction unit is calculated from source (s) prediction unit vector. Scaling operation is performed as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} \cdot \frac{t_t - t_{t,ref}}{t_s - t_{s,ref}}$$

where: x_t, y_t are the scaled vector components for target prediction unit; x_s, y_s are the vector components for source prediction unit; t_t and t_s are the values of Picture Order Count (POC) for the target and the source prediction unit, respectively; $t_{t,ref}$ and $t_{s,ref}$ are POC of the picture referenced by the target and the source unit, respectively. In the case of scaling of vectors derived from neighboring units, t_t is equal to t_s. The motion vector scaling is illustrated in Figure 2. Vectors from neighboring units (red, green, blue) are used to derive prediction candidates for the grey unit. Vectors for the current unit (grey), pointing to a reference frame (darker, POC=1) are searched for. Green neighbor vector points to the same reference picture and can be directly copied, but red and blue neighbor vectors point to different POCs and have to be scaled.

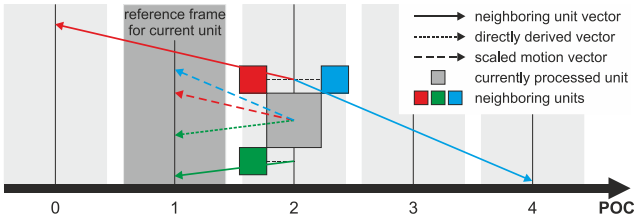


Figure 2. Ilustration of motion vector scaling.

In IPMV-HEVC, for the case where reference pictures from different views are used, we propose inter-view scaling for disparity vectors. Inter-view scaling allows to use disparity vector pointing to the reference picture from a different view as a prediction candidate. The inter-view scaling is performed as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} \cdot \frac{v_t - v_{t,ref}}{v_s - v_{s,ref}}$$

where: x_t, y_t are the scaled vector components for target unit; x_s, y_s are vector components for source prediction unit; v_t and v_s are the view numbers for the target and the source prediction unit, respectively; $v_{t,ref}$ and $v_{s,ref}$ are view numbers of the picture referenced by the target and the source unit, respectively.

3.2. Colocated Units

In HEVC, a prediction candidate can be derived from the colocated units in the reference frame. This is an additional candidate for temporal motion data prediction. The derivation of a motion vector from colocated unit is illustrated in Figure 3. The colocated unit (orange) is located in the first image from the reference list. Next, the motion vector from the colocated unit motion vector is scaled for current unit and the reference frame POCs.

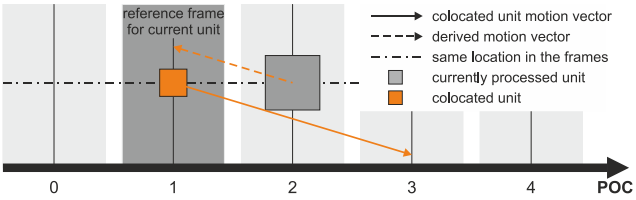


Figure 3. Derrivation of a motion vector from the colocated unit.

In IPMV-HEVC, in order to increase multiview coding efficiency, the inter-view support for colocated units is proposed. There are four possibilities of using colocated units in a coder:

- derive the motion vector from the colocated unit in the temporal reference frame (the only possibility in HEVC and MV-HEVC),
- derive the disparity vector from the colocated unit in the temporal reference frame,
- derive the disparity vector from the colocated unit in the inter-view reference frame,
- derive the motion vector from the colocated unit in the inter-view reference frame.

Depending on the choice of:

- the reference frame selected for the currently processed unit,
- the frame were the colocated units is located,

the encoder choses one of four ways to derive the colocated prediction candidate.

For example, consider the last of these four. Figure 4. shows the process of temporal motion vector derivation from the colocated unit in the inter-view reference frame. This process is used when the colocated unit is taken from inter-view reference frame but the temporal motion vector is required.

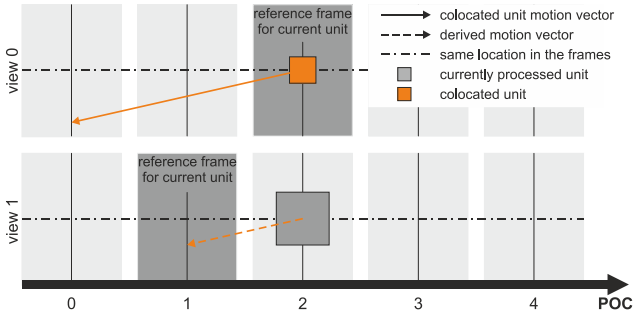


Figure 4. Derivation of a motion vector from the colocated unit in an inter-view reference frame (in IPMV-HEVC).

3.3. Nested Prediction

Nested Prediction (NP) is a new tool introduced to improve prediction of vectors of motion and disparity. The aim of this tool is to avoid situation, when there is no possibility to derive candidates from the neighboring units. The NP is useful in two following cases:

- The currently processed unit is encoded with the use of the reference image in the same view but all neighboring units have their reference frames from side views. Therefore, the motion (temporal) vector is necessary for prediction, but only

disparity (inter-view) vectors are available in neighboring units.

- The currently processed unit is encoded with the use of the reference image in the side view but all neighboring units have their reference frames from same view. Therefore, the disparity (inter-view) vector is necessary for prediction, but only motion (temporal) vectors are available in neighboring units.

Unfortunately, disparity vectors provide no information on motion and motion vectors provide no information on disparity. Therefore, in that case, the coder is unable to derive prediction candidates from neighboring units. In both described cases, the NP can provide a prediction candidate and fill in the sparse prediction candidate list.

The idea behind NP is to track the neighboring unit vector and identify the unit pointed by the neighboring unit vector. Once identified the pointed unit can be used as source of prediction candidate vectors.

Example of NP is shown in Figure 5. The disparity vector of neighboring unit (red) points to another unit (light red) in the reference frame. If the pointed unit has an appropriate motion vector (red dotted vector), the pointed unit vector can be scaled and used as a prediction candidate. In case shown on Figure 5, the NP is able to provide indirectly derived prediction candidate and increase the number of available prediction candidates.

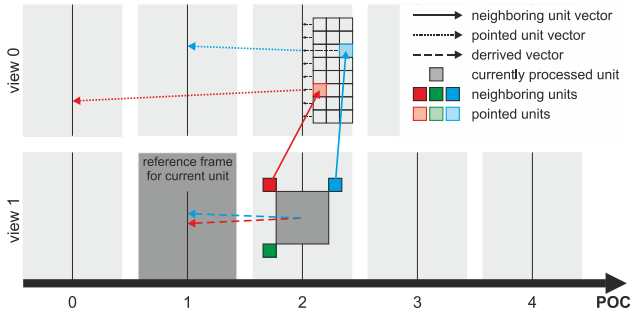


Figure 5. Example of nested prediction (NP).

4. PERFORMANCE EVALUATION AND EXPERIMENTAL RESULTS

The performance evaluation was done under the conditions described in the Call for Proposals (CfP) on 3D Video Coding (3DVC) [4] and in Common Test Conditions [5]. The CfP defines two coding scenarios: 2-view and 3-view case and gives a set of multiview sequences. The test sequences are divided into two classes: Class A – FullHD resolution (Poznan_Hall2, Poznan_Street, Undo_Dancer, GT_Fly) and Class C – XGA resolution (Kendo, Balloons, Lovebird1, Newspaper). The same coding scenarios and test sequences were used to evaluate the performance of responses to CfP on 3D Video Coding.

In both 2- and 3-view cases, the first dependent view was encoded using the base view as a reference. In the 3-view case, the second dependent view was encoded using both the base view and first dependent view as a reference. The available side views served as a reference for both anchor and non-anchor frames.

All the results are gathered for coding conditions conforming random access high efficiency setup (RA-HE) [5]. The BD-rate gain [6] has been calculated.

General comparison between MVC, simulcast HEVC and MV-HEVC are shown in Table 1. Comparison between simulcast HEVC and MVC shown, that simulcast HEVC can provide significantly higher compression than MVC, despite not exploiting inter-view dependencies. Those results are similar to described in [7]. When compared to simulcast HEVC, the multiview HEVC (MV-HEVC) can provide about 22% and 30% bitstream reduction for 2- and 3-view case, respectively. The inter-view texture prediction applied to HEVC has similar efficiency to that in MVC.

The experimental results for IPMV-HEVC are collected in Table 2. Coding results are shown in comparison to MV-HEVC coder [3]. Any of the described modification changes base view encoding process and there is no coding gain for base view. Therefore, Table 2. shows coding efficiency gains for dependent views only.

All proposed tools provide noticeable coding performance gain without increasing the encoding and decoding time. Inter-view disparity vectors scaling itself provides coding performance gain only for 3-view case because this tool can be only used if two inter-view reference images are available on the reference list. Once the inter-view colocated is added, the highest compression gain for a single tool is achieved. This tool is useful in both 2- and 3-view case. The Nested Prediction (NP) also provides coding performance gain for both scenarios. Applying all described modifications allow increasing coding efficiency more than any single modification. The results achieved for this setup shows that both main improvements (inter-view colocated units and NP) complement each other.

Table 1. Bitrate reductions for luma.

	HEVC simulcast vs. MVC	**MV-HEVC vs. HEVC simulcast**	**MV-HEVC vs. MVC**
	2-view case		
Class A	-36.0 %	-26.8 %	-54.5 %
Class B	-47.0 %	-18.6 %	-57.3 %
Overall	-41.5 %	-22.7 %	-55.9 %
	3-view case		
Class A	-30.4 %	-35.4 %	-57.3 %
Class B	-44.3 %	-25.4 %	-59.1 %
Overall	-37.4 %	-30.4 %	-58.2 %

Table 2. Bitrate reductions for variants of IPMV-HEVC vs. MV-HEVC.

	Inter-view Vector Scaling	**Inter-view Colocated Units**	**Nested Prediction**	**All tools used**
	2-view case – dependent view			
Class A	0.00 %	-0.51 %	-0.11 %	-0.44 %
Class B	0.00 %	-0.67 %	-0.44 %	-0.94 %
Overall	0.00 %	-0.59 %	-0.28%	-0.44 %
	3-view case – 1st dependent view			
Class A	0.00 %	-0.77 %	-0.61 %	-1.08 %
Class B	0.00 %	-0.32 %	-0.38 %	-0.63 %
Overall	0.00 %	-0.55 %	-0.50 %	-0.86 %
	3-view case – 2nd dependent view			
Class A	-0.49 %	-1.76 %	-0.32 %	-2.14 %
Class B	-0.33 %	-1.48 %	-0.89 %	-2.02 %
Overall	-0.41 %	-1.62 %	-0.60 %	-2.08 %

5. CONCLUSIONS

In this paper, the multiview HEVC (MV-HEVC) and improvement prediction for multiview HEVC (IPMV-HEVC) was proposed and evaluated. The MV-HEVC inter-view texture

prediction allows to achieve coding gain similar or even better than MVC coding gain. The new improvements of MV-HEVC for motion data prediction have been proposed. The IPMV-HEVC provides an additional bitstream reduction for dependent views without a significant increase in complexity.

Acknowledgement: This work was supported by the public funds as a research project.

6. REFERENCES

[1] ISO/IEC 14496-10, International Standard, "Generic coding of audio-visual objects – Part 10: Advanced Video Coding," 6nd Ed., 2010, [also:] ITU-T Rec. H.264, Edition 5.0, 2010.

[2] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G1103, B. Bross, W.-J. Han, , J.-R. Ohm, G. J. Sullivan, T. Wiegand (Eds.), "High Efficiency Video Coding (HEVC) text specification Working Draft 5", Geneva, November 2011.

[3] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "Multiview HEVC – experimental results" JCT-VC (MPEG/VCEG) Doc. JCTVC-G582, Geneva, November 2011.

[4] Call for Proposals on 3D Video Coding Technology, ISO/IEC JTC1/SC29/WG11, MPEG2011/N12036, Geneva, March 2011.

[5] F. Bossen, "Common test conditions and software reference configurations", JCT-VC (MPEG/VCEG) Doc. JCTVC-F900, Torino, July 2011.

[6] G. Bjøntegaard "Improvements of the BD-PSNR model," ITU-T SG16 Q.6 Document, VCEG-AI11, Berlin, July 2008.

[7] K. Wegner, O. Stankiewicz, K. Klimaszewski, M. Domański "Comparison of multiview compression performance using MPEG-4 MVC and prospective HVC technology", ISO/IEC JTC1/SC29/WG11 MPEG M17913, Geneva, July 2010

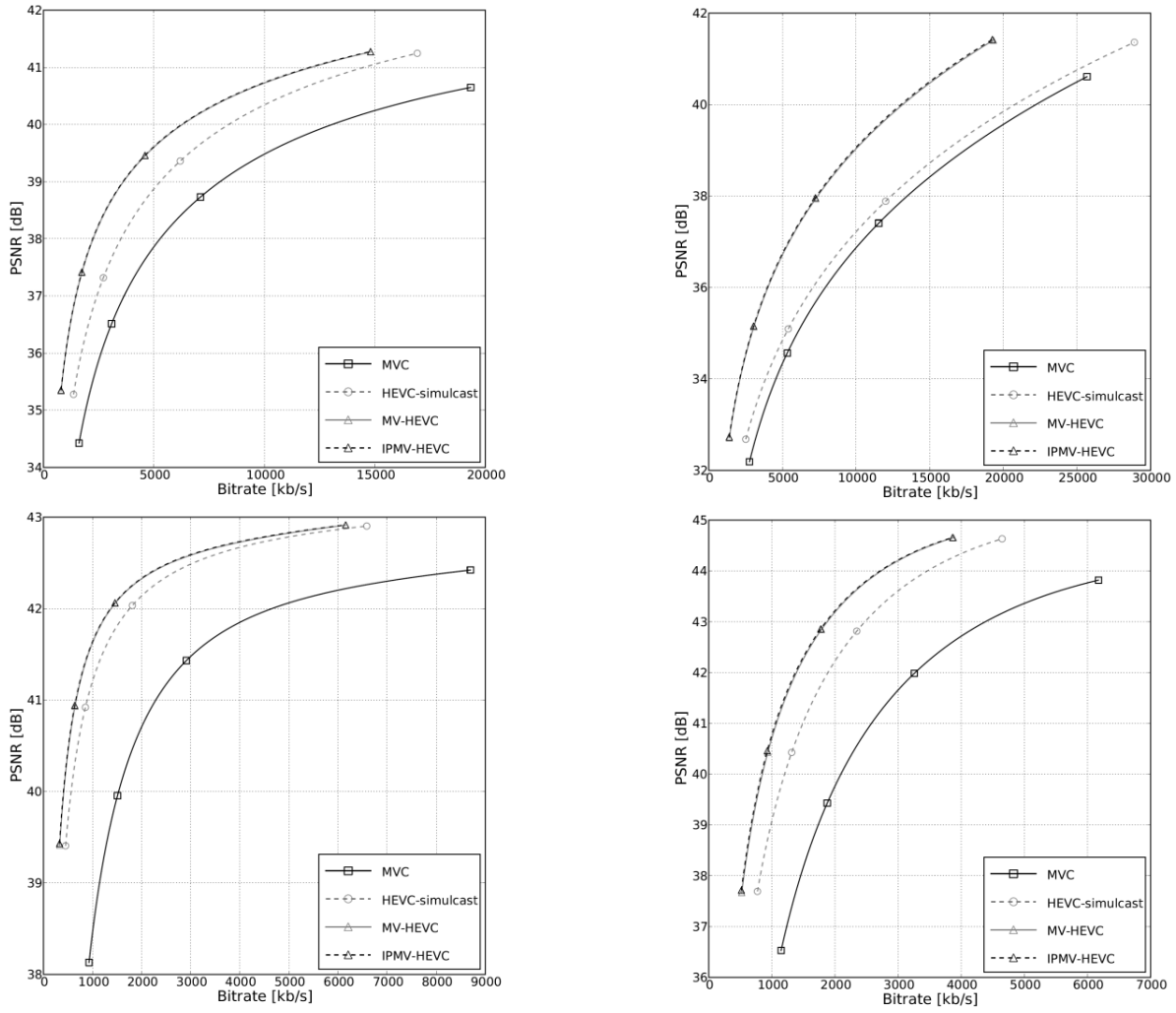


Figure 6. Experimental curves for "Poznan_Street" (left-top), "Poznan_Hall2" (right-top), "Dancer" (left-bottom), "Baloons" (right-bottom) sequences.

High Efficiency 3D Video Coding Using New Tools Based on View Synthesis

Marek Domański, *Member, IEEE*, Olgierd Stankiewicz, Krzysztof Wegner, Maciej Kurc, Jacek Konieczny, Jakub Siast, Jakub Stankowski, Robert Ratajczak, Tomasz Grajek

***Abstract*—We propose a new coding technology for 3D video represented by multiple views and the respective depth maps. The proposed technology is demonstrated as an extension of the recently developed High Efficiency Video Coding (HEVC). One base view is compressed into a standard bitstream (like in HEVC). The remaining views and the depth maps are compressed using new coding tools that mostly rely on view synthesis. In the decoder, those views and the depth maps are derived via synthesis in the 3D space from the decoded base view and from data corresponding to small disoccluded regions. The shapes and locations of those disoccluded regions can be derived by the decoder without any side information transmitted. In order to achieve high compression efficiency, we propose several new tools like Depth-Based Motion Prediction, Joint High Frequency Layer Coding, Consistent Depth Representation and Nonlinear Depth Representation. The experiments show high compression efficiency of the proposed technology. The bitrate needed for transmission of even two side views with depth maps is mostly below 50% of the bitrate for a single-view video.**

***Index Terms*—3D video, coding, compression, MVD representation, HEVC, depth maps.**

I. Introduction

Recently, 3D video technology is evolving towards new systems that include glassless displays and provide realistic impression of depth as well as controllable stereoscopic base-line distance. In such 3D video systems the representation of a 3D scene should be richer than a stereo pair. Therefore, a need has been identified for new video formats and for a respective compression standard. This need was discussed in the ISO/IEC Moving Pictures Experts Group (MPEG) [1], which decided to start the respective standardization project on compression of 3D video in the multiview plus depth format (the MVD format). The features, possibilities and importance of this format have been already discussed in many papers, e.g. [2-9], therefore these issues will be omitted in this paper, which is focused on compression of 3D video in this format.

