HEVC Encoding Assisted with Noise Reduction

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

Abstract—**Optimization of encoding process in video compression is an important research problem, especially in the case of modern, sophisticated compression technologies. In this paper, we consider HEVC, for which a novel method for selection of the encoding modes is proposed. By the encoding modes we mean e.g. coding block structure, prediction types and motion vectors. The proposed selection is done basing on noise-reduced version of the input sequence, while the information about the video itself, e.g. transform coefficients, is coded basing on the unaltered input. The proposed method involves encoding of two versions of the input sequence. Further, we show realization proving that the complexity is only negligibly higher than complexity of a single encoding. The proposal has been implemented in HEVC reference software from MPEG and tested experimentally. The results show that the proposal provides up to 1.5% bitrate reduction while preserving the same quality of a decoded video.**

Keywords—**HEVC encoding, encoder control, model selection optimization, noise reduction**

I. INTRODUCTION

IN the recent years we can observe a steady increase of video streaming in the internet. It is estimated [1] that video on demand (VoD) services like YouTube or NetFlix, attract about $\frac{1}{3}$ of users of the internet. The ability to support vast number of subscribers, like 125 millions [2], pushes forward the development of internet infrastructure, but also is fuel for research on new video compression technologies. Obviously, higher efficiency of video compression enables limiting the required hard disk space, the required bandwidth for transmission, or admitting a higher number of simultaneous users.

One of indisputable homelands of video compression is Motion Picture Experts Groups (MPEG) working on behalf of International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). For over 30 years, MPEG has been developing video and audio technologies which became many successful ISO/IEC standards, often also adapted as standards of International Telecommunication Union (ITU) [3]. The latest video coding technology is HEVC [4]. Comparing to the previous MPEG video compression technology (AVC [5]), HEVC brings about 50% reduction of bitrate required for the same video quality [3], [6]. Of course, such a gain of efficiency required thousands of hours of experiments on developing novel, sophisticated compression tools. These tools, together, provide a variety of options for an encoder, which can be used to represent a video in a compact way. This, however, comes at a cost: the more choices an

encoder has, the more difficult the decision is. For example, it can be estimated that for the largest coding unit (LCU), a HEVC encoder has a choice of about $48 \cdot 10^{18}$ different CU, TU and PU partition combinations, 35 intra prediction modes, and two reference lists with varying pictures for inter prediction [7]. Therefore, the selection of the encoding modes is even more complex than in the case of previous techniques, like AVC. Among all of the available ways, the encoder is searching for an optimal set of modes that maximize the quality of the reconstructed video and minimize the bitrate. Testing of all of the ways of encoding the whole image, often referred to as "brute-force", is practically impossible, and therefore mode selection is most commonly optimized locally at the level of one coding unit, e.g. LCU [8]–[21]. Such an approach is used in HEVC software model (HM) developed by Joint Collaborative Team on Video Coding (JCT-VC) group [22]. JCT-VC is a group of video coding experts from ITU-T Study Group 16 (VCEG) and ISO/IEC JTC 1/SC 29/WG 11 (MPEG).

The most common application of video compression is encoding natural video sequences, captured with real cameras. Such video sequences can be characterized by the presence of noise, produced in electrical circuits of the cameras. From the encoder's point of view, the presence of noise in the images influences the selection of encoding modes. For example, prediction modes and transform coefficients transmitted in the bitstream are optimized for representing a content with a noise. Therefore, the mode selection optimization is disturbed by the presence of unwanted noise components. Moreover, the encoded and reconstructed parts of an image with a noise are often used as prediction source for other parts of the picture. Such a source is highly questionable, because a noise in successive parts of an image is poorly correlated. This phenomenon is a problem both in inter and intra predictions, where parts of an image are predicted from the same image or from images from different time instant, respectively. Moreover, HEVC encoder has a very sophisticated prediction capabilities which usually allow reducing the prediction error to practically sole noise. It can be noted, that the noise is random in nature and cannot be predicted based on the image content. In practice this means that the encoder often unnecessarily loses bits for representation of noise.

A commonly used approach to solve the above-mentioned problem is to reduce noise in the sequence before encoding [23]. In such a scenario, the encoder optimizes encoding modes basing on the original input content without the noise. The advantage of such an approach is that the produced bitstream is significantly reduced. The disadvantage is that the information about the noise is not encoded at all. This is unfortunate, because noise is perceptually important for the

viewer as its absence causes a sense of unnaturalness in the reconstructed image. Also, the noise reduction technique is often imperfect and that can lead to removing important components of and image, which are not noise. A representative example of artifacts produced in this way is absence of small details like sea waves, or blurring of edges of objects.

In paper [24] we have presented a novel hybrid approach for mode selection in video encoding. Its aim was to avoid the problem of direct encoding of noise-reduced sequences, which leads to loss of information about the noise, however ensuring better selection modes unaffected by the occurrence of noise. In this paper we present an extended version of the proposal in which we consider the usage of motion vectors as a separate component in the selection of encoding modes. Moreover, we present the idea in finer details, along with some mentioning of the noise-reduction technique used for experimentation. Also, we present the extended set of experimental results.

