4-9-2017

Information fusion based techniques for HEVC

D. G. Fernández
Universidad Complutense de Madrid

A. A. Del Barrio
Universidad Complutense de Madrid

Guillermo Botella
Universidad Complutense de Madrid

Uwe Meyer-Baese
FAMU-FSU College of Engineering

Anke Meyer-Baese
Florida State University

See next page for additional authors

Authors

D. G. Fernández, A. A. Del Barrio, Guillermo Botella, Uwe Meyer-Baese, Anke Meyer-Baese, and Christos
Grecos

Information fusion based techniques for HEVC

D. G. Fernández[a], A.A. Del Barrio[a], Guillermo Botella[a],
Uwe Meyer-Baese[b], Anke Meyer-Baese[c], Christos Grecos[d]

[a]Dept. Computer Architecture, Universidad Complutense de Madrid. 28040 Madrid (Spain)
{davidgfe, abarriog, gbotella}@ucm.es
[b]Dept. Electrical and Computer Engineering, FAMU-FSU College of Engineering, 32310.
Tallahassee, FL (USA). umeyerbaese@fsu.edu
[c]Dept. Scientific Computing, Florida State University, 32306 Tallahassee, FL (USA)
ameyerbaese@fsu.edu
[d]Dept. Computer Science, Central Washington University, 98926-7520 Washington, WA (USA)
christos.graikos@cwu.edu

ABSTRACT

Aiming at the conflict circumstances of multi-parameter H.265/HEVC encoder system, the present paper introduces the analysis of many optimizations' set in order to improve the trade-off between quality, performance and power consumption for different reliable and accurate applications. This method is based on the Pareto optimization and has been tested with different resolutions on real-time encoders.

Keywords: HEVC, H.265, pre-processing, subjective quality, Pareto optimization.

1. INTRODUCTION

High Efficiency Video Coding (HEVC) is the following step in video compression in order to achieve better compression efficiency in comparison with previous standards. The main goal of HEVC is to keep the same quality while the consumption is less than 50 per cent of the bitstream size when compared with the H264 standard. This can be done even significantly increasing the complexity of the algorithm. The bitstream size reduction means that, at the same quality, a compressed video sequence should occupy less storage space or employ less bandwidth in a transmission. Expressed in another way, for the same storage size the quality of a compressed video should be better than that corresponding to using previous standards. HEVC, aka H.265 [1], is a new coding standardization codec which came up as a project of the Joint Collaborative Team on Video Coding (JCT-VC), a collaboration between the ITU-T VCEG and ISO/IEC MPEG organizations [1]. HEVC, released in 2013, is of immense complexity and requires a great computational effort [2-3], which makes real time execution very difficult to achieve, especially for HD and UHDTV resolutions. As HEVC is tremendously complex, several approaches can be found in literature that have tried to reduce it [4-15].

There are a lot of new tools involved in the HEVC encoding efficiency. HEVC encoders [2, 16] have several input configuration parameters that allow enabling or disabling these tools. In this way, the user can set the configuration which provides the best trade-off for his interest. Basically, the user's objective is to find out the best configuration which allows obtaining the less encoding time, the minimum bitstrem size and a high level of visual quality. Less encoding time allows employing cheaper devices and reducing the power consumption because it is directly related to the computationally complexity of the algorithm. Least amount of data allows the reduction of costs in applications such as live transmissions using satellite links or 3G/4G technology, the visualization of high-resolution content via ADSL connections in real time, teleconferences, etc. The quality is important to achieve a better visual experience from the user's perspective.

Standards for video compression define the syntax or format of a compressed video sequence and a decoding method, however, the design of the encoder is not standardized. The main goal of decoders is to obtain video from a standard bitstream, for this reason decoders do not provide enough quality or compression efficiency. These characteristics, compression efficiency and video quality, directly depends on the encoder side. The standard provides a set of encoding tools but how to use them in an efficient way is a design task that becomes critical. This means that not all video encoders are implemented in the same way.

Sometimes the encoder implementation does not allow us to meet the real time constraints so it is strictly necessary to optimize the code to improve the performance. Apart from the code optimization the best improvements are achieved reducing the complexity of different encoding tools and techniques. When a tool is simplified a bistream size increment and a loss of quality occur, so the designer has to analyze what is the best trade-off which can be a hard task when managing a lot of parameters and optimizations.

The main goal of this article is to provide an analysis to easily find out the optimal encoder configurations to meet the constraints imposed by a user or developer.

The rest of the article is organized as follows. Section 2 is an introduction to HEVC standard, Section 3 focuses on the state of the art and Section 4 describes the metrics and framework used in Section 5 in which the proposed analysis is explained. Our conclusions are shown in Section 6.