Some applications will need numerous views to be available at the receiver. For example, future autostereoscopic displays are expected to present simultaneously even 50 different views corresponding to virtual cameras with parallel optical axes spaced within few human inter-ocular distances. Such dense spacing of the views yields strong similarity between the neighboring views that can be exploited for compression. Moreover, in the receiver many virtual views may be efficiently synthesized using the Depth-Image-Based Rendering (DIBR) [3,10,11] and for transmission the MVD format often may be limited to only 2-3 views accompanied with the corresponding depth maps [1]. In a realistic example of a system with an autostereoscopic display only 3 views with 3 depth maps are transmitted (Fig. 1).

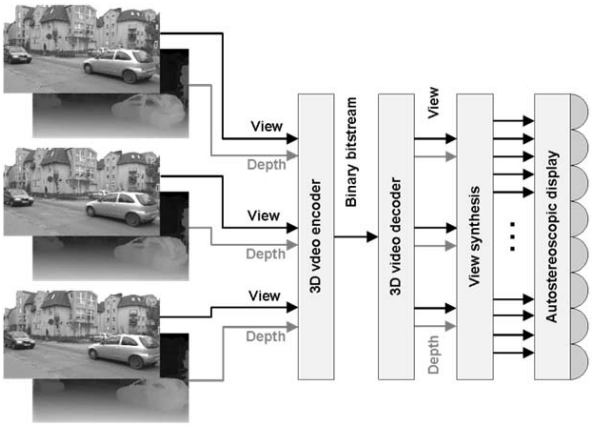


Fig. 1. An example of a 3D video system where 3 views with 3 depth maps are transmitted and used for synthesis of many virtual views.

MPEG has also identified the requirement that the prospective 3D video coding technology should exploit either Multiview Video Coding (MVC) or High Efficiency Video Coding (HEVC) standards [1]. The MVC standard [12] is an extension of the Advanced Video Coding (AVC) standard [13] and it takes advantage of additional inter-view disparity-compensated predictions and provides bitrate reduction of up to 20-30% as compared to the independent AVC-compliant compression of all views. HEVC is currently at the final stage of standardization by ISO and ITU [14]. It provides about 50% bitrate reduction as compared to AVC. Independent HEVC compression of multiple views results in 30% bitrate reduction as compared to MVC [15].

With the aim to search for a new technology that would be potentially used as a base for the prospective international

Manuscript received October 15, 2012; revised March 21, 2013. This work was supported by public funds in the years 2010-2012.

The authors are with the Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology, 60965 Poznań, Poland (e-mail: domanski@et.put.poznan.pl, ostank@multimedia.edu.pl, kwegner@multimedia.edu.pl).

Digital Object Identifier

standard, in early 2011 MPEG announced a Call for Proposals (CfP) on 3D Video Coding Technology [16]. Each proposed technology had to meet the condition that a standard bitstream (either AVC-compliant or HEVC-compliant) for a single view may be extracted from the entire bitstream.

In response to the CfP, over 20 proposals were submitted in two categories: backward compatible with AVC and compatible with HEVC. The proposals were ranked using subjective quality assessment of the decoded and rendered video clips. The assessments were done in a large experiment that involved 12 laboratories and over 600 viewing subjects who assessed about 2700 video clips. The results were disclosed during the MPEG meeting in Geneva in November 2011. In the HEVC-compatible class, the proposals from Fraunhofer Institute – H. Hertz Institute, Berlin, and from Poznań University of Technology, were qualified as the best performing ones.

The latter proposal exploits dependencies between views and depth maps in order to increase the overall coding efficiency (similarly as in [17-19]). The 3D scene information contained in depth maps was also already used to speed up the encoder mode decision [9]. Other approaches use special techniques for depth coding [7,8]. Some of them rely on independent coding of depth that allows to use the standard coding technology for the MVD format but provides limited compression efficiency.

In 2012, the standardization projects for MVD video ware developed by MPEG and by ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V). For MVC, independent depth coding [20] was included into a standard draft [21]. A more sophisticated MVC extension for MVD is also under development [22]. The MVC-like inter-view disparity-compensated prediction was also implemented on top of the HEVC technology. The adoption of MVC coding scheme to HEVC provides coding gains similar or better than MVC gains over AVC [23]. The respective standard draft is under development [24]. In the course of these standardization activities, some new tools described in this paper were also implemented in the test models [25,26] and used as a starting point for development of 3D video coding standards.

II. Assumptions and Overview of the Paper

The paper presents a detailed description of the compression technology proposed by Poznań University of Technology in the response to the call [16] of MPEG. The novelty of this proposal consists in new coding tools and in a the selection of the tools resulting from extensive experiments. Hitherto, this technology was only briefly described in an MPEG document [27] and in conference contributions [28-30] that concerned some selected aspects only. The aim of this paper is to present the entire proposed technology and the most important coding tools. Some tools have been already described in the conference papers [19,31]. For those tools only brief descriptions are given in this paper; more details can be found in the references. This paper also presents the new extensive experimental results obtained by the authors.

The proposed codec is compliant with the requirements that were defined by MPEG in the Call for Proposals (CfP) [1,16], and which resulted from studies of potential applications. In particular, one of the views – called the base view – is coded in compatibility with HEVC syntax, which allows extraction of a base view by a legacy decoder (Fig. 2). The remaining views are called the side views. These views and all depth data are coded with the use of new tools that are described further in this paper. It may be pointed out that encoding and decoding of the side views and depth maps exploits information from the decoded base view.

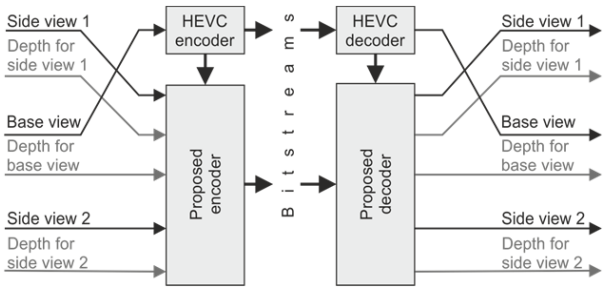


Fig. 2. Proposed codec structure for 3-view MVD (views and their depth maps are denoted in black and gray, respectively).

For the sake of conciseness, further considerations will be given only for the case of 3 views; however. those considerations can be generalized onto an arbitrary number of views. In an independent comparisons of the responses to MPEG CfP [16], the proposed technology has been proved high to have compression efficiency for two coding scenarios: 2 views with 2 depth maps and 3 views with 3 depth maps. For encoding of 3 views, using exhaustive experiments, the we found that the choice of the central view as the base view is usually optimal from the point of view of the overall rate-distortion efficiency. Detailed discussion of this issue is left beyond the scope of the paper.

The next sections of the paper deal with the new tools developed by the authors. These considerations start with Disoccluded Region Coding (Section III) that is the most important coding tool in the proposed 3D video compression technology. Then, two other tools are described: Joint High-Frequency Layer Coding and Depth-Based Motion Prediction (Sections IV and V, respectively). The common feature of these tools is that they all use view synthesis at some stage. Then, two tools for efficient depth representation are considered in Section VI. The first of these tools – Consistent Depth Representation – also exploits view synthesis at a certain step of the depth representation production. The proposed compression technology uses various new tools that exploit view synthesis at some stage. The entire proposed compression technology is described in Section VII. The experimental results for compression performance (Section VIII) are followed by the general conclusions (Section IX).

III. Disoccluded Region Coding

As already mentioned in Introduction, the encoded views are assumed to correspond to the camera locations that are relatively close one to each other. Therefore the views are highly correlated, and the coding tools that bring most of the compression gains are those that exploit inter-view similarities. Of course, those tools are only used in the side views and are not used for the base view that is encoded using the standard HEVC technique. Here, we are going to describe such a new tool called Disoccluded Region Coding, which was proposed by the authors in order to obtain very efficient representations of the side views [29,30].

In Disoccluded Region Coding, view synthesis is used as a primary inter-view prediction mechanism. With reference to the already encoded views and the respective depth maps, a virtual side view is synthesized in the position of the view currently being coded. For the case of 3-view coding, the side views are synthesized from the base view (Fig. 3). For view synthesis the Depth-Image-Based Rendering (DIBR) is used. It is assumed that the intrinsic and extrinsic parameters of cameras are present in the transmitted bitstream for all views. In the experiments (see Section VIII) the we have used a state-of-the-art DIBR implementation developed by MPEG in View Synthesis Reference Software [32].

In the straightforward implementation of the view-synthesis prediction, residuals of the prediction (i.e. prediction errors) are sent to the decoder. Alternatively, the omission of the residuals is signaled in the bitstream. A major drawback of such a solution is considerable yet unjustified transmission payload that either comes from the transmitted residuals or from the signaling of their presence or absence.

Disoccluded Region Coding was proposed in order to omit this drawback. Its idea is based on the observation that a side view consists of two types of distinctive regions (Fig. 3):

- Regions that can be synthesized in the decoder from the decoded base view also usually no prediction error needs to be transmitted.
- Disoccluded regions that cannot be synthesized from the base view and must be transmitted in the bitstream.

The locations and borders of those disoccluded regions can be estimated as a side-product of the DIBR-based synthesis of the side view. After synthesis of a side view, the pixels in an occluded region remain undefined. In a frame of a synthesized view, all pixels with undefined values can be easily detected. In that way, the locations and borders of those disoccluded regions are derived using the already transmitted base view and its depth map. This is doable in the same way both in the encoder and in the decoder. Therefore, there is no need to transmit the locations and borders of the disoccluded regions.

The above-mentioned observation yields that that a side view consists of two distinctive regions:

- Synthesizable regions that are not transmitted but synthesized in the decoder.
- Disoccluded regions are transmitted in the bitstream.

The encoding and decoding of the disoccluded regions is performed in Coding Units (CUs) which are defined in the HEVC draft standard [14] as generalization of macroblocks. The CUs are rectangular blocks of variable size between 64×64 and 8×8 pixels. If a given CU can be entirely synthesized from the base view, it is not transmitted at all. On the contrary, if there are any disoccluded pixels (i.e., pixels that cannot be synthesized from the base view) inside the given CU, it is encoded (Figs. 3 and 4). Because in a typical case most of the scene is the same in all views, only small portions of pictures are disoccluded in subsequently coded views, and thus only a small number of CUs is encoded (Figs. 3 and 4). In the decoder, most of CUs are reconstructed using view-synthesis prediction from the base view. Missing disoccluded regions are filled with the content transmitted in CUs in the bitstream.

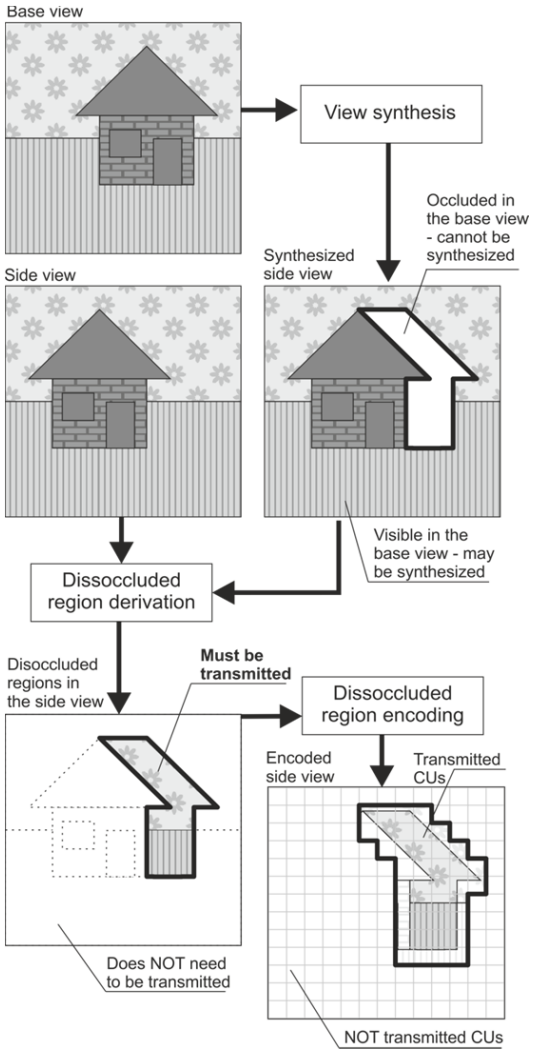


Fig. 3. The principle of Disoccluded Region Coding.

The Coding Units with some disoccluded pixels are encoded using the inter-view disparity-compensated prediction well-known from the MVC extension of the AVC [12]. Note that prediction is only applicable to the disoccluded Coding Units in the side views, i.e. to a relatively small portion of pictures.

Adaptation of the CU size is an important optimization mechanism in HEVC, and it is also exploited by Disoccluded Region Coding. Roughly, the goal is to choose the numbers and sizes of the Coding Units corresponding to disoccluded regions in such a way that all those Coding Units jointly cover the shapes of disoccluded regions (see Figs. 3 and 4). Both the encoder and the decoder use the same view synthesis algorithm with the same reference views. Therefore also the CU adaptation can be performed exactly in the same way both in the encoder and in the decoder. Moreover, the positions and the sizes of the Coding Units that contain pixels from the disoccluded regions can be derived in the decoder in exactly the same way as in the encoder. Therefore, there is no need for additional signaling in the bitstream.

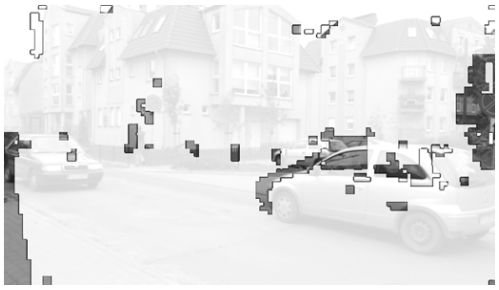


Fig. 4. Disoccluded Coding Units (CUs) transmitted for example *Poznan Street* test sequence.

Application of view-synthesis with Disoccluded Region Coding leads to very efficient representation of video in the side views. That property has been verified experimentally despite of the fact that an adaptive arithmetic encoder is likely to suffer from a small amount of context updating information when the number of the encoded CUs is small. Also, because the number of encoded CUs is highly reduced, and thus representation of the side views and depth maps is so compact, both encoder and decoder can run much faster, resulting in overall reduction of computational complexity.

Note that a reconstructed view, resulting from such encoding, is composed of two distinct areas which contain artifacts of different nature. The first area consists of synthesized regions (Fig. 3) which suffer from view synthesis artifacts. The second area consists of disoccluded regions coded in CUs which suffer from typical compression artifacts. On the boundaries between those two areas, annoying artifacts may occur. In order to remove these artifacts, a technique similar to in-loop deblocking filter is applied inside the loop of the codec.

The above-mentioned in-loop processing is performed as linear interpolation of pixel values within CUs containing disoccluded regions. The linear interpolation provides smooth fading between coded and synthesized regions (Fig. 5).

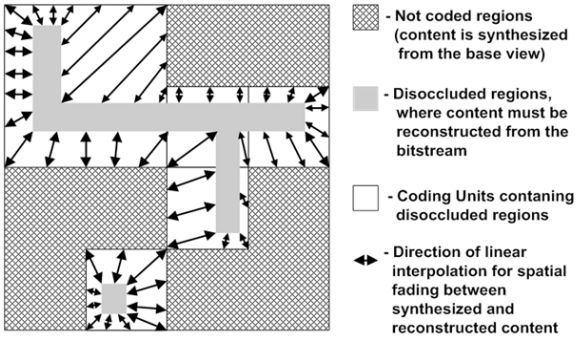
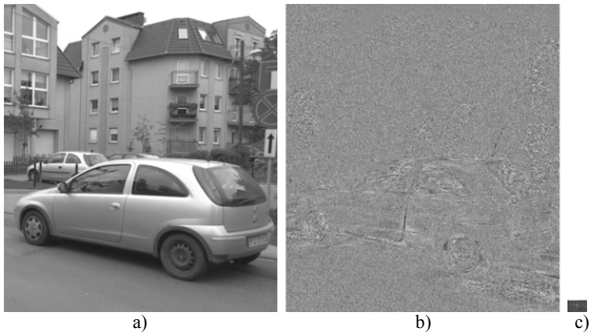


Fig.5. Illustration of in-loop processing for reduction of artifacts caused by Disoccluded Region Coding.

IV. JOINT HIGH-FREQUENCY LAYER REPRESENTATION

In video, temporal high-frequency components are mostly moderate or small (Figs 6a and 6b) but their autocorrelation functions usually exhibit very small values. Therefore these components are not compressed efficiently using standard video coding methods. On the other side, temporal high-frequency components are important for natural appearance of video. Therefore a special tool was proposed to represent efficiently temporal high-frequency components of multiple views.



a) b) c)

Fig. 6. A cropped frame from the test sequence *Poznan Street*: the low-frequency layer (a), the high-frequency layer – for presentation the sample values are amplified 8 times (b) and spatial distribution of energy (c).

The idea is to divide the input image into two spectral bands (layers) in the temporal domain and to encode them separately. The low-frequency layer is obtained by temporal motion-adaptive low-pass FIR filtering [33]. The high-frequency layer is the residual. Frequency division is selected in a way that the high-frequency layer contains only noise.