II. THE IDEA

The main idea of the paper focuses on a method for video encoding, basing on two versions of the encoded sequence: the original one, and the noise-reduced one. The noise-reduced version of a sequence (denoised) is used for selecting the encoding modes. The original, unaffected version of a sequence (containing noise) is used for calculating the encoded content, e.g. quantized transform coefficients, and thus for production of the output bitstream.

Implementation of the method (Fig. 1) uses two modified encoders, each working on different version of a sequence being encoded, the original input one, and the denoised one.
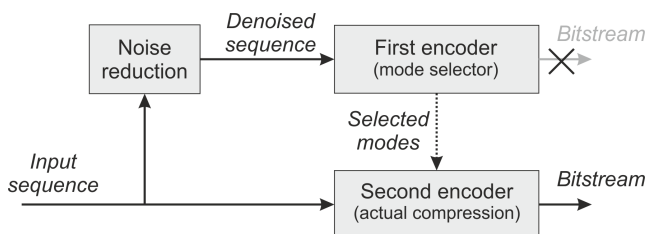

Fig. 1. The idea of proposed mode selection.

The first encoder (top) works with the denoised version of a sequence. It implements all typical encoder functionalities with the exception of entropy coding module. During the encoding, the encoder searches for the optimal coding modes, which are then given to the second encoder. As mentioned in the introduction, the optimization is performed locally at the level of LCU units. The difference between the proposed scheme and classical approach it that the encoder decisions are not influenced by the presence of noise, because it has been removed.

The second encoder (bottom) works on unaffected version of the input sequence. Implementation of this encoder is very simple, because it does not perform the expensive optimization of the encoding modes. Instead, it uses exactly the encoding modes provided by the first encoder. These encoding modes are used to generate prediction signal for each encoded block

(e.g. LCU) and to generate transform coefficients. All of the generated syntax components are then entropy-encoded and outputted in the final bitstream.

Summing up, the produced bitstream represents the original, unaffected input sequence (with noise) but represents it with the encoding modes selected basing on the noise-reduced version of a sequence.

III. ENCODER IMPLEMENTATION

We have implemented the proposed method by making modifications in the HM software model [22] (Version 13.0) for HEVC technique developed by JCT-VC group. Obviously, direct realization (Fig. 1) would lead to the use of two encoders, each of them encoding a single sequence. As shown on Fig. 2, such a redundancy is not required and can be got around. The optimization of the implementation can be done basing of two merits.

- The first encoder (which is working on the noise-reduced version of the sequence) does not generate any binary stream. Therefore, in that case entropy encoding module is not used and can be omitted. This constitutes important optimization, because it can be noticed that in modern video compression technologies, like HEVC, entropy encoding is performed with the use of very sophisticated techniques. In the case of HEVC, CABAC technique (Context-Adaptive Binary Arithmetic Coding) is used, which constitutes a substantial portion of the runtime of the encoder [25].

- The second encoder (which is working with the unaltered, original sequence) does not perform encoding mode selection. Instead, it takes encoding modes provided from the first encoder. Because the encoding mode selection, which among others includes motion vector search, is practically a second substantial part of the encoding in the terms of computational complexity, complementary to the entropy coding, omitting it also constitutes a significant optimization.

The usage of the two optimization, allows us for reducing the total complexity of the both encoders used in the proposal to complexity similar to a single, classical encoder.

IV. NOISE REDUCTION

The proposed encoding scheme is general and can be used with any noise reduction technique. For the sake of experimentation, we have decided to use an already developed motion compensation package called "mv-tools" [26], which is a plug-in for VirtualDub/AviSynth video scripting framework [27]. It allows noise reduction of video sequences in near real time. Although there are many other known noise reduction techniques, we have found out that even such a simple technique is sufficient to achieve good results.

The general idea of the algorithm in mv-tools is to reduce noise by averaging motion-compensated blocks from the sequence (Fig. 3). The details are as follows.

For each frame, block-based motion estimation is performed in order to find motion vectors pointing to frames neighbouring in time (3 previous and 3 following ones in our experiments).
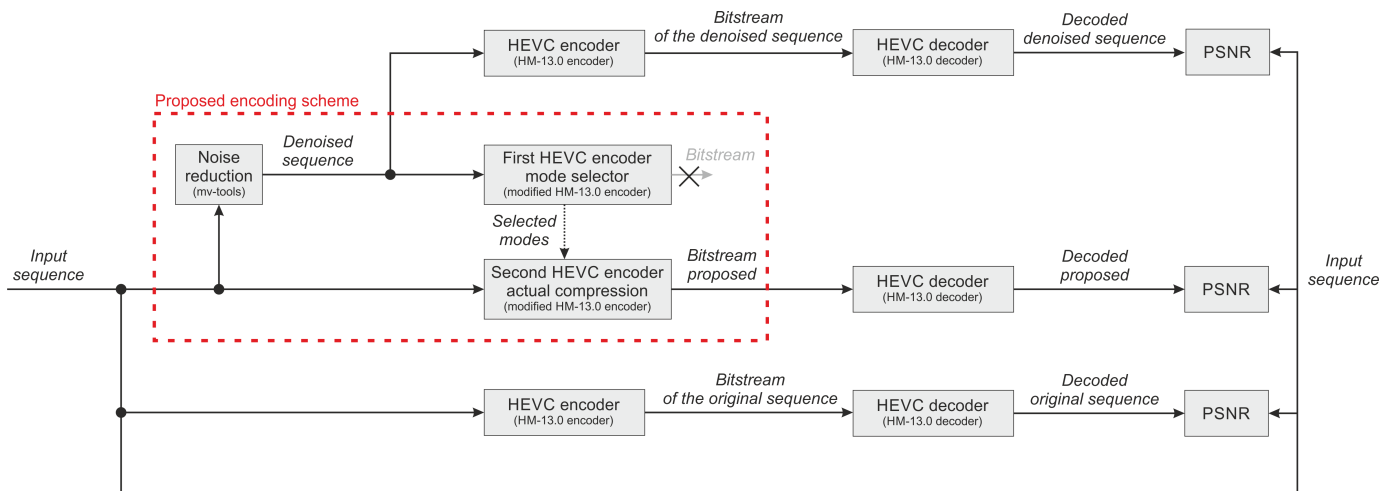
Fig. 2. The idea of proposed experiments.

There are 6 motion vectors as a result. The blocks from neighbouring frames that are pointed by these vectors are candidates for averaging. The compensated blocks are first compared with the original contents of the current frame and only if the best candidates are similar enough (basing on Sum of Squared Differences criterion) are fed to the average block. Otherwise, they are omitted. Therefore, averaging may be performed on various numbers of blocks, from 1 (only the current frame) to 7 (the current frame, 3 previous and 3 next frames).
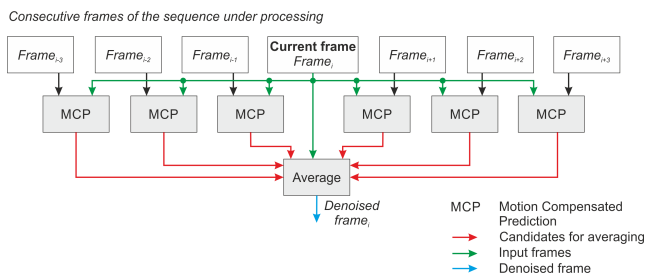

Fig. 3. The idea of the used noise reduction technique.

V. EXPERIMENTAL RESULTS

We have carried the experimental verification of the proposed technique using a set of test sequences commonly used (for the purpose) in the literature. These sequences has been developed by MPEG/JCT-VC groups during their works on HEVC technology standard [28]. The sequences are grouped (Table I) in six classes (A, B, C, D, E, F) with varying resolution (from 416×240 to 2560×1600) and frame-rate (from 20 to 60 frames per second).

In Fig. 4, 5, 6 and 7 we present exemplary pictures from the used test sequences. We also show fragments of these exemplary pictures, enlarged and with enhanced contrast, in order to allow comparison between the original version of the encoded sequence (b) and the denoised one.

All of the performed experiments followed the scheme shown in Fig. 2. Each sequence has been encoded with various

TABLE I
TEST SEQUENCES USED IN EXPERIMENTATION

Class	Number of sequences	Resolution	Frame rates
A	3	2560×1600	30; 60
B	5	1920×1080	24; 50
C	4	832×480	30; 50; 60
D	4	416×240	30; 50; 60
E	3	1280×720	60
F	4	1024×768	20; 30

quantization settings. In particular, quantization parameter index was set to QP=[22,27,32,37,42,47]. The encoding was performed in four variants: (α), (β), (γ) and (δ), as described below.

The first variant (α) is a reference for the other results. Here, a single HEVC encoder (unmodified HM software in version 13.0) encodes the input unaltered sequence.

The second variant (β) reflects the state-of-the-art known from the literature. Here, the input sequence is noise-reduced with the technique implemented in mv-tools [26] and described above. The results of noise reduction are then encoded with unmodified HM 13.0.

The third variant (γ) is in general the proposal presented in [24], but with some minor improvements. Here, two encoders are used as presented in Fig. 1. The first encoder selects encoding modes basing on noise-reduced version of the input sequence. The second encoder uses encoding modes selected by the first encoder to compress the original, unaltered input sequence. Both encoders implement the approach presented in Section III and are based on HM 13.0 software. Due to the use of optimizations mentioned in Section III, the complexity of both encoders is practically the same as the complexity of a single HM 13.0 encoder, as it is shown in Fig. 8.

The fourth variant (δ) is the same as (γ) with small difference related to motion vectors. In this variant, the motion vectors are also calculated in the second encoder, and the better one (either from the first or from the second encoder)

Fig. 4. Sample picture from the "BQTerrace" sequence (a), an enlarged and contrast-enhanced fragment (b) and its noise-reduced version (c).