The low-frequency layer is encoded like video but the high-frequency layer is modeled as a non-stationary random process. There are two components of the model that need to be encoded (Fig. 7): spatial energy distribution (SDE) and spectral envelope. The spatial energy distribution is estimated for each frame. For this purpose, a frame from the high-frequency layer is divided into rectangular non-overlapping blocks. In each of those blocks energy is measured. Energy values, associated with respective blocks, constitute a frame of spatial energy distribution, whose resolution is smaller than resolution of the input video (for example, it fits into one CU for a HD frame, see Fig. 6c).

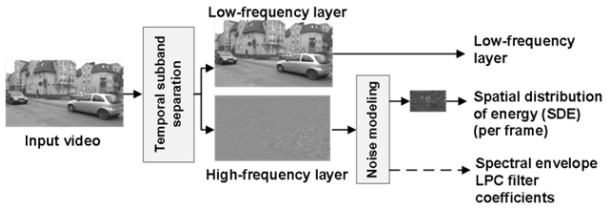


Fig. 7. High-Frequency Layer Representation in an encoder.

This estimated spatial distribution of energy is used in order to normalize the high-frequency layer.

For high-frequency layer, the second component is spectral envelope. It is estimated from energy-normalized high-frequency subband using a technique similar to LPC. The resulting set of separable IIR filter coefficients (in horizontal and vertical direction) is encoded using LAR (log-area-ratio [34]) 8-bit representation. A set of those filter coefficients is estimated for each frame and transmitted to the decoder.

Parameters of the noise model are highly correlated among the views. The frames of the spatial distribution of energy of all views are mapped through view synthesis to a position of the base view, and then averaged. This operation results in only one joint spatial distribution of energy (SDE). Similarly, the energy envelopes of all of the views are averaged, resulting in one joint spectral envelope.

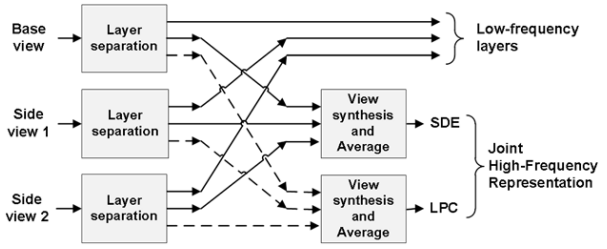


Fig. 8. Layered representation of views: low-frequency layer and Joint High-Frequency Layer Representation (SDE and LPC) – video and parameters are denoted by solid and dashed lines, respectively.

In a decoder, pseudo-random white noise is generated and then modulated by the upsampled spatial energy distribution transmitted in the bitstream and then filtered with IIR filters that reflect the envelope of the original high-frequency layer spectrum. The resultant video, which resembles the original high-frequency subband, is added to the reconstructed low-frequency layer in order to restore the high-frequency components.

V. Depth-Based Motion Prediction (DBMP)

All of the views in a multview video sequence are projections of the same scene. Therefore, motion of the objects in the subsequent viewsis almost the same. The basic idea of this tool is to reuse motion field of the base view in the side views.

The concept of DBMP was previously described in [19] for the MVC codec. Further the tool was extended [35] as a coding tool for the HEVC-based multiview video codec [36,37]. In the proposed 3D video coding technology, some modifications of this DBMP tool are introduced to maximize its performance.

During encoding of the base view sparse block-based motion field is estimated and transmitted as a part of motion compensation prediction in the HEVC encoder. This motion field of a base view (along with the corresponding depth map) is used for synthesis of dense motion field for side views. Such synthetized motion field is then used during the encoding of the side view.

Before encoding and decoding of a frame, for each pixel in the side view, motion information (motion vector and reference picture indices) is synthesized directly from already encoded CUs in the base view at the same time instant (Fig. 8). As a consequence, the motion vectors and reference indices do not need to be transmitted as they can be derived through the DIBR technique at the decoder from the base view. For the CU that uses synthesized motion field, motion compensation is done on pixel basis.

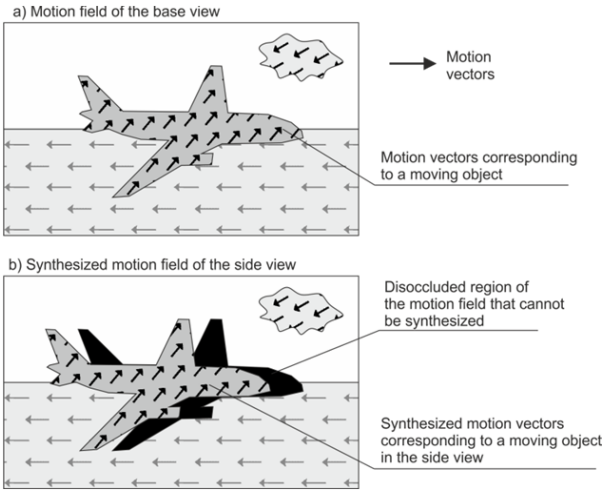


Fig. 9. Independent derivation of motion information for each point of encoded CU from a corresponding point in the reference view.

In HEVC, the most efficient modes for encoding motion information are based on the concept of block merging [35]. In this approach motion vectors and reference picture indices for each currently encoded CU are inferred from motion information of a reference CU. As a result, no motion information or reference picture indices need to be transmitted in the bitstream for the encoded CU. For this reason, DBMP tool has also been implemented as an additional merge candidate. Obviously, the proposed DBMP merge candidate is not added in the base view, which is HEVC-compatible.

VI. Depth Representation

One of the challenges in 3D video coding is efficient representation of the depth data. Many depth views may exist the MVD format, however, the depth maps acquired at the same time instant, but from different views, usually exhibit a significant level of spatial inconsistency. This inconsistency results in different depth values for the corresponding points in the individual views. Also, the depth value at a point may be

considered as noisy, especially after classical coding which introduces uniform quantization error. This is especially disappointing for objects that reside in close distance to the camera. We propose two tools that cope with these problems: Uniform Depth Representation and Nonlinear Depth Representation.

A. Consistent Depth Representation

As mentioned before, the inter-view prediction based on view synthesis is widely used for both video and depth. Unfortunately, if the view synthesis algorithm uses inconsistent depth maps, it renders very annoying artifacts in the synthesized video. For that reason, the first step of the 3D video compression algorithm is the depth map inter-view consistency refinement that produces Consistent Depth Representation. The refinement technique consists of four steps: depth map smoothness improvement, depth map synthesis, inter-view depth information exchange and depth value restoration.

The first step of this processing is to improve depth map smoothness by using the Mid-Level Hypothesis algorithm [38]. This algorithm increases sub-pixel precision of the artificially estimated depth maps and provides some level of alignment between the depth and the corresponding texture. This algorithm is described in more detail in [31].

The depth map synthesis is performed similarly to the view synthesis. Each depth map is synthesized from each other depth map independently. For 3-view MVD, this results in two alternative depth maps for each view.

The inter-view depth information exchange consists in median filtering of alternative depth values for each view. This results in depth values that are aligned consistently in the views.

The depth value restoration step ensures that the depth modifications imposed by the algorithm do not affect the overall 3D model of the scene represented by the depth maps.

In general, these four steps could be repeated iteratively but the experimental results [31] indicate that the major improvement is made in the first iteration. In real-life video coding case, it is enough to do one or two iterations.

B. Nonlinear Depth Representation

Human perception of depth depends on absolute distance between viewed objects and the observer. A typical observer is more sensitive to depth variations close at hand than to those far away. Also, view synthesis algorithms are much more vulnerable to depth mismatch in the foreground (which results in doubling of borders of the objects) than in the background (typically resulting merely in a small displacement of the whole plane). Therefore, we propose to employ a tool called Nonlinear Depth Representation. This tool produces nonlinear representation of the depth values for compression, which consequently is equivalent to non-uniform quantization of depth.

Depth samples are transformed according to a nonlinear function. An inverse transformation is done after decoding (Fig. 9). This impacts other coding and decoding operations in the tools that rely on depth values. For example, the encoder and the decoder have to be aware about the depth transformation when performing view synthesis.

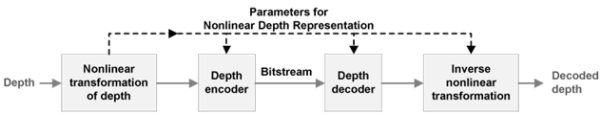


Fig. 10. Nonlinear Depth Representation scheme.

The internal depth representation is non-linear, i.e. closer objects are represented more accurately than distant ones. Internal depth sample values are defined by the following power-law expression, similar as in the case of the well-known video gamma correction:

$$d_{nonlinear} = \left(\frac{d_{linear}}{M_{linear}}\right)^{\gamma} \cdot M_{nonlinear} \quad (1)$$

where: $d_{nonlinear}$ is a nonlinear disparity value used internally inside the codec, normalized to the range from 0 to $M_{nonlinear}$, d_{linear} is a linear disparity value at the input and at the output of the codec, normalized to the range from 0 to M_{linear} and γ is a constant parameter that controls non-linearity of the process.

The exponent γ is automatically chosen by the encoder and sent to decoder in the encoded bitstream. In our experiments it was set in range 1.2÷1.6, depending on the quantization step. The depth range was assumed to be constant, but recently it was proposed to update it on frame basis [39].

VII. Integration of the Tools into the Codec Structure

For both video and depth, hierarchical view coding structure similar to MVC [13] is used: the already coded views are used as references for the coding of the subsequent views. The proposed 3D video codec includes the new tools described in this paper. These tools exploit the presence of the depth information using the approaches known from computer vision. These tools are integrated with the MVC structure and basic low-level HEVC compression tools like intra-frame prediction, inter-frame motion-compensated prediction, transform coding, in-loop filtering and others. The encoder and decoder structures are depicted in Fig. 11.

In general, the composite 3D-video bitstream consists of 4 types of video sub-streams (see Fig. 11):

- low-frequency video layer of the base view,
- low-frequency video layers of the side views (more than one such sub-stream may exist),
- depth maps for individual views (more than one such sub-stream may exist),
- high-frequency layer representations – may be sent for individual views but, as already mentioned, even all views may share one such component.

For each time instant, video and depth frames from the base view are encoded first, then the depth maps are encoded for the side views. Afterwards, the video for side views and the residual layer are encoded.

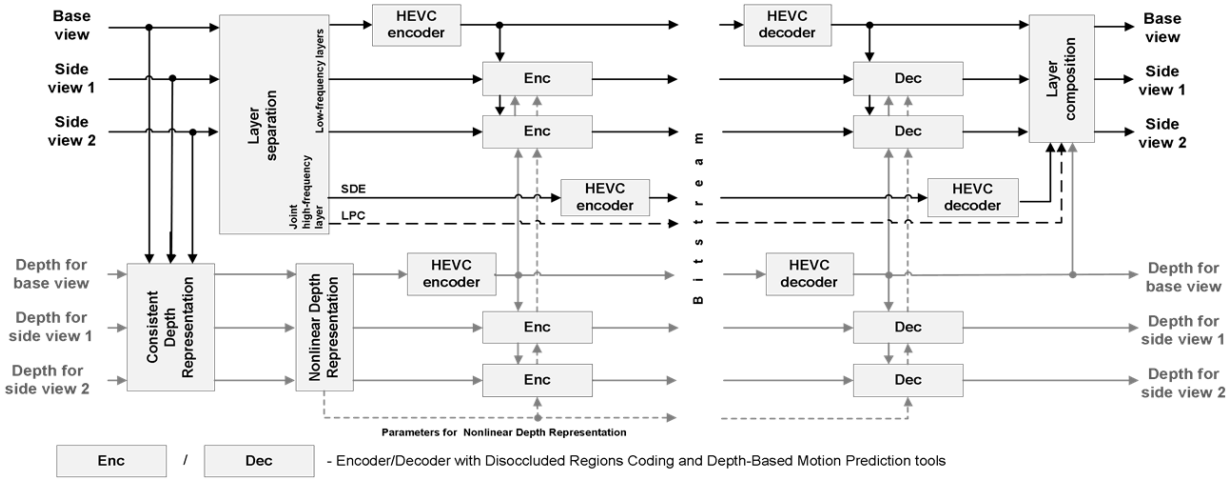


Fig. 11. Proposed encoder (left) and decoder (right) structures.

The encoder produces a bitstream in the form of a sequence of NAL units. As mentioned before, the bitstream of the base view is fully compliant with the HEVC syntax. Other streams, which are not HEVC-compatible, are encapsulated in transparent NAL units, so that they can be skipped by a basic HEVC decoder. A full 3D decoder can use them to decode all of the input views or only some of them.

VIII. Experimental Assessment and Subjective Evaluation

The final versions of individual tools and the codec structure have been obtained by extensive iterative experiments that are beyond the scope of this paper. Here, we report the experimental results that allow to estimate the overall compression performance of the technology as well as to estimate the efficiency of the individual tools. This is done using subjective and objective tests of decoded video quality. The methodology resembles the one used by MPEG for evaluation of the responses to Call for Proposals on 3D Video Coding. For that purpose MPEG has selected 8 test MVD sequences. Four of them were in 1920×1080 resolution (*Poznan Hall 2, Poznan Street, GT Fly* and *Undo Dancer*). Other four test MVD sequences were in the 1024×768 format. For the sake of brevity, in this paper, detailed results will be presented for the 1920×1080 sequences only (Table 1).

Table 1. MVD test sequences used for experiments.

Name	Length	Type	Provider
S01 *Poznan Hall2*	8 secs	Natural - indoor	Poznań University of Technology [40]
S02 *Poznan Street*	10 secs	Natural – outdoor	
S03 *Undo Dancer*	10 secs	Synthetic – indoor	Nokia Corporation [41,42]
S04 *GT Fly*	10 secs	Synthetic – outdoor	



Fig. 12. The arrangement of the views: the v are marked in black while the views synthesized in the receiver are marked in gray.

For our experiments we used 3 views from each sequence (with texture depth maps) and encoded them at four different bitrates. Based on the decoded data we synthesized six virtual views (Fig. 12 – "v1", ... ,"v6") uniformly placed between the original views (Fig 12 – "1", ... ,"3"). Similarly, we have synthesized six virtual views at the same spatial positions from uncompressed data as a reference for average PSNR calculations for luma (Tables 2 and 3, Fig. 11). Average bitrate reductions versus HEVC simulcast were calculated using the Bjontegaard formula [43]. Synthetic references were used instead of the video captured by real cameras because our aim was to assess the quality degradation caused by the coding technology, not caused by the view synthesis algorithm itself.

For the view synthesis algorithm, we used MPEG Synthesis Reference Software [32] with the default configuration.

In all cases, original (not pre-preprocessed) sequences have been used as references for quality measurement – both objective (PSNR) and subjective (MOS).

The subjective tests have been carried out [44] in accordance with the general rules of ITU Recommendation BT.500 [45]. A total number of 62 young persons were viewing each stereopair (composed from virtual views "v3" and "v4", see Fig. 12) on a 46'' Hyundai S465D polarization monitor. The Double Stimulus Method was selected for the subjective quality assessment that followed the rules used by the MPEG for evaluation of the proposals for the 3D video coding technology in 2011 [16].

In our experiments, the number of subjects involved was higher than in the official MPEG evaluation. The high number of subjects yielded that 95% confidence intervals were of order of ±(0.1÷0.25), i.e. very small. Therefore, those intervals were not depicted on the plots (Fig. 13).

Figure 13 shows objective evaluation results (PSNR versus bitrate - BD-rates - in Table 2) and Figures 17-20 show subjective evaluation results (11-point MOS versus bitrate - BD-rates – Table 3) for virtual synthesized views for all four testes sequences.

Note that both subjective and objective quality assessments lead to somewhat similar conclusions. Application of Nonlinear Depth Representation (Tables 2 and 3 – column A) may result in more than 20% bitrate reduction.