Fig. 5. Sample picture from "SteamLocomotive" sequence (a), an enlarged and contrast-enhanced fragment (b) and its noise-reduced version (c).
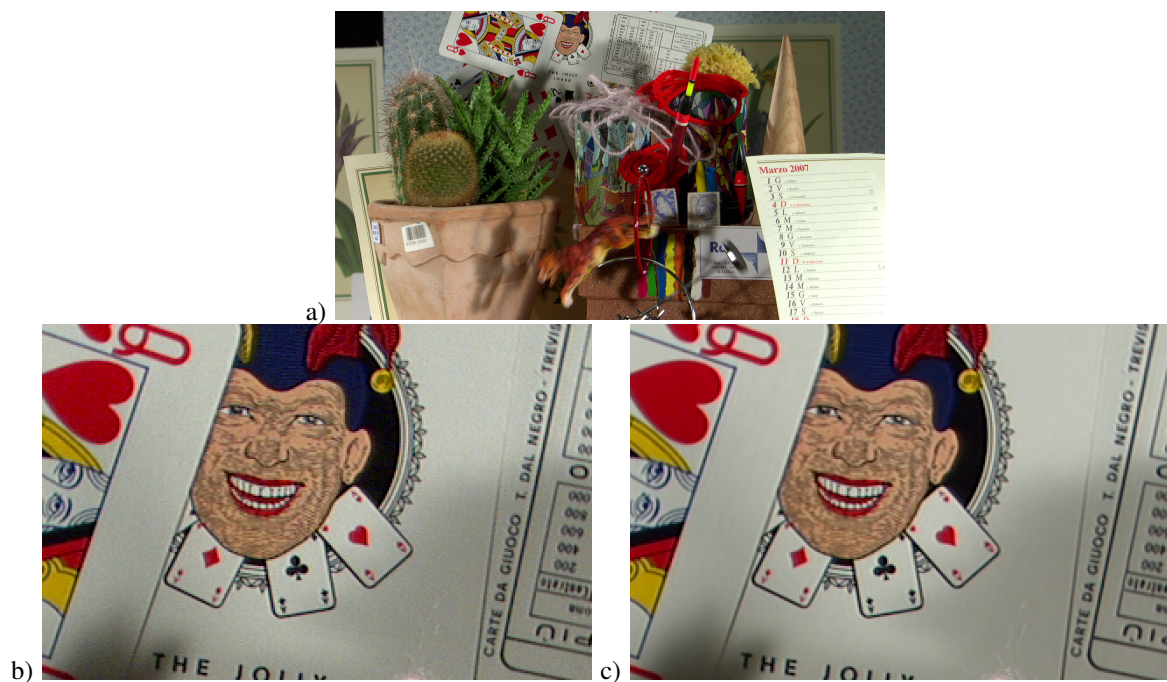
Fig. 6. Sample picture from "Cactus" sequence (a), an enlarged and contrast-enhanced fragment (b) and its noise-reduced version (c).
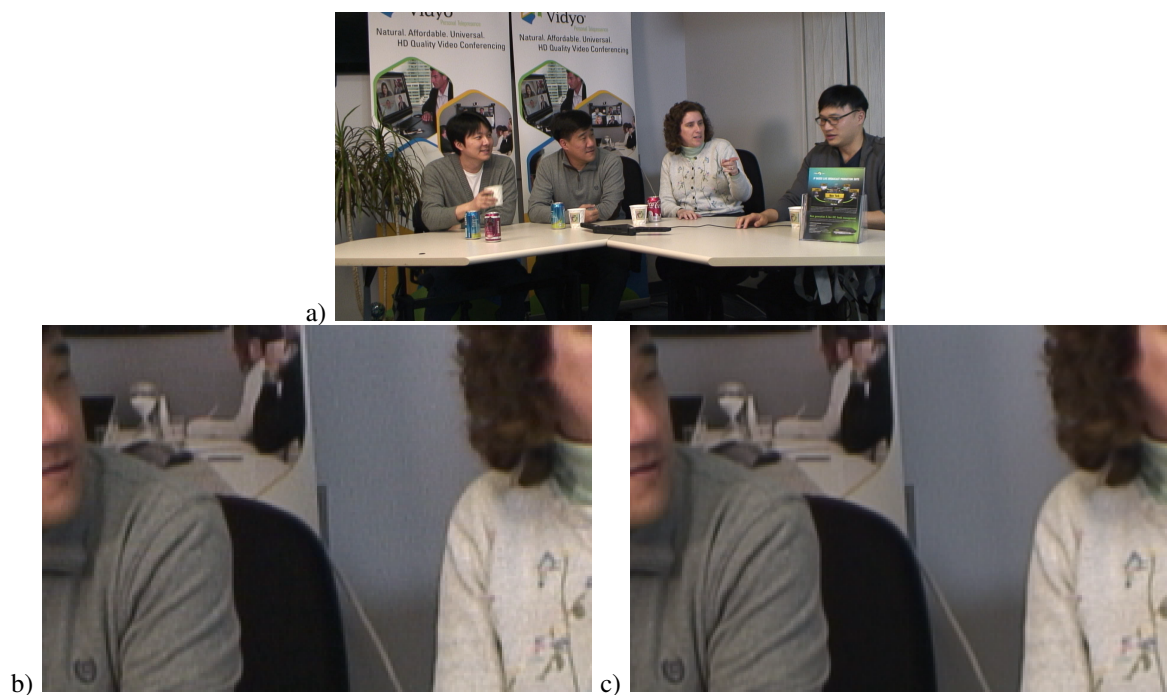