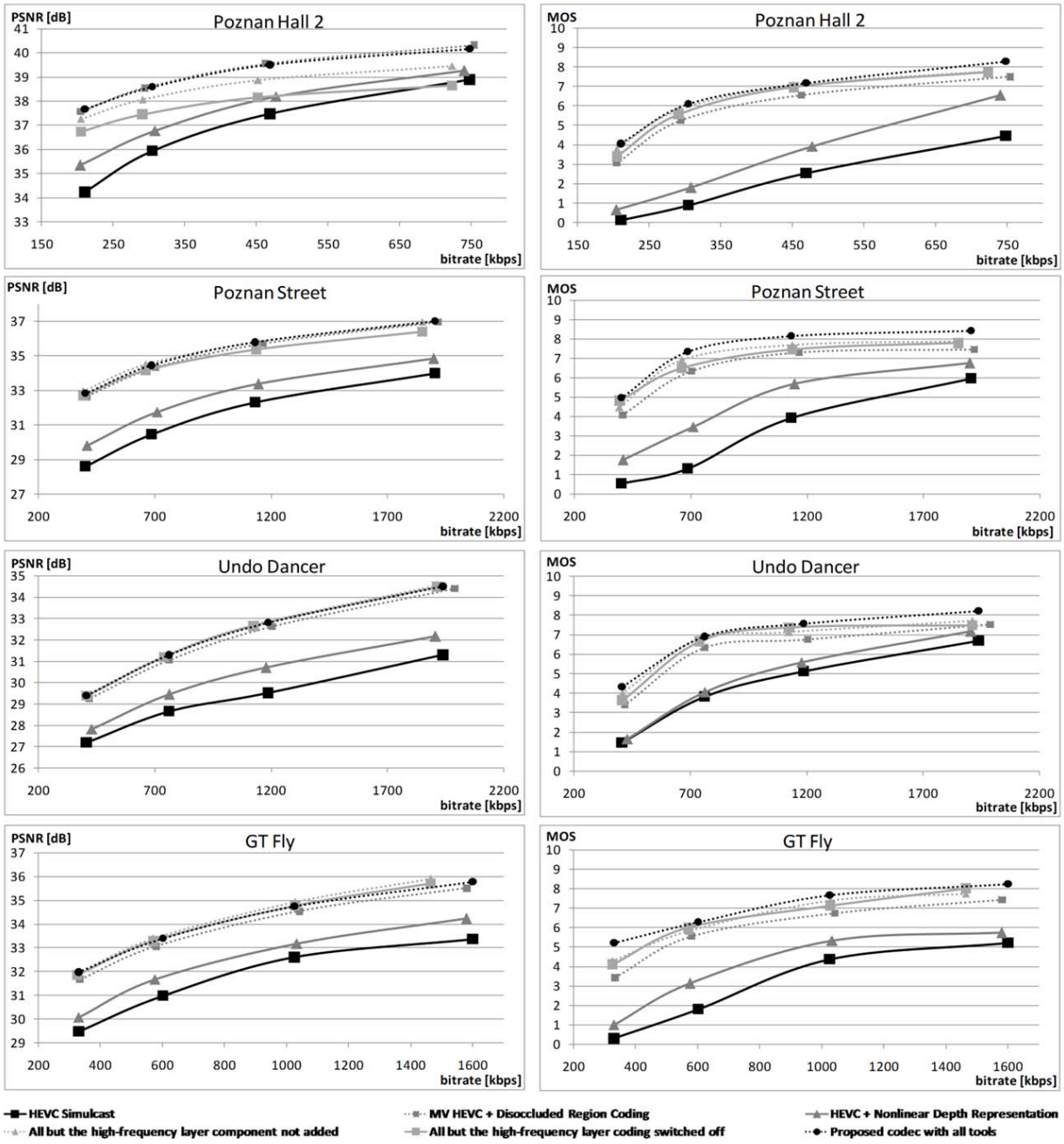


Fig. 13. Objective (left) and subjective (right) evaluation results for S01-S04 sequences.

View-synthesis inter-view prediction combined with MVC-toolset (column C) yields about 50 – 60 % bitrate reduction. Disoccluded Region Coding implemented in a standard HEVC without the MVC toolset provides similar bitrate reductions (column B) of about 45%. The discrepancy between the results obtained by subjective and objective video quality assessment is the most noticeable in the case of *Poznan Hall 2* sequence (S01) which probably results from low quality of associated depth maps.

The application of all devised tools except the Joint High-Frequency Layer Representation results in 50% or more of bitrate reduction (column D). When High-Frequency Layer coding is used, we can consider two cases: column E - when the high frequency layer is modeled but finally not reconstructed, which is more reliable for objective evaluation (PSNR comparison with synthetic noise would be irrelevant) - the gains here are about 50% (objective quality measures) and over 60% (subjective quality assessment).

In another case - column F in Tables 2 and 3 - the synthetic noise is additionally summed with the output video (more important for subjective viewing), which improves the overall gains by about 1 percent. Such gain is not much, but also adequate to the bitrate cost of High-Frequency Layer (Fig. 13).

Therefore, the main gains of the whole proposal come from: Disoccluded Region Coding, Nonlinear Depth Representation and disparity-compensated prediction (MVC-like). It is also worth noting that the remaining tools together contribute a substantial gain of approximately 2÷4% of the overall bitrate as compared to simulcast HEVC (Tables 2 and 3).

The decoded stereoscopic sequences for 2-view and 3-view scenarios and bitstreams (for 8 test sequences), and decoder executables can be found at the website: http://3d-codec.multimedia.edu.pl.

Table 2. Average bitrate reductions calculated as Bjontegaard rates for luma PSNR [dB] versus original (not preprocessed) sequences.

Test sequence	Average (over bitrates and sequences) bitrate reduction versus HEVC simulcast					
	A. HEVC + Nonlinear Depth Representation	**B.** HEVC + Disoccluded Region Coding	**C.** MV HEVC + Disoccluded Region Coding	**D.** All but Joint High-Freq.Layer Repr. switched off	**E.** All but high-freq. layer not added	**F.** Proposed codec with all tools
S01	-19.6	-20.3	-26.1	-14.7	-16.9	-23.7
S02	-27.2	-55.7	-56.8	-58.0	-62.8	-59.8
S03	-29.1	-57.0	-58.0	-60.9	-61.1	-60.7
S04	-23.2	-48.8	-49.4	-54.0	-55.4	-53.7
Avg.	**-24.8**	**-45.4**	**-47.6**	**-49.1**	**-49.1**	**-49.5**

Table 3. Average bitrate reductions calculated as Bjontegaard rates for Mean Opinion Score (MOS) versus original (not preprocessed) sequences.

Test sequence	Average (over bitrates and sequences) bitrate reduction versus HEVC simulcast				
	A. HEVC + Nonlinear Depth Representation	**C**. MV HEVC + Disoccluded Region Coding	**D.** All but Joint High-Freq.Layer Repr. switched off	**E.** All but high-freq. layer not added	**F.** Proposed codec with all tools
S01	-24.5	-65.2	-67.2	-69.4	-70.1
S02	-35.7	-67.5	-72.2	-72.6	-74.8
S03	-8.0	-52.3	-57.4	-61.4	-62.7
S04	-29.6	-62.0	-69.0	-68.8	-67.2
Avg.	**-24.5**	**-61.7**	**-66.4**	**-68.1**	**-68.7**

IX. CONCLUSIONS

In the paper, we have proposed a 3D video coding technology which consists of several new coding tools. The benefits and costs of individual tools have been discussed with the reference to the respective experimental results. The proposed compression technology provides bitrate reduction of the order of 60% as compared to HEVC simulcast. This figure was obtained by systematic subjective tests. It proves high compression performance of the proposed technology that allows very efficient coding of the side views. The bitrate needed for even two side views with the corresponding depth maps is mostly below 50% of the bitrate for single-view video.

REFERENCES

[1] "Applications and Requirements on 3D Video Coding", ISO/IEC JTC1/SC29/WG11, Doc. N11829, Geneva, CH, March 2011.

[2] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. of IEEE Int. Conf. Image Proc.(*ICIP'07), Austin TX, USA, September 2007, pp.I-201-204.

[3] P. Kauff, K. Müller, A. Smolic, et al. "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," Signal Proc.: Image Comm,, vol. 22, no. 2, 2007.

[4] F. Shao, G. Jiang, M. Yu, K. Chen, Y.-S. Ho, "Asymmetric Coding of Multi-View Video Plus Depth Based 3-D Video for View Rendering," IEEE Trans. on Multimedia, vol. 14, 2012, pp. 157- 167.

[5] K. Müller, P. Merkle, T. Wiegand, "3-D video representation using depth maps," *in Proc. of the IEEE,* vol. 99, no.4, 2011, pp. 643-656.

[6] "Overview of 3D video coding," ISO/IEC JTC1/SC29/WG11, Doc.N9784, Archamps, FR, May 2008.

[7] P. Merkle, Y. Morvan, A. Smolic, et al. "The effects of multiview depth video compression on multiview rendering," Signal Processing: Image Communication, vol. 24, no. 1+2, January 2009, pp. 73-88.

[8] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. of the IEEE,* vol. 99, no.4, 2011, pp. 643-656.

[9] Y. Lin, and J. Wu, "A depth information based fast mode decision algorithm for color plus depth-map 3D videos, " *IEEE Trans.* on Broadcasting, vol.57, no.2, June 2011, pp.542-550.

[10] D. Tian, P.-L. Lai, P. Lopez, C. Gomila, "View Synthesis Techniques for 3D Video," App. of Digital Image Proc. XXXII. Proc, of the SPIE, Volume 7443 (2009), 2009.

[11] A. Smolic, K. Müller, T. Wiegand, et al. "Intermediate View Interpolation Based on Multiview Video Plus Depth for Advanced 3D Video Systems," IEEE Int. Conf. on Image Processing, ICIP2008,SanDiego, CA, USA, October 2008.

[12] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC Standard," *Proc. of the IEEE*, vol. 99, no.4,2011, pp. 626-642.

[13] Generic coding of audio-visual objects – Part 10: Advanced Video Coding, ISO/IEC 14496-10, 6nd Ed., 2010, [also:] ITU-T Rec. H.264, Edition 5.0, 2010

[14] "High-Efficiency Video Coding (HEVC) text specification draft 10", JCTVC Document, JCTVC-L1003, Geneva, CH, Jan. 2013.

[15] K. Wegner, O. Stankiewicz, K. Klimaszewski, M. Domański "Comparison of multiview compression performance using MPEG-4 MVC and prospective HVC technology", ISO/IECJTC1/SC29/WG11 MPEG M17913, Geneva, CH, July 2010.

[16] "Call for proposals on 3D video coding technology", ISO/IEC JTC1/SC29/WG11, Doc. N12036, Geneva, CH, March 2011.

[17] J. Seo, H. Wey, S. Lee, K. Sohn, "Motion information sharing mode for depth video coding," *in Proc. of* 3DTV Conference: The True Vision-Capture,Transmission and Display of 3DVideo(3DTV'2010), 2010.

[18] K. Klimaszewski, K. Wegner"Joint Intra Coding of Video and Depth Maps" IEEE Int. Conf. Signals and Electronic Systems – ICSES 2010, Gliwice, PL, Sept. 2010.

[19] J. Konieczny, M. Domański, "Extended inter-view direct mode for Multiview Video Coding", IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 845-848, Prague, CZ, May 2011.

[20] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," *in Proc. of* IEEE Intern. Conf. on Image Processing(ICIP'07), TX, USA, 2007, pp.I-201-204.

[21] T. Suzuki, M. M. Hannuksela, Y. Chen, S. Hattori , "Text of ISO/IEC 14496-10:2012/PDAM 2 MVC extensions for inclusion of depth maps", W12731, MPEG meeting, Geneva, May 2012.

[22] M. M. Hannuksela, Y. Chen, T. Suzuki, J.-R. Ohm, G. J. Sullivan, "3D-AVC draft text 5", ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc.JCT3V-C1002, Geneva, CH, Jan. 2013.

[23] M. Domański, K. Klimaszewski, O. Stankiewicz, K. Wegner et al. "Multiview HEVC – experimental results" JCT-VC (MPEG/VCEG) Doc. JCTVCG582,Geneva, November 2011.

[24] G. Tech, K. Wegner, Y. Chen, M. Hannuksela, J. Boyce, " MV-HEVC Draft Text 3", JCT3V Document, JCT3V-C1004, Geneva, CH, Jan. 2013.

[25] D. Rusanovskyy, F. C. Chen, L. Zhang, T. Suzuki, "3D-AVC Test Model 5", ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11,Doc. JCT3V-C1003, Geneva, CH, Jan. 2013.

[26] H. Schwarz, K. Wegner, " Test Model under Consideration for HEVC based 3D video coding", ISO/IEC JTC1/SC29/WG11/ MPEG2011/N12559, San Jose, CA, USA, Feb. 2012.
[27] M. Domański, O. Stankiewicz, K. Wegner, et al. "Technical Desciption of Poznan University of Technology proposal for Call on 3D Video Coding Technology", ISO/IEC JTC1/SC29/WG11, Doc. M22697, Geneva, CH, November 2011.
[28] M. Domański, O. Stankiewicz, K. Wegner et al. "New Coding Technology for 3D Video within Depth Maps as Proposed for Standardization within MPEG," Int. Conf. Systems, Signals and Image Processing IWSSIP 2012, Vienna, AT, April, 2012, pp. 415-418.
[29] M. Domański, O. Stankiewicz, K. Wegner et al. "Coding of Multiple Video+Depth Using HEVC Technology and Reduced Representations of Side Views and Depth Maps," 29th Picture Coding Symposium, Kraków, PL, May 2012, pp. 5-8.
[30] M. Domański, O. Stankiewicz, K. Wegner et al. "3D video compression by coding of disoccluded regions," 2012 IEEE Int. Conf. Image Proc., ICIP 2012, Orlando, Florida, USA, Sept. – Oct. 2012.
[31] M. Kurc, O. Stankiewicz, M. Domański, "Depth map inter-view consistency refinement for multiview video", Picture Coding Symposium, Kraków, PL, 2012.
[32] ISO/IEC JTC1/SC29/WG11, "View synthesis algorithm in view synthesis reference software 3.0 (VSRS3.0)," Doc. M16090 , Feb.2009.
[33] M. Fizick (A. Balakhnin), Tsp, TSchniede "Mv-tools web-page", http://avisynth.org.ru/mvtools/mvtools.html.
[34] J.A. Martins "Low bit rate LPC vocoders using vector quantization and interpolation" International Conference on Acoustics, Speech, and Signal Processing, 1991, vol. 1 pp. 597- 600.
[35] J. Konieczny, M. Domański, "Depth-based inter-view motion data prediction for HEVC-based multiview video coding", Picture Coding Symposium, PCS 2012, Kraków, PL, May 2012, pp.33-36.
[36] M. Domański et al., "Multiview HEVC – experimental results", ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc.JCTVC-G582, Geneva, CH, 2011.
[37] J. Stankowski, M. Domański, O. Stankiewicz, K. Wegner et al."Extensions of the HEVC technology for efficient Multiview Video Coding", IEEE Intern. Conf. on Image Processing, Orlando, USA, 2012.
[38] O. Stankiewicz, M. Domański, K. Wegner, "Stereoscopic Depth Refinement by Mid-Level Hypothisis", IEEE Int. Conf. on Multimedia & Expo, Singapore, SG, July 2010.
[39] I. Lim, H. Wey, D. Park, "3D-CE7.a Improved Nonlinear Depth Representation", ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-C0094, Geneva, CH, Jan. 2013.
[40] M. Domański, O. Stankiewicz, K. Wegner et al. "Poznań multiview video test sequences and camera parameters," ISO/IEC JTC1/SC29/WG11, Doc. M17050, Xian, China, October 2009.
[41] J. Zhang, Ri Li, H. Li, D. Rusanovskyy and M.M. Hannuksela, "Ghost Town Fly 3DV sequence for purposes of 3DV standardization," ISO/IEC JTC1/SC29/WG11, Doc. M20027, Geneva, CH, March 2011.
[42] D. Rusanovskyy, P. Aflaki and M. M. Hannuksela, "Undo Dancer 3DV sequence for purposes of 3DV standardization," ISO/IEC JTC1/SC29/WG11, Doc. M20028, Geneva, CH, March 2011.
[43] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," ITU-T SG16, Doc. VCEG-M33, April 2001.
[44] F. Lewandowski, M. Paluszkiewicz, T, Grajek, K. Wegner, "Subjective Quality Assessment Methodology for 3D Video Compression Technology", Int. Conf, of Signals and Electronic Systems, 2012.
[45] Methodology for the subjective assessment of the quality of television pictures, ITU-R Rec. BT.500-11, 2002.



Marek Domański (M'91) received M.S., Ph.D. and Habilitation degrees from Poznań University of Technology, Poland in 1978, 1983 and 1990 respectively. Since 1993, he is a Professor at Poznań University of Technology, where he leads the Chair (Department) of Multimedia Telecommunications and Microelectronics. Since 2005 he is the head of Polish delegation to MPEG. He authored highly ranked technology proposals submitted in response to MPEG calls for scalable video compression (2004) and 3D video coding (2011). He also leaded the team that developed one of the very first AVC decoders for tv set-top boxes (2004) and various AVC and AAC HE codec implementations and improvements. He is an author or co-author of 3 books and over 200 papers in journals and proceedings of international conferences. The contributions were mostly on image, video and audio compression, image processing, multimedia systems, 3D video and color image technology, digital filters and multidimensional signal processing. He was General Chairman and host of several international conferences: Picture Coding Symposium, PCS 2012; European Signal Processing Conference, EUSIPCO 2007; 73rd Meeting of MPEG; Int. Workshop on Signals, Systems and Image Processing, IWSSIP 1997 and 2004; Int. Conf. Signals and Electronic Systems, ICSES 2004 and others. He served as a member of various steering, program and editorial committees of international journals and international conferences. He served as Area Editor of Signal Processing: Image Communications journal in 2005-2010.



Olgierd Stankiewicz received his M. Sc. degree from the Faculty of Electric Engineering, Poznań University of Technology in 2006. In 2005 he won the second place in IEEE *Computer Society International Design Competition (CSIDC*), held in Washington D.C. Currently, he is an assistant at the Chair of Multimedia Telecommunications and Microelectronics, where he is working toward the Ph.D. degree. He is co-author of several papers on free view television, depth estimation and view synthesis and papers related to hardware implementation in FPGA. His professional interests include signal processing, video compression algorithms, computer graphics and hardware solutions. He is also involved in ISO standardization activities where he contributes to the development of the 3D video coding standards.



Krzysztof Wegner received the M.Sc. degree from Poznań University of Technology in 2008. Currently he is working towards his Ph.D. there. He is co-author of several papers on free view television, depth estimation and view synthesis. His professional interests include video compression in multipoint view systems, depth estimation form stereoscopic images, view synthesis for free view television, face detection and recognition. He is involved in ISO standardization activities where he contributes to the development of the future 3D video coding standards.

Jacek Konieczny received the M.Sc. and Ph.D. degrees from Poznań University of Technology, Poznań, Poland, in 2008 and 2013, respectively. He has been involved in several projects focused on multiview and 3D video coding. His research interests include representation and coding of multiview video scenes, free-viewpoint video, and 2-D and 3-D video-based rendering. He is involved in ISO standardization activities where he contributes to the development of the 3D video coding standard.

Maciej Kurc received his M.Sc. (2008) from the Faculty of Electronics and Telecommunications, Poznań University of Technology, PL, where he currently is a Ph.D. student. His main areas of research are video compression and FPGA logic design.

Jakub Siast received the M.Sc. degree (2009) from the Faculty of Electronics and Telecommunications, Poznań University of Technology, PL, where he is Ph.D. student. His current research interests include image processing and coding, developing of video coding algorithms, FPGA and microprocessor architecture design.

Jakub Stankowski received the M.Sc. degree (2009) from the Faculty of Electronics and Telecommunications, Poznań University of Technology, PL, where he is a Ph.D. student. His current research interests include video compression, performance optimized video processing algorithms, software optimization techniques.

Robert Ratajczak received the M.Sc. degree from the Faculty of Electronics and Telecommunications, Poznań University of Technology, PL, in 2010, where he is currently a Ph.D. student. His current research interests include stereoscopic images processing and coding, 3D surface reconstruction, object classification and detection.

Tomasz Grajek received his M.Sc. and Ph.D. degrees from Poznań University of Technology in 2004 and 2010 respectively. At present he is an assistant at the Chair of Multimedia Telecommunications and Microelectronics. He is author and co-author of several papers on digital video compression, entropy coding and modeling of advanced video encoders. He has been taking part in several projects for industrial research and development projects. His current researches encompass implementing and optimizing video codecs and modeling of advanced video coders.

Fast and Energy-Efficient Watermarking of HEVC-Compressed Video Bitstreams

Jakub Stankowski
Institute of Multimedia Telecommunications
Poznań University of Technology
Poznań, Poland
ORCID: 0000-0002-5105-5090

Marek Domański
Institute of Multimedia Telecommunications
Poznań University of Technology
Poznań, Poland
ORCID: 0000-0002-9381-0293

Tomasz Grajek
Institute of Multimedia Telecommunications
Poznań University of Technology
Poznań, Poland
ORCID: 0000-0002-7142-9416

***Abstract*— Most research results in video watermarking refer to the embedding of messages into uncompressed video, which needs full decoding and then full encoding of the content. In some applications, such a method is too complex, and watermarking of compressed bitstreams is required. Watermarking of compressed bitstreams is more difficult as the number of bits is very limited, and the compressed video is very sensitive to all modifications that may yield visible artifacts in the decoded watermarked video. In the paper, an original watermarking method is described for HEVC-compressed video. The main idea of the technique is to embed a message into nonzero residual transform coefficients in such a way that potential drift is limited. The watermark detection is based on reference, unwatermarked video. The experimental verification demonstrates the strong robustness of the embedded watermarks.**

Keywords*— *robust watermarking, HEVC, transform-domain watermarking, bitstream watermarking

I. Introduction

We are witnessing continuous development in video compression methods. About every 9-10 years new compression standard is released by ISO/IEC [1]. Typically video compression standards do not have support for functionalities such as authentification, copyright protection, and fingerprinting. Therefore, whenever a new standard is deployed on the market, a great effort is made to develop the above-mentioned functionalities. They are very crucial because with improvements in flat-panel displays and Internet streaming platforms, creating high-quality copies of video data is getting easier and easier every year.

One of the most popular solutions for illegal copy protection is watermarking [2, 3]. Generally, watermarks may be visible or invisible, fragile or robust. For illegal copy protection, a watermark should be robust (it has to survive after strong image processing including camcording and recompression) and invisible/unnoticeable by potential viewers [4-7].

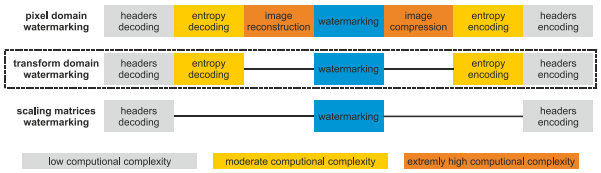


Fig. 1. Comparison of most significant classes of compressed domain watermarking. The method proposed in this paper belongs to the transform domain watermarking class [5].

In order to survive after recompression and/or camcording, a watermark should be hidden in image samples. Generally, three main paths to modify image samples according to watermark may be distinguished (Fig. 1): pixel domain, transform domain, and scaling matrices domain. The scaling matrices domain is the less complex approach because only entropy decoding and re-encoding are needed. However, it allows only for very weak control of the watermark embedding process as there is no direct access to image samples. At the other end, we have the pixel domain watermarking. This is the most complex approach (full decoding and re-encoding are needed), but it offers full control of the watermark embedding process including drift compensation. Between these two mentioned approaches, the transform domain watermarking is placed. It is characterized by moderate complexity (partial decoding and re-encoding are needed) and gives indirect access to image samples through transform coefficients [2, 3].

This paper focuses on HEVC (MPEG-H part 2 and H.265) [8], as it is observed to replace AVC (MPEH-4 Part 10 and H.264) [9]. The HEVC watermarking is not new but is still a very active research area [6, 7, 10-35]. Mainly because most of the methods designed for previous video compression standards cannot be directly applied to HEVC, except that requiring full decoding and reencoding (i.e. pixel domain watermarking). Moreover, most of the proposed methods are fragile against full decoding and further processing of video samples, so they will be useless to track illegal camcording and distribution on the internet.

II. General Idea

The idea of the described solution is to propose a technology that allows hiding some watermark data in HEVC-compressed bitstream. The user of the technology should have the possibility to mark each stream with a dedicated watermark. Therefore, the watermark embedding should have relatively low computational complexity, and memory bandwidth requirements and generate low energy consumption. This leads us to only one reasonable approach which is the transform domain watermarking (Fig. 1). Moreover, the authors previously developed a set of watermarking solutions for AVC [9] based on scaling matrices watermarking [4] and transform domain watermarking [5].

In the proposed approach the watermarking technique is treated as a way of creating a reliable communication channel hidden in video content that is robust against common attacks. The general idea of the proposed system has been shown in Fig. 2.

In general, the watermark is embedded by altering some of the low-frequency DCT coefficients transmitted as residual data. This leads to modification of reconstructed frame frequency characteristics i.e. a slight increase of energy of selected components. These changes are unnoticeable for the viewer but can be detected and decoded into watermark bits.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

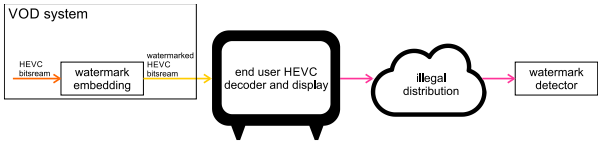


Fig. 2. The general idea and use case of the proposed watermarking system.

III. Description of the Proposed Method

A. Frame selection

The transform domain watermarking allows embedding watermark into any frame. However, there are several factors rendering the watermarking of all frames inefficient.

The first reason is computational complexity. As measured and described in [36], the I frames are the biggest in the bitstream i.e. in a 32-picture GOP (Group of Pictures) containing only one I frame, the I frame takes about 30% of a bitstream. The remaining 31 frames are 70% which makes them more suitable for processing. Since the B and P frames are smaller (and entropy encoding and decoding are significant factors in watermarking computational complexity) the processing of P and B frames only allows for reducing latency introduced by watermarking process.

The second factor is drift control. Any change in the I frame propagates to almost all frames in GOP since the I frame is the main source of motion compensated prediction. The same happens to P and B frames at the beginning of the GOP. The propagation of watermark thru motion-compensated prediction increases distortion and can cause visible artifacts. This leads to the idea of processing only the second half of any GOP in order to reduce drift. Moreover, the availability of non-watermarked pictures (1st half of the GOP) allows to model the influence of watermark attacks and helps in increasing the reliability of watermark detection.

The examples of picture selection in the proposed method are presented in Fig. 3.

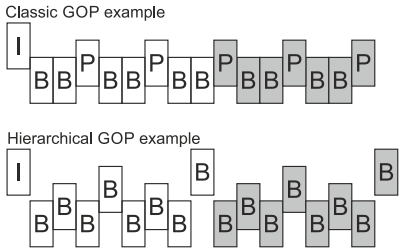


Fig. 3. Visualization of picture selection for classic and hierarchical GOP. Pictures marked as grey are processed during watermark embedding.

B. Distortion-optimized selection of TUs

In HEVC [8], the P and B frames can contain areas encoded using intra and inter prediction. As shown in [36], the B slices in HEVC bitstream contain a noticeable amount of bits representing intra-prediction modes. This leads to the conclusion that even in B frames, a significant part of PUs is encoded in intra mode. Unfortunately, the presence of intra-coded areas makes watermark embedding much more difficult. The intra-coded blocks use the edge of previously encoded blocks as a prediction source. In the case of embedding leading to alternate edge pixels, the intra prediction protracts the distorted values over the entire block. This leads to emerge of errors looking like streaks or dump patches. To make things worse, these distorted areas tend to flicker which makes them noticeable and annoying to viewers. Therefore, significant effort has to be taken to avoid the distortion introduced by intra-coded Prediction Units (PU).

The most efficient way of avoiding the above-mentioned errors is to keep intra-prediction sources unaltered. This leads to the development of a dedicated drift-optimized embedding algorithm. The proposed algorithm is designed to reduce computational complexity. At the first step, some rough check is made and more compute-heavy steps are taken only if necessary.

The watermarked picture is processed at the Coding Unit (CU) level. In the first step, all intra-encoded CUs are excluded from embedding. In the next step, inter-encoded CUs are examined. For every CU the neighboring blocks are identified. The intra prediction uses below-left, left, above-left, above and above-right neighboring unit edges as prediction signals. Therefore, for each watermarked CU, below, below-right, and right neighboring CUs have to be identified. If non of mentioned neighboring CUs is intra-encoded, the processed CU can be watermarked without any precautions.

In the opposite situation, the embedder checks if there is a risk of introduction of intra-related distortion. In the first step, the intra-prediction mode of all neighboring (below, below-right and right) intra-encoded CUs are checked. The information about the intra-prediction mode is then used to estimate if the currently processed block is a prediction source for neighboring units. For example: if a right neighbor of the currently processed block uses intra mode 34 (so-called right-top, see [37]) the currently processed block will not be a prediction source and the change of edge pixel values does not influence the right neighbor CU. If the investigation of all neighboring units intra mode leads to the conclusion that there is no risk of drift introduction, the processed CU can be also watermarked.

In the case where both above-mentioned paths are not available, the watermark embedding process is continued at the Transform Unit (TU) level. The inter-coded CU can have one or more TUs containing prediction errors. In the case of CU divided into more than one TU, the TUs are processed as follows: If TU does not belong to the right or bottom edge of CU, the watermark is embedded. If TU contacts the right or bottom edge, the previously described procedure is used to determine if the alternation of samples covered by TU can lead to intra-related drift. The decision of watermarking processed TU is made based on the calculated risk of drift introduction.

C. Sign hiding

In HEVC the new coding tool called Sign Data Hiding (SDH) was introduced [38]. The idea of SDH is to reduce the number of transmitted transform coefficient signs. In general, quantized transform coefficients are separated into sign and module and both sets of values are transmitted separately. Since the signs are unpredictable, SDH tries to hide one sign value in parity of transmitted coefficient amplitudes. This tool allows significantly increased compression efficiency. However, it reduces the flexibility and introduces some restrictions on transmitted coefficient amplitudes.

When SDH is enabled the sign of the last nonzero coefficient is embedded in the parity of the sum of the levels of the Coefficient Group (CG) [38]. Therefore, the parity of

the sum of the levels of the CG has to be preserved during the watermarking process.

D. Coefficient selection and watermark strength consideration

As mentioned in Section II the proposed method reuses the idea of alternating the low-frequency AC components. In [5] the rationale of this approach was described. In HEVC [8] four transform sizes were defined: 4x4, 8x8, 16x16, and 32x32. All transforms used in inter-coded blocks are DCT based. In order to preserve watermark consistency, the same range of spatial frequencies has been chosen to be alternated. In the case of the 4x4 transform, this means the lowest AC coefficients with indexes (0,1) and (1,0). Similarly, for 8x8 transform the corresponding coefficients are (0,2), (0,3), (2,0) and (3,0). Coefficients for 16x16 and 32x32 transforms have been selected following the same approach.

During the watermark embedding process, the selected AC coefficient is modified to change the distribution of spatial frequencies in a decoded picture. The modification is done by multiplying the quantized transform coefficient by the watermark strength (*ws*) factor. The multiplication result has to be rounded (up or down) since quantized transform coefficients are represented as integer values. Moreover, the value of the modified coefficient needs to be adjusted in order to preserve sign information hidden in the sum of the levels of the Coefficient Group (CG).

The *ws* factor can be used to control the strength of the watermark. The higher the *ws* value, the stronger the watermark is, but at the cost of higher distortion. During the development of the described algorithm, we decided to evaluate different values of *ws* and check its influence on watermark robustness and image quality degradation.

E. Watermark detection

The watermark detection uses the method described in [4]. The watermark detector operates in the pixel domain and requires two signals the reference one (original, non-watermarked video) and investigated one. Detector processes every sequence extracting per frame data about the energy of spatial frequencies within the frame. Those statistics are rather compact (several values per frame) and can be stored instead of the original video.

The detector averages the spectral energy statistics for images from a watermarked group and images from non-watermarked groups surrounding the watermarked one. In the next step, detection concentrates on spatial energy related to coefficients modified during the watermarking process. The difference between the energy of watermarked and non-watermarked parts of GOP in the investigated sequence is calculated and normalized by the same difference in the reference sequence. The increase of energy corresponding to selected horizontal or vertical frequencies is interpreted as the appearance of 0 or 1 in watermarked signal.

IV. Experimental evaluation of the proposed method

The proposed method has been evaluated experimentally. To perform the evaluation, a set of test sequences was prepared. The watermarking process was performed by software designed and developed by the authors. The software includes HEVC entropy decoder/encoder (VLC for headers and CABAC for slice data) and watermarking module processing the entropy decoded data. The software does not perform image reconstruction (decoding) so image pixels were not used at any stage of watermarking process. The embedding software has been developed to evaluate the proposed method and cannot be considered optimized in any way, regardless of the throughput of ~130Mbit/s was achieved using a single core of consumer-grade CPU.

A. Evaluation framework

The proposed method has been evaluated using a set of 6 diversified video sequences. Each sequence has a FullHD (1080p) resolution and a length of 12-15 minutes. The entire sequence has been used during the evaluation. The used sequences can be described as: anim1 and anim2 - computer-animated adventure comedy films; epic1 – vintage, epic, historical drama; soap1 and soap2 – soap opera; war1 – modern war movie. Due to the low density of embedded watermark, we can not use typical test sequences recommended by the MPEG group since these sequences are too short (up to 12 seconds). The watermark strength (*ws*) multiplier values from set {1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8 3.0} have been used.

The experiment has been performed starting from an uncompressed sequence. In the first step, HEVC compressed bitstreams were prepared using the x265 [39] encoder. The x265 has been set to produce 8Mbit/s bitrate, operate using a "veryslow" preset, and use PSNR as a quality metric for rate-distortion optimization. The chosen x265 parameters guaranteed high compression efficiency and very good quality. The quality (measured as luma PSNR) and precise bitrate of each encoded sequence have been measured.

In the next step, all previously encoded sequences have been watermarked using all previously mentioned watermark strength values. Moreover, for each sequence and each *ws* value, a set of four random messages was used. As a result, we prepared a set of 240 watermarked bitstreams (corresponding to about 45 hours of video material).

Since the proposed method aims to create a robust watermark we decided to perform some attacks and measure the watermark's robustness. Experiments described in [4] and [5] show that one of the most difficult scenarios is when content is displayed on a TV and recorded by a camera (or camcorder). In such a case the decoded watermarked video is a subject of very strong distortion caused by: post-processing performed by TV (i.e. additional deblocking, contrast amplification, etc), geometrical distortion, dynamic range lost during capturing by a camera, camera side pre-processing, and lossy compression. All these sources of distortion lead to the creation of a scenario that is very demanding for robust watermarking. Unfortunately, the high amount of video material makes it very difficult to perform experiments using content recorded from TV by a camcorder. Therefore, we decided to simulate such an attack. The test recordings and the simulation framework based on subjective evaluation were prepared. The simulation framework aimed to mimic the distortion caused by the TV-to-camera path. The simulation includes: a strong deblocking filter (in 8x8 regions), slight image rotation, dynamic range clip to 0.1-0.9 of the original range, dynamic range expansion (contrast enhancement), HEVC encoding with ½ of original bitrate with the usage of "fast" preset.

All watermarked and attacked sequences have been processed by a watermark detector. The robustness of the watermark has been evaluated by measuring the percent of

correctly recovered bits. Moreover, the watermarked sequences have been decoded and their quality was measured (PSNR) to evaluate the influence of the watermarking process on image quality. The entire evaluation framework has been summarized in Fig. 4.

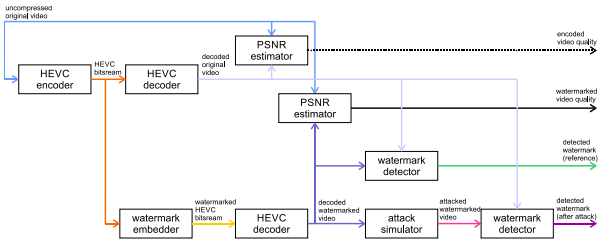


Fig. 4. Visualization of experimental evaluation workflow.

As a result watermark robustness was tested against two scenarios:

- Scenario1 – corresponding to screen-captured content (i.e. VODRip): HEVC reencoding with ½ bitrate and "fast" preset,
- Scenario2 – corresponding to camera recorded content (i.e. CamRip/Telecine) uses the previously mentioned attack simulation framework.

B. Evaluation results

The watermark robustness evaluation results were summarized in TABLE I. The results are averaged over all four messages. For the Scenario1 the *ws* values as low as 1.6 provide acceptable robustness. If camera recording is considered (Scenario2), the required *ws* has to be increased to the 2.2-2.4 range.