Fig. 7. Sample picture from "FourPeople" sequence (a), an enlarged and contrast-enhanced fragment (b) and its noise-reduced version (c).
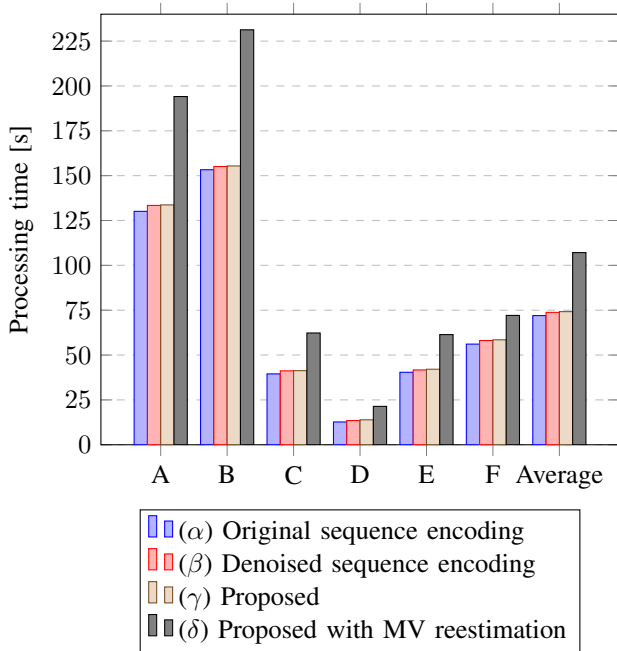
Fig. 8. Comparison of the processing time for all tested cases.

is selected. Obviously, such modification implies increase of computational complexity, because motion estimation has to be performed twice - both in the first and in the second encoder. As it can be seen in Fig. 8, the measured increase of complexity is about 50%.

After the encoding in the four mentioned variants, the produced bitstreams are then decoded and compared to the original, unaltered input sequence (with noise) with the use of objective PSNR metric (Fig. 2).

Some of the obtained results, in the form of a rate-distortion curves are illustrated in Fig. 9, 10 and 11. Fig. 9 presents results for an exemplary sequence from class B (BQTerrace). Fig 10 presents the average over all sequences in class B. Fig. 11 presents the average for all considered test sequences.

Basing on the attained R-D curves, it has been measured how big bitrate reduction can be attained with the use of the three latter variants (β), (γ), (δ) as compared to the reference (α). For this purpose we have used Bjøntegaard metric [29]. Δ(BD-RATE) was calculated over attained results for quantization parameter index range QP 27...42. The results are presented in Table II and visualized in Fig. 12 and 13. The negative Δ(BD-RATE) values indicate a reduction in bitrate, related to the reference (α).

As it can be noticed from Table II, Fig. 12 and 13, the results are consistent for nearly all of classes of the test sequences. In general, the resulting reduction of the bitrate of the proposed method is about 1.5%, indepedently whether (γ) or (δ) of the proposal is used. This provides evidence, that the method that we have described in [24] cannot be significantly improved by performing additional motion-compensation in the second encoder.

Obviously, the encoding of the noise-reduced version of the test sequence, without any noise information encoding, known from the literature, brings a higher bitrate savings up to an
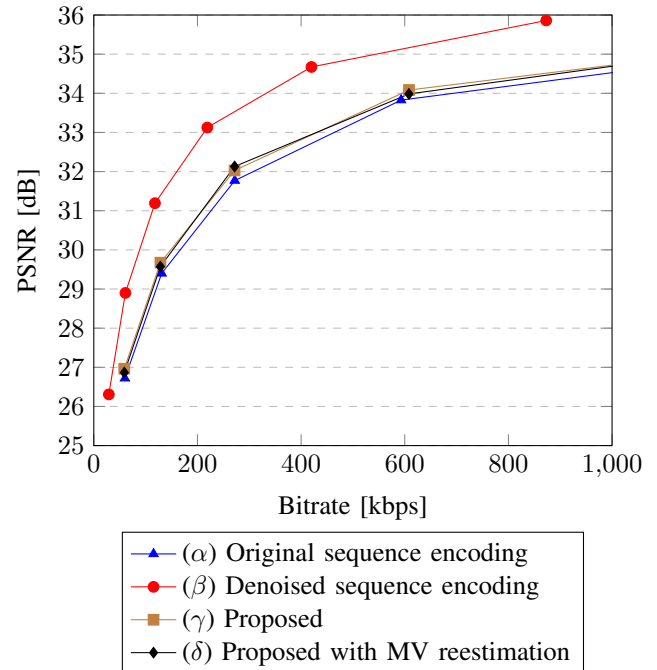