TABLE I. DETECTION RESULTS (AS A PERCENT OF CORRECTLY DETECTED BITS)

	Correctly detected bits for Scenario 1[%]									
	Watermark strength									
Sequence	***1.2***	***1.4***	***1.6***	***1.8***	***2.0***	***2.2***	***2.4***	***2.6***	***2.8***	***3.0***
anim1	53.7	77.5	92.7	93.8	92.6	97.9	98.0	98.0	98.0	98.1
anim2	52.5	72.9	89.9	90.5	89.3	97.6	97.7	98.5	98.3	97.7
epic1	56.4	81.0	98.7	98.5	98.2	99.8	99.8	99.8	99.8	99.8
soap1	52.8	70.0	95.0	96.0	94.5	99.9	99.9	99.9	99.9	99.9
soap2	54.4	78.5	95.4	96.1	94.9	99.9	99.9	99.9	99.9	99.9
war1	53.3	67.3	84.2	85.0	83.9	95.1	95.1	95.6	95.6	95.6
Average	53.9	74.5	92.6	93.3	92.2	98.4	98.4	98.6	98.6	98.5

	Correctly detected bits for Scenario 2[%]									
	Watermark strength									
Sequence	***1.2***	***1.4***	***1.6***	***1.8***	***2.0***	***2.2***	***2.4***	***2.6***	***2.8***	***3.0***
anim1	50.3	60.8	77.0	79.0	76.7	92.9	92.9	93.8	93.8	93.8
anim2	50.9	61.0	76.2	77.7	75.9	91.5	91.4	92.5	92.5	92.4
epic1	52.1	71.3	97.1	97.0	97.2	99.5	99.5	99.6	99.5	99.5
soap1	51.2	58.3	73.9	75.9	73.2	96.6	96.7	97.5	97.7	97.6
soap2	51.4	57.3	68.9	71.1	68.5	95.4	95.2	96.4	96.6	96.7
war1	50.7	57.6	71.0	73.3	70.6	90.2	90.3	91.1	91.3	91.3
Average	51.1	61.0	77.3	79.0	77.0	94.3	94.3	95.1	95.2	95.2

The TABLE II. and TABLE III. summarize the bitrate and quality changes caused by watermarking process.

Assuming that *ws* =1.6 and *ws*=1.8 are the most suitable for Scenario1, the watermarking-related bitrate increase is negligible (~0.3%) and can be neglected. The quality degradation is about 0.3dB which can be considered very low.

For Scenario2 the most suitable watermark strengths are *ws*=2.2 and *ws*=2.4. In this case, the watermarking process increases the resulting bitrate by ~1.8% and causes ~1.1dB quality loss. The bitrate increase is not significant and quality loss seems to be an acceptable tradeoff for camera recording robustness.

TABLE II. RELATIVE BITRATE INCREASE

	Bitrate increase [%]									
	Watermark strength									
Sequence	***1.2***	***1.4***	***1.6***	***1.8***	***2.0***	***2.2***	***2.4***	***2.6***	***2.8***	***3.0***
anim1	0.005	0.064	0.272	0.298	0.253	1.567	1.564	1.728	1.745	1.765
anim2	0.003	0.057	0.285	0.305	0.265	1.579	1.577	1.745	1.765	1.779
epic1	0.004	0.060	0.394	0.415	0.373	3.187	3.185	3.404	3.424	3.440
soap1	0.005	0.061	0.277	0.300	0.255	1.709	1.706	1.880	1.901	1.919
soap2	0.006	0.077	0.312	0.342	0.285	1.635	1.630	1.829	1.854	1.879
war1	0.004	0.042	0.201	0.217	0.188	1.064	1.062	1.184	1.196	1.210
Average	0.004	0.060	0.290	0.313	0.270	1.790	1.787	1.962	1.981	1.999

TABLE III. WATERMARKING INFLUENCE INTO IMAGE QUALITY

	Decoded image quality change (delta PSNR) [dB]									
	Watermark strength									
Sequence	***1.2***	***1.4***	***1.6***	***1.8***	***2.0***	***2.2***	***2.4***	***2.6***	***2.8***	***3.0***
anim1	-0.011	-0.088	-0.235	-0.319	-0.262	-1.008	-1.039	-1.330	-1.422	-1.524
anim2	-0.006	-0.069	-0.210	-0.277	-0.227	-0.909	-0.936	-1.225	-1.317	-1.406
epic1	-0.009	-0.091	-0.322	-0.404	-0.342	-1.630	-1.656	-1.999	-2.087	-2.168
soap1	-0.011	-0.088	-0.251	-0.331	-0.270	-1.083	-1.114	-1.418	-1.511	-1.595
soap2	-0.013	-0.111	-0.303	-0.413	-0.335	-1.174	-1.215	-1.590	-1.712	-1.842
war1	-0.009	-0.066	-0.182	-0.247	-0.201	-0.722	-0.748	-0.980	-1.047	-1.135
Average	-0.010	-0.085	-0.251	-0.332	-0.273	-1.087	-1.118	-1.424	-1.516	-1.612

C. Objective quality of watermarked sequences

The huge amount of test materials rendered the subjective evaluation very difficult. Requiring viewers to evaluate 45h of video materials is almost impossible. Therefore, we decided to perform a simplified subjective evaluation.

The subjective test participants were unaware of the type of artifacts introduced by watermarking process. Viewers were asked to watch one sequence of each type for different watermark strengths (1.6, 1.8, 2.2, 2.4) and decide if they see any noticeable distortion or annoying artifacts. No artifacts were noticed and most participants considered differently watermarked versions of one sequence as repetition of the same content.

V. CONCLUSIONS

In the paper, an original watermarking method is described for HEVC-compressed video. The main idea of the technique is to embed a message into nonzero transform coefficients in such a way that potential drift is limited. The complexity of the proposed solution is significantly lower than pixel domain watermarking (so it is a relatively fast approach), leading to huge energy savings by omitting full decoding and re-encoding. The experimental verification has been performed on six test sequences for different watermark strengths against two strong attacks modeling modifications caused by typical techniques used to create illegal copies of videos. The experimental verification demonstrates the strong robustness of the embedded watermarks with unnoticeable quality degradation.

ACKNOWLEDGEMENT

The research was supported by the Ministry of Education and Science of the Republic of Poland under the subvention for research SBAD.

REFERENCES

[1] D. Karwowski, T. Grajek, K. Klimaszewski, O. Stankiewicz, J. Stankowski, K. Wegner, "20 Years of Progress in Video Compression - from MPEG-1 to MPEG-H HEVC. General View on the Path of Video Coding Development," Book: Image Processing and Communications Challenges 8, Editors: Ryszard S. Choras, Advances in Intelligent Systems and Computing: No. 525, Springer International, pp. 3-15, ISBN: 978-3-319-47273-7, ISSN: 2194-5375, 2016.

[2] G. Doërr, J.-L. Dugelay: "A guide tour of video watermarking", Signal Processing: Image Communication vol. 18, no. 4, pp.263-282, Elsevier 2003.

[3] F. Hartung, B. Girod: "Watermarking of Uncompressed and Compressed Video", Signal Processing, vol. 66, no. 3, pp. 283-301, Elsevier, 1998.

[4] J. Stankowski, T. Grajek, M. Domański, "Fast watermarking of MPEG-4 AVC/H.264 encoded HDTV video bitstreams", 29th Picture Coding Symposium, PCS 2012, Kraków, Poland, 7-9 May 2012, pp. 265-268.

[5] J. Siast, J. Stankowski, T. Grajek, M. Domański, "Digital Watermarking with Local Strength Adjustment for AVC-Compressed HDTV Bitstreams", 30th Picture Coding Symposium, PCS 2013, San Jose, CA, USA, 8-11 December 2013, pp. 53-56.

[6] Y. Tew, K. Wong, R. C.-W. Phan, and K. N. Ngan, "Separable authentication in encrypted HEVC video," in Multimedia Tools and Applications, pp. 1–20, 2018.

[7] A. A. Elrowayati, M. F. L. Abdullah, A. A. Manaf, and A. S. Alfagi, "Tampering detection of double-compression with the same quantization parameter in HEVC video streams," in 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 174–179, 2017.

[8] ISO/IEC 23008-2, MPEG-H Part 2, High-Efficiency Video Coding, Edition 4, 2020, also: ITU-T Rec. H.265 Edition 8.0, 2021.

[9] ISO/IEC 14496-10, MPEG-4 Part 10, Generic coding of audio-visual objects, Advanced Video Coding. Edition 9, 2020, also: ITU-T Rec. H.264 Edition 14.0, 2021.

[10] T. Dutta and H. P. Gupta, "An efficient framework for compressed domain watermarking in P-frames of high-efficiency video coding (HEVC)–encoded video," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 13, no. 1, p. 12, 2017.

[11] P.-C. Chang, K.-L. Chung, J.-J. Chen, C.-H. Lin, and T.-J. Lin, "A DCT/DST-based error propagation-free data hiding algorithm for HEVC intra-coded frames," Journal of Visual Communication and Image Representation, vol. 25, no. 2, pp. 239–253, 2014.

[12] S. Swati, K. Hayat, and Z. Shahid, "A watermarking scheme for high efficiency video coding (HEVC)," PloS one, vol. 9, no. 8, pp. 1–8, 2014.

[13] S. Gaj, A. Sur, and P. K. Bora, "A robust watermarking scheme against re-compression attack for H.265/HEVC," in Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4, 2015.

[14] K. Ogawa and G. Ohtake, "Watermarking for HEVC/H.265 stream," in IEEE International Conference on Consumer Electronics (ICCE), pp. 102–103, 2015.

[15] Y. Tew, K. Wong, and R. C.-W. Phan, "HEVC video authentication using data embedding technique," in IEEE International Conference on Image Processing (ICIP), pp. 1265–1269, 2015.

[16] T. Dutta and H. P. Gupta, "A robust watermarking framework for high efficiency video coding (HEVC)–encoded video with blind extraction process," Journal of Visual Communication and Image Representation, vol. 38, pp. 29–44, 2016.

[17] A. A. Elrowayati, M. F. L. Abdullah, A. A. Manaf, and A. S. Alfagi, "Robust HEVC video watermarking scheme based on repetition-BCH syndrome code," International Journal of Software Engineering and Its Applications, vol. 10, no. 1, pp. 263–270, 2016.

[18] Y. Tew, K. Wong, R. C.-W. Phan, and K. N. Ngan, "Multi-layer authentication scheme for HEVC video based on embedded statistics," Journal of Visual Communication and Image Representation, vol. 40, pp. 502–515, 2016.

[19] S. Gaj, A. Kanetkar, A. Sur, and P. K. Bora, "Drift-compensated robust watermarking algorithm for H.265/HEVC video stream," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 13, no. 1, pp. 1–24, 2017.

[20] G. Jiang, J. Wang, M. Yu, and F. Chen, "Zero-watermarking registration and detection method for HEVC video streaming against requantization transcoding," U.S. Patent 9,947,065.

[21] K. Joa, W. Lei, and Z. Lic, "A watermarking method by modifying QTCs for HEVC," Computing, Performance and Communication systems, pp. 8–16, 2018.

[22] G. Kaur, S. S. Kasana, and M. Sharma, "An efficient watermarking scheme for enhanced high efficiency video coding/H.265," Multimedia Tools and Applications, pp. 1–23, 2018.

[23] Y. Liu, H. Zhao, S. Liu, C. Feng, and S. Liu, "A robust and improved visual quality data hiding method for HEVC," IEEE Access, vol. 6, pp. 53 984–53 997, 2018.

[24] A. A. Mohammed and N. A. Ali, "Robust video watermarking scheme using high efficiency video coding attack," Multimedia Tools and Applications, vol. 77, no. 2, pp. 2791–2806, 2018.

[25] A. Joshi, V. Jain, S. Ladda, and R. Kumar, "Real-time implementation of blind and robust watermarking for HEVC video coding," in 2018 IEEE International Symposium on Smart Electronic Systems (iSES), pp. 58–63, 2018.

[26] C. Wang, R. Shan, and X. Zhou, "Anti-HEVC recompression video watermarking algorithm based on the all phase biorthogonal transform and SVD," IETE Technical Review, vol. 35, no. sup1, pp. 42–58, 2018.

[27] X. Yu, C. Wang, and X. Zhou, "A hybrid transforms-based robust video zero-watermarking algorithm for resisting high efficiency video coding compression," IEEE Access, vol. 7, pp. 115 708–115 724, 2019.

[28] T. Shanableh, "Data embedding in HEVC video by modifying the partitioning of coding units," IET Image Processing, vol. 13, no. 11, pp. 1909–1913, 2019.

[29] G. Kaur, S. S. Kasana, and M. Sharma, "An efficient watermarking scheme for enhanced high efficiency video coding/H.265," Multimedia Tools and Applications, vol. 78, no. 9, pp. 12 537–12 559, 2019.

[30] G. Kaur, S. S. Kasana, and M. Sharma, "An efficient authentication scheme for high efficiency video coding/H.265," Multimedia Tools and Applications, vol. 78, no. 15, pp. 21 245–21 271, 2019.

[31] D. R. Galiano, A. Del Barrio, G. Botella, and D. Cuesta, "Efficient embedding and retrieval of information for high-resolution videos coded with HEVC," Computers & Electrical Engineering, vol. 81, p. 106541, 2020.

[32] M. N. Favorskaya and V. V. Buryachenko, "Authentication and copyright protection of videos under transmitting specifications," in Computer Vision in Advanced Control Systems-5, pp. 119–160, 2020.

[33] M. Z. Konyar, O. Akbulut, and S. Öztürk, "Matrix encoding-based high-capacity and high-fidelity reversible data hiding in HEVC," Signal, Image and Video Processing, pp. 1–9, 2020.

[34] S. Gaj, A. Sur, and P. K. Bora, "Prediction mode based H.265/HEVC video watermarking resisting re-compression attack," Multimedia Tools and Applications, pp. 1–31, 2020.

[35] Y. Zhou, C. Wang, and X. Zhou, "An intra-drift-free robust watermarking algorithm in high efficiency video coding compressed domain," IEEE Access, vol. 7, pp. 132 991–133 007, 2019.

[36] J. Stankowski, D. Karwowski, T. Grajek, K. Wegner, J. Siast, K. Klimaszewski, O. Stankiewicz, M. Domański, "Bitrate Distribution of Syntax Elements in the HEVC Encoded Video", International Conference on Signals and Electronic Systems, ICSES 2014, Poznań, Poland, 11-13 September 2014, ISBN: 978-1-4799-7009-4.

[37] J. Lainema, F. Bossen, W. Han, J. Min and K. Ugur, "Intra Coding of the HEVC Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1792-1801, Dec. 2012, doi: 10.1109/TCSVT.2012.2221525.

[38] J. Sole *et al*., "Transform Coefficient Coding in HEVC," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1765-1777, Dec. 2012, doi: 10.1109/TCSVT.2012.2223055.

[39] x265 HEVC Encoder / H.265 Video Codec, https://www.x265.org/

Fast Watermarking of MPEG-4 AVC/H.264 Encoded HDTV Video Bitstreams

Jakub Stankowski, Tomasz Grajek, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology,
Poznan, Poland
{jstankowski, tgrajek}@multimedia.edu.pl, domanski@et.put.poznan.pl

Abstract— **The paper describes watermarking of compressed video bitstreams. This technique is proven to be very fast and robust against various attacks including camcorder recording attack. The watermark is embedded using indirect modifications of transform coefficients in selected P- and B-frames. In the watermarked bitstream there exist sequences of watermarked pictures interleaved with sequences of unmodified pictures. The unmodified pictures serve as a reference for the watermark detection. The experimental results described for real-world MPEG-4 AVC/H.264 encoded HDTV video bitstreams are included.**

Keywords- watermarking, compressed domain watermarking, HDTV, MPEG-4 AVC, H.264, camcorder recording attack.

I. INTRODUCTION

Great improvement achieved in high quality flat-panel display and digital video camera technologies allows users to record videos displayed on HDTV with little quality degradation. Distribution of illegal copies, made using such a method, may have a very significant economic influence on the "movie market". Digital watermarking is one of the tools for protecting digital copyrights. This paper refers to watermarking for detection of illegal camcording of video content. Therefore, the watermarks should resist camcording, i.e. camcording attacks.

In literature one may find several studies dealing with camcorder attacks [4-7]. Unfortunately, most of these solutions are of high computational complexity mainly due to using uncompressed video content in a watermark embedder. Therefore, full decoding and re-encoding of video content is needed [7,8]. Moreover, in the vast majority of papers all experiments are conducted on low resolution video materials and achieved results cannot be directly referred to HDTV video materials.

In this paper, we propose a very fast watermarking method for MPEG-4 AVC/H.264 [1] bitstreams robust to very strong attacks, e.g. video camcorder recording. This method does not need full decoding and again re-encoding of the video content. Even full entropy decoding and re-encoding are not needed. Only selected headers of the selected pictures have to be modified, regardless of the used context-adaptive entropy method (CAVLC or CABAC) [2]. Watermarking is designed to have minimal influence on video quality but, on the other hand, watermark has to be detectable after strong attacks including camcorder recording attack.

The main idea behind the proposed method is to embed a watermark into a compressed MPEG-4 AVC/H.264-compliant bitstream, without the need of full entropy de- and encoding of the bitstream. The proposal is to achieve this goal by computationally inexpensive indirect modifications of residual transform coefficients obtained by changing the quantization matrices. To be precise, only non-zero transform coefficients can be changed. Quantization matrix modification affects the whole image, there is no possibility to control the changes locally. This operation affects the values of residual signal in the decoder and thus modifies the reconstructed image. The idea is to change coefficients in the first row and the first column of the transform block.

As a matter of fact, the idea of watermarking by using quantization matrices has been already presented for MPEG-2 bitstreams [4,5]. However, as compared to the abovementioned methods, the proposed method provides an original approach to selection of the embedded coefficients and watermark detection.