Fig. 9. Rate-distortion curves attained for comparison of three tested encoding variants: (α), (β), (γ) and (δ) for an exemplary BQTerrace sequence from class B.
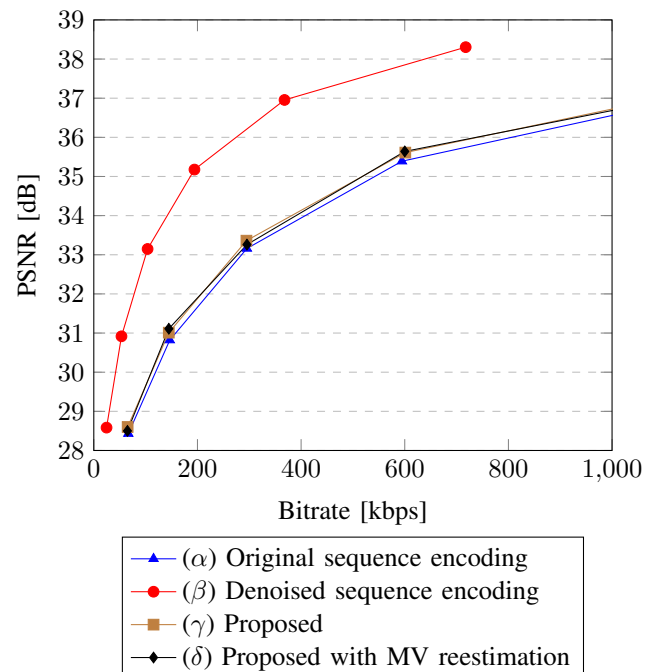
Fig. 10.   Rate-distortion curves attained for comparison of three tested encoding variants: (α), (β), (γ) and (δ) averaged over all sequences from class B.
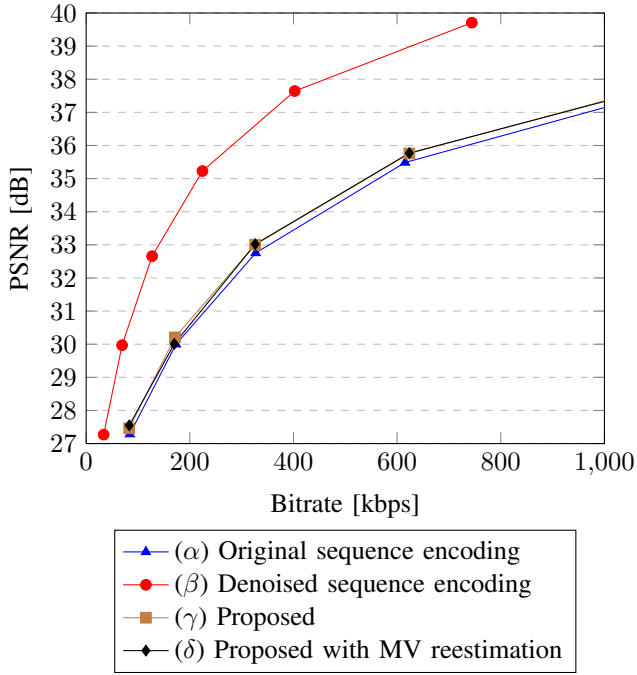
Fig. 11. Rate-distortion curves attained for comparison of three tested encoding variants: (α), (β), (γ) and (δ) averaged over all of the test sequences.
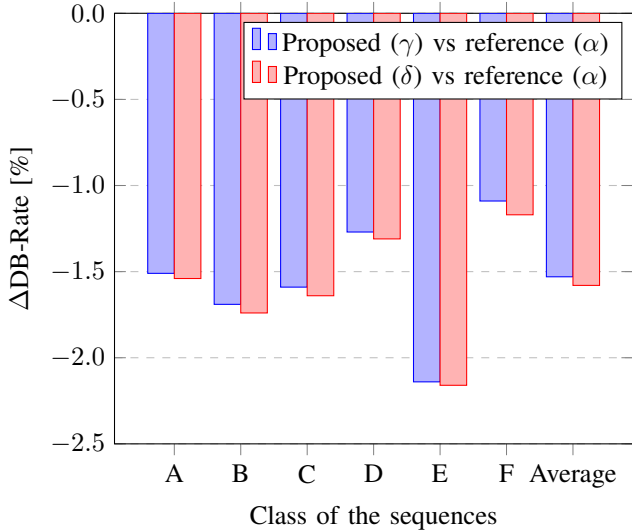

Fig. 12. Summary of Bjøntegaard deltas [29] of the proposed (γ) versus original sequence encoding (α) and the proposed with motion vectors reestimation (δ) versus original sequence encoding (α). Negative values indicate bitrate reductions.

average of about 60%. This indicates that the encoding modes selected based on the noise-reduced version of the sequences are closer to the optimal choice, but still does not allow for bitrate reduction comparable to the case in which the noise is totally omitted.