II. PROPOSED METHOD

A. General idea

In MPEG-4 AVC/H.264 standard terminology, equivalent for the quantization matrix is scaling matrix. Therefore, in this article those terms will be used interchangeably.

The MPEG-4 AVC/H.264 standard defines default scaling matrices (quantization matrices) for luma and chroma components. Nevertheless, in the High profile, the standard provides a possibility to use custom quantization matrices for transform coefficients. Such matrices can be added to the bitstream in the sequence and/or picture parameter sets. A custom scaling matrix could be sent separately for intra and inter macroblocks and separately for luma and chroma components.

B. Profile issue

In MPEG-4 AVC/H.264, the quantization matrix (scaling matrix) transmission is possible in the High profile. In HDTV

broadcasts, sequences compliant with the High profile are mostly used. If the Main profile bitstream is to be watermarked, this bitstream has to be converted to the High profile. The set of compression tools of the High profile is a superset of those for the Main profile [3]. Therefore, every Main profile bitstream can be converted to the High profile by modifying the *profile_idc* in Sequence Parameter Set (SPS) and adding some flags required in the High profile. This modification does not have any effect on the reconstructed image but makes the stream decodable only by the High profile compatible decoder (standard HDTV decoder is compatible with the High profile of MPEG-4 AVC/H.264).

C. Watermark embedding

In MPEG-4 AVC/H.264, the syntax element containing scaling matrix is called scaling list [1]. In the Main profile bitstream (converted to the High profile), the watermark may be embedded only with use of 4x4 scaling matrices (only 4x4 transform size is available). In the High profile bitstream, the watermark can be embedded with use of both 4x4 and 8x8 scaling matrices in order to achieve sufficient watermark detectability.

For watermarking purposes, low-frequency AC coefficients were chosen, because we found that lower frequencies are more robust to attacks such as camcorder recording or transcoding. This observation is compliant with those already reported in the references [9]. Therefore, (0,1) and (1,0) coefficients have been selected for the 4x4 transform. These coefficients are not distorted by harmonics produced from other coefficients in non-linear systems that model the attacks. In 4x4 and 8x8 scaling matrices, the modified coefficients should correspond to similar spatial frequencies. Therefore, (0,2), (0,3), (2,0) and (3,0) coefficients have been selected for the 8x8 transform (see Fig. 1).

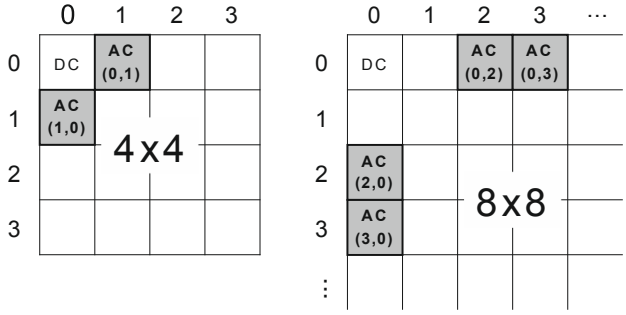


Figure 1. The modified transform coefficients, for 4x4 and 8x8 block size, respectively.

The watermark may be inserted into I-, P-, or B-frames. The scaling matrices for luma or/and chroma may be modified separately. After several experiments, we decided to embed the watermark only in the luma component in P- and B-frames. Watermarking of the chroma components cannot provide sufficient energy of watermark because the vast majority of transform coefficients are equal to zero. The main advantage of P-frame watermarking is a sufficient number of non-zero transform coefficients that can be modified in most of the macroblocks and limited subjective-quality image degradation. In B-frames, there is only a small number of non-zero transform coefficients in most of the macroblocks, but watermarking of those frames provides better detection results without noticeable image quality degradation. Watermarking of I-frames by modification of scaling matrices results in significant image quality degradation mainly due to drift caused by intra prediction. It should be kept in mind that intra predicted macroblocks may be used by the encoder in all types of frames. Thus, the intra predicted blocks might be present in the watermarked sequences. The proposed method does not guarantee avoiding of intra prediction drift but it is possible to keep it unnoticeable by properly adjusting the watermark strength.

In MPEG-4 AVC/H.264, scaling matrix may be sent in Picture Parameter Set (PPS). In our proposal, modified scaling matrices are sent in additional PPSs. Once modified, the watermarked bitstream contains one original PPS and two additional PPSs. One of the latter contains a scaling matrix with increased vertical frequency coefficients (PPS-V) and the other contains a scaling matrix with increased horizontal frequency coefficients (PPS-H). This set of PPSs may be used to embed the watermark. Embedding process is basically a modification of slice headers in order to point out one of the three PPSs. For the unmodified slices the original PPS has to be indicated whereas for the watermarked slices one of the modified PPS (PPS-H or PPS-V) has to be chosen. Change of the PPS is done by *pic_parameter_set_id* field modification in slice header. Indicated PPSs will be used to decode the given slice, and modify appropriate coefficients. An example of the original and watermarked bitstream is shown in Fig.2.

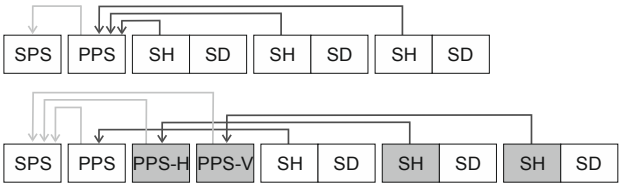


Figure 2. Bitstream modifications made during watermarking process: orginal bitstream (top), watermarked bitstream (bottom).

D. Embedding complexity and bitstream overhead

Because embedding is a simple slice header PPS swaping, the proposed watermarking technique is of very low computational complexity. Only PPSs and slice headers (SH) must be decoded, modified and encoded, but these units are very small (in tested bitstreams: SPS ~26 bytes, PPS ~5 bytes and SH ~5 bytes) and easy to encode. These units are encoded using fast Exp-Golomb code and may be processed without using CABAC or CAVLC [1]. Assuming structure of group of pictures (GOP) typical for broadcasting, PPS and SPS transmitting before every I-frame (to be precise – IDR-frame) and 30 fps, there is only about 360 bytes per second to tdecode. The computational requirements for the embedder are below 1 MIPS.

Bitstream overhead caused by the watermarking process is also very small. The additional PPS (with added scaling list syntax) has about 50 bytes, which means that about 200 bps

have to be added to the bitstream. For 10 Mbps bitstream, the watermarking overhead will be about 0.016%.

E. Watermark detection

Watermark detection is performed using the original and the investigated video sequence. Both sequences must be temporally aligned, but spatial alignment is not critical. Even temporal alignment does not need to be extremely accurate – shifts of 1-2 frames may be accepted if the sequences of watermarked frames are long enough. Temporal synchronization may be obtained using any of the known techniques.

During watermark embedding (0,2) and (0,3) or (2,0) and (3,0) coefficients have been modified (see section II.C). Therefore, the proposed detection algorithm uses transform coefficients energy to retrieve the watermark. The detector's input sequences (the original and the investigated) have to be analyzed in spectral domain. We decided to use the 8x8 transform as defined in MPEG-4 AVC/H.264. The averaged energy of each coefficient over 8x8 blocks in a frame is estimated as follows:

$$\overline{E_{i,j}} = \frac{1}{N}\sum_{k}(a_{i,j}^{k})^2 \tag{1}$$

where $a_{i,j}$ is the amplitude of the (i,j)-*th* transform coefficient in *k-th* block, k is block number, N is the number of blocks in a single frame.

We assumed that the analyzed video sequence contains consecutive groups of watermarked and non-watermarked images. An example of image sequence from investigated video is shown in Fig. 3.

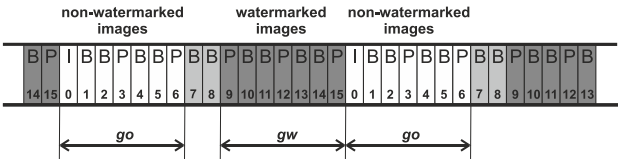


Figure 3. Example of investigated video sequence. Dark frames have been watermarked. Sequence contains non-watermarked groups of frames (*go1* and *go2*) and watermarked groups of frames (*gw*).

In order to detect even a very slight watermark, strongly masked by marked data, content influence on the watermark has to be eliminated. In the proposed method it is assumed that transform coefficient energy distribution in transform coefficient being under investigation is quasi stable in consecutive frames. Therefor only modification of energy distribution is caused by watermarking. We compare transform coefficient energy distribution of investigated sequence with energy distribution of original sequence. First, the energy of coefficients is averaged separately for images from a watermarked group and for images from non-watermarked groups surrounding the watermarked group:

$$G_{i,j}^{W} = \sum_{l\in gw}\overline{E_{i,j}}^{l}, \tag{2}$$

$$G_{i,j}^{O} = \frac{1}{2}\left(\sum_{l\in go1}\overline{E_{i,j}}^{l} + \sum_{l\in go2}\overline{E_{i,j}}^{l}\right) \tag{3}$$

where: $G^{W}_{i,j}$ is an averaged energy for the watermarked images, $G^{O}_{i,j}$ is an averaged energy for non-watermarked images, l is the image number, i,j are coefficients' coordinates in transform block. Next, changes of the energy distribution between original (org) and investigated (cam) sequences are calculated as follows:

$$S_{i,j} = \frac{G_{i,j}^{w}(cam)}{G_{i,j}^{O}(cam)} - \frac{G_{i,j}^{w}(org)}{G_{i,j}^{O}(org)} \tag{4}$$

where: $S_{i,j}$ is a differential energy of the (i,j)-*th* coefficient.

Detector response bases on selected $S_{i,j}$ coefficients corresponding to the watermarked set of transform coefficients. As mentioned before, watermark was embedded by modifying (0,2) and (0,3) or (2,0) and (3,0) coefficients. Moreover, the camcorder recording causes that the watermark energy partially flows to (1,2), (1,3) or (2,1), (3,1) transform coefficients. Therefore, detector response has been defined as a weighted average of individual decision results:

$$\begin{aligned} r \;=\; & a\cdot\mathrm{sgn}(S_{0,2}-S_{2,0}) + b\cdot\mathrm{sgn}(S_{0,3}-S_{3,0}) \;+ \\ + \; & c\cdot\mathrm{sgn}(S_{1,2}-S_{2,1}) + d\cdot\mathrm{sgn}(S_{1,3}-S_{3,1}) \end{aligned} \tag{5}$$

where: r is detector response; *sgn(x)* is sign function; a, b, c and d are weight matched to data statistics. Detector response $r<0$ means that vertical frequency coefficients have been modified, $r>0$ means that horizontal frequency coefficients have been modified.

III. EXPERIMENTAL RESUSTS

A. Testing method

Tests have been performed on HDTV (1920x1080, 25 fps) video sequences of the length of 10 minutes each. Test sequences were encoded using x264 [10] encoder, configured to produce the MPEG-4 AVC/H.264 High-profile-compliant bitstreams. The prepared bitstreams had been watermarked and displayed on an LCD HDTV set. The displayed sequences were recorded by a HD camcorder. Afterwards, the recorded movie has been decoded and passed to the input of the watermark detector.

The watermark was embedded by changing the scaling matrix for P- and B-frames in the consecutive groups of pictures (GOP). For test purposes, two different random bit streams (A and B in Table 1) were used as the watermark payload. The watermark embedder was also set to mark the second half of GOP only (for example: when GOP length is equal to 16 only the last 3 P-frames and last 4 B-frames are modified – see Fig. 4). Each watermarked GOP corresponds to one bit from the watermark payload.



Figure 4. Example of watermarked GOP. Dark P- and B-frames are watermarked.

B. *Perceptual quality*

A watermark embedded using the proposed method is rather not visible. The most common effect would be seen as flickering spots related to positions where intra predicted macroblocks were used. These defects are hardly visible in still images but, when watermark is strong, they are visible and disturbing in video sequences.

Watermark visibility tests were performed and in most cases the watermark was rated as unnoticeable. Intra prediction related errors were evaluated weak and not frequent. In general, the watermark is not disturbing and does not stand out while watching test sequences.

C. *Robustness*

Watermark robustness tests were performed by recovering watermark payload from watermarked sequence. The effectiveness of watermark detection was calculated as a comparison of both: the detector response and embedded watermark data, in order to illustrate how many bits were detected correctly. Four different scenarios were tested to check robustness against most popular attacks:

- Scenario 1 (attack simulation): Low-pass filtering, 10% image cropping and 10% image resizing,
- Scenario 2: Camcording using HD camera,
- Scenario 3: Camcording using HD camera and downscaling to SD resolution,
- Scenario 4: Camcording using HD camera and framerate reduction (from 25 to 12,5 fps).

The effectiveness of watermark detection was collected in Table 1.

TABLE I. EXPERIMENTAL RESULTS

Sequence type	**Payload bitstream**	**Properly detected bits [%] in scenario:**			
		1.	***2.***	***3.***	***4.***
Soap opera 1	A	97.7	88.5	83.3	88.4
	B	97.4	87.6	82.0	87.8
Soap opera 2	A	98.4	88.5	88.5	88.5
	B	98.7	92.4	87.3	91.1
Historical novel	A	99.8	97.4	96.3	96.7
	B	99.9	99.6	98.7	99.2
Computer-animated film	A	96.6	87.2	82.7	84.7
	B	97.0	87.7	82.8	86.0
War film	A	94.9	76.2	72.4	73.8
	B	95.4	75.0	73.9	74.6
Average	A	97.5	87.6	84.6	86.4
	B	97.7	88.5	84.9	87.7

In Scenario 1, were combination of several attacks was simulated, more than 95% of bits was properly detected. For scenarios, were real HD camcorder was used, properly detected bits ratio is not lower than 85%.

IV. CONCLUSIONS

The main advantage of the proposed method is very low computational complexity of the embedder. Therefore, it may be considered as a very fast watermarking solution. Additionally, increase of the watermarked bitstream size in comparison to original one is negligibly small. The results show that proposed method is robust against blurring, camcording, downscaling to SD resolution and downsampling to lower framerate. Our method is suitable for implementation on low processing power devices or systems, unlike other watermarking methods known from literature.

REFERENCES

[1] ISO/IEC 14496-10: Inform. Tech. – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding (2010).

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560 –576 (2003).

[3] D. Marpe, S. Gordon, T. Wiegand, "H.264/MPEG4-AVC fidelity range extensions: tools, profiles, performance, and application areas", IEEE International Conference on Image Processing, ICIP 2005, pp. 593 – 596, Genova, Italy, (2005).

[4] M. Celik, J. Talstra, A.N. Lemma, S. Katzenbeisser, "Camcorder Capture Robust Low-Complexity Watermarking of MPEG-2 Bit-Streams", IEEE International Conference on Image Processing, ICIP 2007, pp. 489-492, San Antonio, Texas, USA (2007).

[5] D. Choi, H. Do, H. Choi, T. Kim, "A blind MPEG-2 video watermarking robust to camcorder recording", Signal Processing, vol. 90, no. 4, pp. 1327-1332, Elsevier (2010).

[6] Y. Wang, A. Pearmain, "Blind MPEG-2 video watermarking robust against geometric attacks: a set of approaches in DCT domain", IEEE Transactions on Image Processing, vol. 15, no. 6, pp. 1536-1543 (2006).

[7] H. Do, D. Choi, H. Choi, T. Kim, "Digital Video Watermarking Based on Histogram and Temporal Modulation and Robust to Camcorder Recording", IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2008, pp. 330-335, Sarajevo, Bosnia & Herzegovina (2008).

[8] F. Hartung, B. Girod, "Watermarking of Uncompressed and Compressed Video", Signal Processing, vol. 66, no. 3 (Special issue on Watermarking), pp. 283-301, Elsevier (1998).

[9] E. Nezhadarya, Z.J. Wang, R.K. Ward, "Watermark survival chance (wsc) concept for improving watermark robustness against JPEG compression", IEEE International Conference on Image Processing, ICIP 2010, pp. 3697-3700, Hong Kong (2010).

[10] Source code and description available on http://x264.nl/

Digital Watermarking with Local Strength Adjustment for AVC-Compressed HDTV Bitstreams

Jakub Siast, Jakub Stankowski, Tomasz Grajek, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology
60-965 Poznań, Polanka 3 Street, POLAND
{jsiast,jstankowski,tgrajek}@multimedia.edu.pl

***Abstract*— A computationally inexpensive watermark schema for MPEG-4 AVC/H.264 streams that use local mark strength adjustment minimizing visible artifacts is presented. The local mark strength adjustment mechanism and watermark embedding mechanism operate on residual data. This approach minimize the number of operations needed for video bitstream decoding and subsequent watermarked video data encoding. The experimental results included herein, described for real-world MPEG4 AVC/H.264 encoded HDTV video bitstreams prove its robustness to complex attacks, especially camcorder recording attack.**

I. INTRODUCTION

Along with development of digital TV systems and digital video cameras, quality of video material recorded from HDTV became better and better. Nowadays, it is possible to make a copy of video displayed on HDTV set with little quality degradation. Therefore, a method to protect movies with many-millions budget against illegal camcording and distribution in internet is highly desired. Watermarking is one of the methods that has been proposed to address this issue [1]. This article refers to watermarking for detection of illicit camcording of video content. Therefore, the watermarks should be robust against camcording, i.e. camcording attacks. This work is continuation of our previous research on this topic described in [2].

Generally watermark can be embedded either into uncompressed or compressed video. There are many approaches to embed watermark into uncompressed (raw) video e.g. [3,4]. Unfortunately, full decoding and re-encoding will be required to apply them into compressed video bitstream (see Fig. 1). In many applications it is not desired or even not possible to perform full decoding and again encoding, due to high computational complexity of such an approach. Consequently, watermarking of compressed video bitstreams has gained lot of attention e.g. [5-12]. Among compressed domain watermarking schemes, there are only several studies dealing with camcorder attacks e.g. [6-8] mostly for older video compression standards such as MPEG-2. Moreover, in the vast majority of papers all experiments are conducted on low resolution video materials and achieved results cannot be directly referred to HDTV video materials.

In this paper, we propose a fast watermarking method for MPEG-4 AVC/H.264 [13,14] bitstreams robust to very strong attacks, e.g. video camcorder recording. It was developed based on scaling matrices watermarking scheme described in our previous paper [2]. The computational complexity of scaling matrices watermarking was very low (see Fig 1.), but there was no possibility for local watermark adjustment. The proposed method is more complex because it requires partial decoding of bitstream to obtain prediction error transform coefficients (which are changed in the proposed watermark method). However, the strength of the watermark can be locally adjusted based on the values of residual transform coefficients. Additionally, this method provides an original approach to selection of the embedded coefficients. In contrary to scaling matrices watermarking, this method corresponds to transform domain watermarking (black dotted rectangle in Fig. 1).

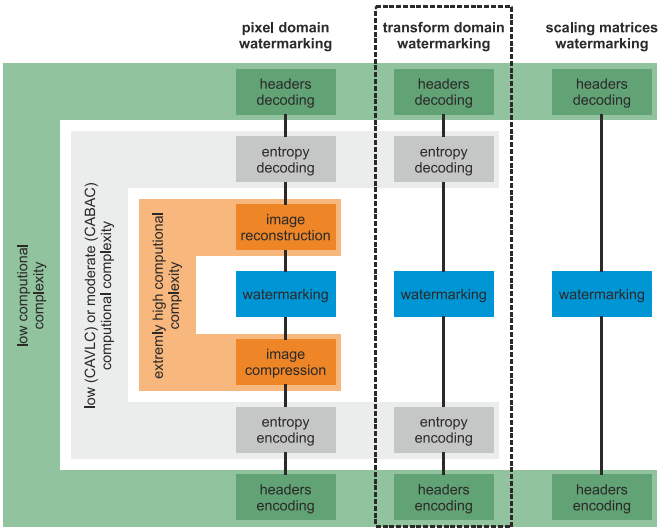


Figure 1. Three most common approaches to watermarking of compressed video sequences (pixel domain, transform domain and scaling matrices domain). For every approach the reqired operations and coresponding complexity on embedder side are shown.

II. WATERMARK EMBEDDING

In the proposed method a watermark is embedded by direct modifications of the value of low frequency DCT coefficients of transformed residual signal. The increase of energy of selected residual DCT coefficient corresponds to increased

energy of corresponding frequency in reconstructed picture. In proposed approach, energy of vertical or horizontal coefficients is changed according to bit value of embedded data. In order to embed invisible watermarks with sufficient data capacity enough to identify time and place of camcording watermark can be embedded into selected luma transform blocks, called Embedding Block (EB). Moreover, EBs can belong to selected frames called Embedding Frame (EF).

A. Frame selection

In order to preserve compatibility with embedding method described in [2], the watermark is embedded into P- and B-frames. Moreover modifying P- and B- frames allow to reduce the computational complexity and improve locality of the watermark in sequence of pictures. The complexity issue will be discussed later. Appropriate selection of EFs allow to control presence of watermark in selected periods of sequence. As in [2] the sequence of 8 watermarked images has been interleaved with the sequence of 8 non-watermarked images (Fig. 2.).



Figure 2. The location of watermarked and non-watermarked frames in watermarked video sequence. The frames marked in dark are watermarked (from [2]).

B. Embedding Block selection

The simplest approach is to watermark all blocks in the frame. Such a method will be very similar to scaling matrix watermarking [2]. However, direct access to transform coefficients allows for more advanced watermarking approach.

In MPEG-4 AVC/H.264 in P- and B- frame the both inter and intra prediction can be used. In case when intra prediction is not constrained to use another intra predicted block only (*constrained_intra_pred_flag* in PPS is equal to 0) the inter predicted block may be used as intra prediction source. This could cause drift and escalation of "desired error" introduced by watermarking. Moreover, the intra drift is the most disturbing kind of introduced distortion and occurs when watermarked blocks are used by intra predicted blocks.

In order to avoid the intra prediction issue the EB selection algorithm has been developed. The algorithm was designed to reduce visibility of watermarking artifacts. To limit the intra prediction issue every candidate block is investigated to check if any other block uses it as intra prediction reference. Such candidate blocks are excluded from EBs list. Therefore, embedder is able to reduce visibility of watermarking artifacts.

C. DCT coefficiens selection

As previously mentioned, the watermark is embedded by changing the energy of low frequency DCT coefficients. Increased energy of vertical coefficients corresponds to embedding information bit 1, and increased energy of horizontal coefficients corresponds to embedding information bit 0. For watermarking purposes, low-frequency AC coefficients have been chosen due to its camcorder attack robustness, as reported in the references [2]. Therefore, the so called 'embedding coefficients' (EC) have been selected. These are (0,1) and (1,0) coefficients for the 4x4 transform and (0,2), (0,3), (2,0) and (3,0) coefficients for the 8x8 transform (see Fig. 3). The selected 4x4 ECs and 8x8 ECs correspond to the same spatial frequencies.

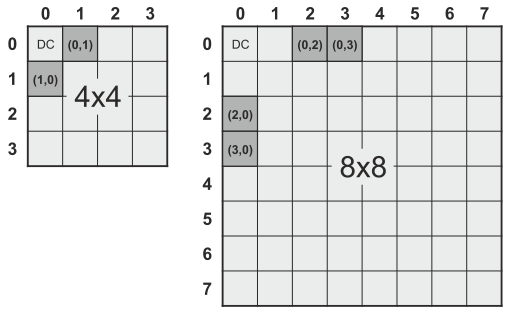


Figure 3. The modified transform coefficients, for 4x4 and 8x8 block size, respectively (from [2]).

D. Watermark strength adjustment

To utilize embedding block characteristic and its capacity for invisible watermark energy, embedding watermark strength adjustment mechanism was developed.

It is well known that highly textured blocks has higher watermark embedding potential than plain ones [5]. In proposed watermarking technology no texture reconstruction process is done and no straight block pattern classification can be done to utilize this phenomena. From the other hand we observed, that nonzero coefficients in MPEG-4 AVC/H.264 residual data are strongly correlated with boundaries of objects or strong objects patterns. Therefore, value of residual coefficients is used as control data for watermark strength adjustment mechanism.

Visibility of watermark in MPEG-4 AVC/H.264 compressed domain depends not only on watermarked block change but also on a drift that it cause. Watermark intra drift artifacts are reduced as described in subsection B but developed watermark strength adjustment mechanism takes inter drift into consideration also. In proposed approach P frames has stronger constraints for watermark embedding, because they are a source for B frames and therefore source of inter drift artifacts.

In details, developed watermark strength adjustment mechanism is summarized in Table 1.

TABLE I. EB WATERMARK STRENGTH ADJUSTMENT

	4x4 block	**8x8 block**
B-frame	If(nonzero EC): EC' = EC*2 Else if(nonzero C): EC' = 3.5/qstep Else: EC' = 0	If(Nonzero EC): EC' = EC*2 Else: EC' = 0
P-frame	If(Nonzro EC): EC' = EC*2 Else: EC' = 0	

For blocks in P and B frames if embedding coefficient value is different from zero it is doubled. For B frames only if block has any nonzero coefficient (C) but embedding coefficient (EC)

value is zero the nonzero value is inserted into EC position. The new EC value is computed as 3.5 divided by quantization step (dividing is a consequence of quantized representation for residual signal in a bitstream). In other situation embedding coefficients remains unchanged, effectively excluding EB from being watermarked. Proposed watermark strength adjustment method impose that energy of a watermark vary for sequences with different spatial frequency characteristic.

E. Embedding process

Watermark embedding is performed by changing the energy of residual coefficients (RC). Before watermark embedding RCs have to be extracted from the bitstream. In MPEG-4 AVC / H.264 partial entropy decoding of slice is not possible and whole slice has to be entropy decoded. After entropy decoding phase the embedder selects EBs from all available luma blocks. The EBs are modified by changing the selected RCs. In proposed approach only one bit per frame is embedded. Therefore all EBs in picture (slice) are modified in the same way (by changing coefficients corresponding to horizontal or vertical frequencies). After embedding process the slice is entropy encoded.

F. Embedding computional complexity

The most complex part of embedding process is entropy decoding and re-encoding. The MPEG-4 AVC / H.264 standard defines two different entropy encoders CABAC and CAVLC [14]. CABAC is based on binary arithmetic encoder enhanced by probability modeling system. CAVLC uses switched tables with predefined variable length codes. When compared to CAVLC, CABAC provides better adaptation to video characteristics and higher compression efficiency, however it has significant higher computational complexity and lower throughput. Therefore, the entropy decoding process complexity depends on entropy encoder used during generation of watermarked bitstream. The CABAC decoding complexity is rather significant but still it is only small part of full decoding burden.

The complexity of EBs selection and ECs modification is very low and can be neglected.

The entropy encoding complexity is similar to entropy decoding. However for CABAC encoded sequences the encoding complexity can be significantly reduced by changing the entropy encoder. The proposed approach is to use CAVLC instead CABAC to encode watermarked slices. In MPEG-4 AVC / H.264 switching between entropy encoders is allowed on slice level, and conformant bitstream can contain interleaved slices encoded using either CABAC or CAVLC.

G. Bitstream overhead

The bitstream increase caused by watermarking process is very small. During embedding process in P slices only existing coefficients are modified and no new coefficients are introduced. In B slices also new coefficients may be inserted in residual block containing any nonzero coefficient. It should be kept in mind that in MPEG-4 AVC / H.264 transform coefficients coding scheme the increase of coefficient absolute value causes less bitstream expansion than introduction of new coefficients [15].

The CABAC entropy encoding has higher compression efficiency than CAVLC [15]. Therefore, slice conversion from CABAC to CAVLC entropy encoder causes noticeable bitstream increase. In proposed approach only some part of slices are marked and transcoded to CAVLC (see Fig. 2). Moreover, the embedder modifies P and B slices only which are much smaller than I slices. Therefore, the average bitstream expansion is limited.

The total bitstream expansion has been investigated experimentally. For case when CABAC encoded slices are converted to CAVLC during watermarking process the obtained average bitstream expansion was about 3.8%.

III. WATERMARK DETECTION

In order to maintain compatibility between scaling matrices watermarking described in [2], the watermark detection process is performed in the same way as for scaling matrices watermarking method. The detector uses both original and investigated video sequence (in pixel domain) to detect the watermark. Both sequences must be temporally aligned but spatial alignment is not critical. The detector calculates the spatial frequency characteristics for all pictures in original and investigated sequence. The differences in frequency characteristic caused by watermarking process are calculated.

In the watermarking process only some pictures are watermarked, therefore watermarked sequence contains interleaved groups of watermarked and original pictures. The presence of original pictures allows to eliminate attacks influence (different changes in frequency characteristics) in detection process. More details about watermark detection algorithm can be find in [2].

IV. RESULTS

The watermarking and the detection techniques were implemented and tested for various attacks robustness and visibility. Four real-world HDTV sequences (Blu-ray quality) were selected for tests: *Soap opera 1*, *Soap opera 2*, *Historical novel*, *Computer-animated film*. Proposed sequence set represents material with quite different spatial frequency characteristic. This is important because of watermark strength adjustment method as described in section II D. First two sequences can be described as first-plan-in-focus, historical novel is all-in-focus sequence, and the last one is animated movie with low noise and weak textures.

All sequences used in the tests were FullHD resolution (1920x1080), 25 frames per second and 10 minutes long. Sequences were encoded by MPEG-4 AVC/H.264 encoder conformant to High profile (bitrate - 15Mbit/s). The prepared bitstreams have been watermarked. Two different random bit streams (A and B in Table 2) have been used as the watermark payload. Watermarked sequences have been displayed on an LCD HDTV set.

Watermark visibility tests have been performed on group of non-expertise students. Viewers assessed overall quality of the

watermarked as well as original (not watermarked) video material. Analysis of the subjective test results shown that the distortion introduced by watermarking process was rated as unnoticeable.

In order to rate robustness of the proposed watermarking method four different attack scenarios have been tested:

- Scenario 1: HD camcorder recording,
- Scenario 2: HD camcorder recording and downscaling to SD resolution,
- Scenario 3: Camcorder recording and frame rate reduction by frame dropping (from 25 to 12,5 fps).

All scenarios start with HD camcorder recording of watermarked sequence displayed on a LCD HDTV set and further processing of captured sequence. Camcorder and LCD HDTV set were configured to obtain as good quality of recorded sequence as possible. Therefore, automatic mechanism in both devices were disabled (e.g. auto white balance, autofocus, auto exposure) and day light strength in a test room was limited. The watermarked and attacked sequence has been decoded and passed to the input of the watermark detector. The percentage of properly decoded bits from the watermark payload is shown in Table 2.

TABLE II. EXPERIMENTAL RESULTS

Sequence type	**Payload**	**Correctly detected bits [%] in scenario:**		
		1.	***2.***	***3.***
Soap opera 1	A	86.8	82.6	85.5
	B	87.4	82.6	86.1
Soap opera 2	A	82.6	75.6	82.2
	B	86.9	79.3	85.3
Historical novel	A	90.9	87.9	88.0
	B	97.2	92.9	95.4
Computer-animated film	A	79.0	73.1	77.0
	B	76.9	70.0	75.6
Average		86.0	80.5	84.4

Robustness of proposed watermarking method depends mostly on sequence spatial frequency characteristic. Computer-animated film as a one with the lowest low spatial frequency characteristic and the lowest embedded watermark energy has the lowest correctly detected bits. Historical novel with the highest energy of low spatial frequencies has the highest detection ratio. Comparing different scenarios using camcorder recording without further processing (scenario 1) and scenarios with further processing (scenario 2 and 3) one can see that differences in correctly detected bits are low with maximum of 7%. That makes proposed method robust to captured sequence post processing. Average correctly detected bits ratio above 80% and minimum of 70% shows that with proper correction codes watermark payload can be correctly decoded. Based on experimental results we may say that proposed watermarking method is robust to very strong and complex attacks.

V. CONCLUSIONS

The main advantage of the proposed method is local watermark strength adjustment performed without full video content decoding. Therefore, the watermark can be embedded with moderate computational burden and without significant bitstream overhead. Additionally, distortion introduced by the watermarking process in perceptually unnoticeable for the viewer. The experimental results show that on average over 80% of embedded data bits are correctly decoded. This number of correctly decoded bits are sufficient for the modern forward error correction (FEC) techniques to restore transmitted information. Proposed method provides sufficient robustness against strong attacks including camcorder recording.

REFERENCES

[1] G. Doërr, J.-L. Dugelay: "A guide tour of video watermarking", Signal Processing: Image Communication vol. 18, no. 4, pp.263-282, Elsevier 2003.

[2] J. Stankowski, T. Grajek, M. Domański: "Fast Watermarking of MPEG-4 AVC/H.264 Encoded HDTV Video Bitstreams", 29th Picture Coding Symposium, PCS 2012, pp. 265-268, Kraków, Poland, 2012.

[3] A. Koz and A. A. Alatan: "Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system", IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 3, pp. 326–337, Mar. 2008.

[4] F. Hartung, B. Girod: "Watermarking of Uncompressed and Compressed Video", Signal Processing, vol. 66, no. 3, pp. 283-301, Elsevier, 1998.

[5] M. Noorkami, R.M. Mersereau: "Digital Video Watermarking in P-Frames With Controlled Video Bit-Rate Increase", IEEE Transactions on Information Forensics and Security, vol. 3, no 3, pp. 441-455, 2008.

[6] M. Celik, J. Talstra, A.N. Lemma, S. Katzenbeisser: "Camcorder Capture Robust Low-Complexity Watermarking of MPEG-2 Bit-Streams", IEEE International Conference on Image Processing, ICIP 2007, pp. 489-492, San Antonio, USA (2007).

[7] D. Choi, H. Do, H. Choi, T. Kim: "A blind MPEG-2 video watermarking robust to camcorder recording", Signal Processing, vol. 90, no. 4, pp. 1327-1332, Elsevier, 2010.

[8] H. Do, D. Choi, H. Choi, T. Kim: "Digital Video Watermarking Based on Histogram and Temporal Modulation and Robust to Camcorder Recording", IEEE International Symposium on Signal Processing and Information Technology, pp. 330-335, Sarajevo, 2008.

[9] Y. He, G. Yang, N. Zhu: "A real-time dual watermarking algorithm of H.264/AVC video stream for Video-on-Demand service", International Journal of Electronics and Communications, vol. 66, no. 4, pp. 305–312, Elsevier, 2012.

[10] J. Zhang, A.T..S. Ho, G. Qiu, P. Marziliano: "Robust Video Watermarking of H.264/AVC", IEEE Transactions on Circuits and Systems—II: Express Briefs, vol. 54, no. 2, pp. 205-209, 2007.

[11] C.-V. Nguyen, D.B.H. Tay, G. Deng: "A Fast Watermarking System for H.264/AVC Video", IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2006, pp. 81-84, Singapore, 2006.

[12] P.-C. Su, M.-L. Li, I.-F. Chen: "A Content-Adaptive Digital Watermarking Scheme in H.264/AVC Compressed Videos", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP '08, pp. 849-852, Harbin, China, 2008.

[13] ISO/IEC 14496-10: Inform. Tech. – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding, 2012.

[14] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560 –576, 2003.

[15] D. Marpe, H. Schwarz, T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", IEEE Transactions on Circuits and Systems for Video Technology vol. 13, no. 7, pp. 620–636, 2003.