VI. CONCLUSIONS

The paper presents a method for encoding modes selection, on the example of HEVC coding technology. In the presented method, two encoders are used. The first encoder works with

TABLE II
ATTAINED BITRATE REDUCTION COMPARED TO THE REFERENCE PRESENTED AS BJØNTEGAARD DELTAS (ΔBD-RATE) [29]. NEGATIVE (ΔBD-RATE) VALUES INDICATE BITRATE REDUCTIONS

| Class	Sequences	Δ(BD-RATE)		
		Proposed (γ) vs reference (α)	Proposed (δ) with MV reestimation vs reference (α)	Denoised (β) vs reference (α)
A	Traffic	-1.77%	-1.81%	-58.93%
A	PeopleOnStreet	-1.44%	-1.46%	-64.21%
A	SteamLocomotive	-1.31%	-1.35%	-20.62%
Average class A		-1.51%	-1.54%	-47.92%
B	Kimono1	-1.01%	-1.11%	-62.80%
B	ParkScene	-1.50%	-1.54%	-67.47%
B	Cactus	-1.80%	-1.87%	-61.02%
B	BQTerrace	-2.40%	-2.38%	-49.08%
B	BasketballDrive	-1.74%	-1.81%	-78.37%
Average class B		-1.69%	-1.74%	-63.75%
C	RaceHorses	-1.44%	-1.45%	-63.09%
C	BQMall	-1.43%	-1.47%	-67.04%
C	PartyScene	-1.61%	-1.63%	-49.09%
C	BasketballDrill	-1.88%	-2.00%	-63.37%
Average class C		-1.59%	-1.64%	-60.65%
D	RaceHorsesLow	-0.30%	-0.31%	-58.30%
D	BQSquare	-1.88%	-1.91%	-56.77%
D	BlowingBubbles	-1.70%	-1.75%	-62.31%
D	BasketballPass	-1.21%	-1.28%	-72.50%
Average class D		-1.27%	-1.31%	-62.47%
E	FourPeople	-1.88%	-1.98%	-56.35%
E	Johnny	-2.50%	-2.48%	-56.08%
E	KristenAndSara	-2.03%	-2.01%	-55.95%
Average class E		-2.14%	-2.16%	-56.13%
F	BasketballDrillText	-1.80%	-1.88%	-64.51%
F	ChinaSpeed	-1.22%	-1.28%	-63.58%
F	SlideEditing	-0.35%	-0.46%	-49.53%
F	SlideShow	-0.98%	-1.07%	-71.00%
Average class F		-1.09%	-1.17%	-62.16%
Average all		-1.53%	-1.58%	-59.65%
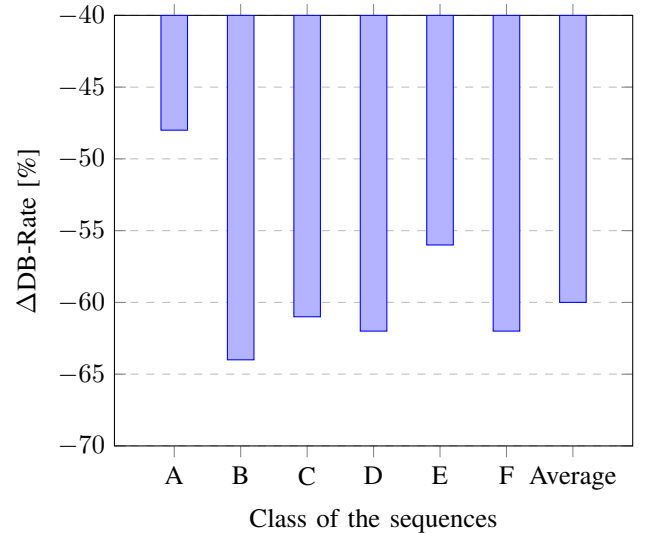
Fig. 13. Summary of Bjøntegaard deltas [29] of the denoised (β) sequence encoding versus original sequence encoding (α). Negative values indicate bitrate reductions.

noise-reduced version of the input sequence and generates encoding modes which are then used in the second encoder. Although the second encoder worked with the original, unaltered version of the input sequence (with noise), the selected encoding modes are optimized for the content without the influence of the noise. The proposal was implemented with the use of HM (version 13.0) software, which is a HEVC model software created by JCT-VC group. As an addition to our previous paper in the subject [24], we have presented extended results which confirm already attained conclusions. The results show that the proposed method enables up to 1.5% reduction in the bitstream. This indicates that the encoding modes selected based on the noise-reduced version of the sequences are closer to the optimal choice, but still does not allow for bitrate reduction comparable to the case in which the noise is totally omitted. Here, we additionally tested a variant in which motion vectors are recalculated in the second encoder and it turned out that also such a modification does not bring significant gain. This means that the gain coming from directly compressing the noise-reduced version of the sequences, without the noise, comes mainly from omitting transform coefficients related with noise components in the encoded signal and to a lesser extent, with a better choice of encoding modes.

REFERENCES

[1] https://www.omnicoreagency.com/youtube-statistics/
[2] https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/
[3] 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, vol. 525, pp. 3-15, Springer, Heidelberg 2017.
[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, 2012.
[5] 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.
[6] J.-R. Ohm, G.J. Sullivan, H. Schwarz, TK 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, 2012.
[7] F. Bossen, B. Bross, K. Suhring, D. Flynn, "HEVC Complexity Analysis and Implementation", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1685-1696, 2012.
[8] S. Cho, M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 9, pp. 15551564, 2013.
[9] W. Jiang, H. Ma, Y. Chen, "Gradient Based Fast Mode Decision Algorithm for Intra Prediction in HEVC", in Proceedings of 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 1836-1840, 2012.
[10] S. Jung, H. Park, "A Fast Mode Decision Method in HEVC using Adaptive Ordering of Modes", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 10, pp. 1846-1858, 2016.
[11] K. Lim, J. Lee, S. Kim, S. Lee, "Fast PU Skip and Split Termination Algorithm for HEVC Intra Prediction", IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 8, pp. 1335-1346, 2015.
[12] B. Min, R.C.C. Cheung, "A Fast CU Size Decision Algorithm for the HEVC Intra Encoder", IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 5, pp. 892896, 2015.
[13] D. Palomino, E. Cavichioli, A. Susin, L. Agostini, M. Shafique, J. Henkel, "Fast HEVC Intra Mode Decision Algorithm Based on New Evaluation Order in the Coding Tree Block", Picture Coding Symposium (PCS), 2013.
[14] L. Shen, Z. Zhang, Z. Liu, "Effective CU Size Decision for HEVC Intracoding", IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 42324241, October 2014.
[15] J. Siast, J. Stankowski, M. Domaski, "Hierarchical fast selection of intraframe prediction mode in HEVC", International Journal of Electronics and Telecommunications, Vol. 62, No. 2, pp. 147-151, 2016.
[16] T. L. Silva, L. A. Silva Cruz, L. V. Agostini, "HEVC Intra Mode Decision Acceleration Based on Tree Depth Levels Relationship", Picture Coding Symposium (PCS), 2013.
[17] G. Tian, S. Goto, "Content Adaptive Prediction Unit Size Decision Algorithm for HEVC Intra Coding", Picture Coding Symposium (PCS), pp. 405-408, 2012.
[18] D. Xu, Q. Lin, X. Dong, "Fast intracoding unit size decision algorithm for high-efficiency video coding", Journal of Electronic Imaging, vol. 23, no. 1, January 2014.
[19] H. Zhang, Q. Liu, Z. Ma, "Priority Classification Based Fast Intra Mode Decision for High Efficiency Video Coding", Picture Coding Symposium (PCS)2013.
[20] W. Zhao, T. Onoye, T. Song, "Hierarchical Structure based Fast Mode Decision for H.265/HEVC", IEEE Transactions on Circuit and Systems for Video Technology, vol. 25, no. 10, pp. 1651-1664, 2015.
[21] L. Zhao, L. Zhang, S. Ma, and D. Zhao, "Fast mode decision algorithm for intra prediction in HEVC", IEEE Visual Communications and Image Processing (VCIP), pp. 1-4, 2011.
[22] "HM 15.0 Reference Software". 2014. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO / IEC JTC 1 /SC29/WG11, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-15.0.
[23] B.C. Song, K.W. Chun, "Motion-compensated temporal prefiltering for noise reduction in a video encoder", 2004 International Conference on Image Processing, ICIP '04, pp. 1221-1224, 2004.
[24] O. Stankiewicz, K. Wegner, D. Karwowski, J. Stankowski, K. Klimaszewski, T. Grajek, "Encoding mode selection in HEVC with the use of noise reduction", International Conference on Systems, Signals and Image Processing (IWSSIP), Poznan, Poland, 2017.
[25] 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.
[26] M. Fizick, A. Balakhnin, T. Schniede "Mv-tools web-page" http://avisynth.org.ru/mvtools/mvtools.html - online 1 Dec 2013.
[27] R. Berg, K. Post, W. Dijkhof, E. Peche, B. Rudiak-Gould, "Avi-Synth webpage", www.avisynth.org - online 1 Dec 2013.
[28] F. Bossen. "Common test conditions and reference software configurations", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO / IEC JTC1 / SC29 / WG11, Doc. JCTVC-J1100, 2012.
[29] G. Bjøntegaar, "Calculation of average PSNR differences between RD curves", ITU-T SG16 / Q6, Doc. VCEG-M33, 2001.