2. HEVC OVERVIEW

HEVC standard is based on the well-known block-based hybrid coding architecture, combining both transform coding and motion-compensated prediction with entropy coding. As can be seen in Figure 1, each video frame is divided into blocks of pixels so the processing of the video frame can be done at block level. Residual data are obtained by subtracting the input block with the result of the corresponding predicted block. Spatial redundancies that exist within the video frame are exploited in the intra prediction. Temporal dependencies that exist between blocks in successive frames are used in inter prediction so that only the changes between successive frames need to be encoded. These dependencies are obtained by using motion estimation in which a search is performed in the previously encoded frames (reconstructed images in Figure 1) to determine the motion. The remaining spatial redundancies that exist within the video frames are exploited by coding residual blocks through transform, quantization and entropy coding. The inverse quantization and inverse transform allow the obtaining of the reconstructed image after adding its output to the predicted block. This reconstructed frame is a faithful representation of the image that is obtained on the decoder side. The deblocking and the sample adaptive offset (SAO) filters are used to remove blocking artifacts caused by the encoding process.
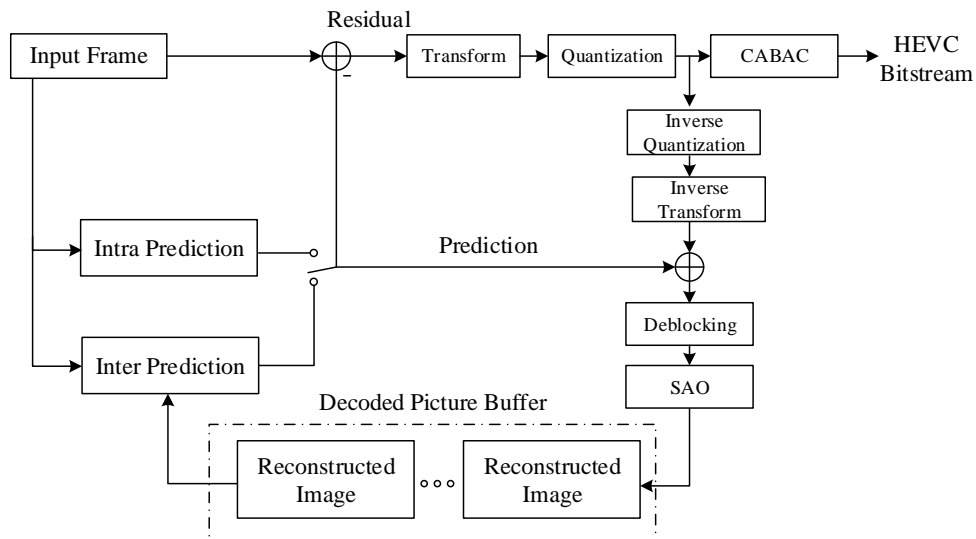

Fig. 1. HEVC coding architecture.

In contrast to previous video coding standards, HEVC uses a flexible quad-tree coding block partitioning structure that enables the use of large and multi-sized coding, prediction, and transform blocks. Compared with H264, HEVC employs an adaptive motion prediction, an enhanced intra prediction, a new loop filter (SAO) and an improved version of Context-Adaptive Binary Arithmetic Coding (CABAC) entropy coding. All of these improvements considerably increase the complexity of the encoding process. HEVC also includes new high-level structures for parallel processing.

As mentioned above, HEVC includes a new quad-tree coding block partitioning structure in which each input video frame is split into tiles and/or slices, which are divided into Coding Tree Units (CTUs). Intra slices use spatial prediction (intra prediction) within the same slice, without dependency of other slices. Inter slices, on the contrary, exploit temporal and spatial redundancy (inter prediction) using previously encoded slices. CTUs are similar to the macroblocks in previous standards. However, three different sizes are allowed, 64x64, 32x32 and 16x16 pixels for CTUs, whereas 16x16 is the only size allowed for macroblocks. In the new quad-tree partitioning structure, the CTUs are further partitioned into square regions, named Coding Units (CUs), whose sizes can be 64x64, 32x32, 16x16 and 8x8 pixels [17, 3]. The CU is configured to use a specific prediction mode (intra or inter) [2] and is represented by the leaf node of a quad-tree structure. CUs are, in turn, subdivided into one or more Prediction Units (PUs) according to a particular partition mode [17, 2]. In inter prediction mode, each Prediction Unit is predicted by using previous reference pictures in encoding time. Every PU has a set of motion parameters in order to represent the movement of the current PU as regards its reference. In order to characterize the movement, a motion estimation algorithm is applied to the reconstructed images which are stored in the decoded picture buffer (Figure 1). In intra prediction mode, every PU is predicted from neighboring pixels, exploiting spatial redundancy. The size of PUs range from 64x64 to 4x4 pixels. Finally, the last block concept existing in HEVC is the Transform Unit (TU), specified by another quad-tree structure, which is the basic unit for transform and quantization. The TUs are always square, varying the size from 32x32 to 4x4.

As Figure 1 shows, after prediction, residual data are transformed using either a Discrete Cosine Transform (DCT, Transform block in Figure 1), or a Discrete Sine Transform (only for Intra Prediction and 4x4 TU size). Finally, the quantized transform coefficients, prediction and partitioning information and other headers are encoded using CABAC, which generates the resulting bitstream.

The last concept to be introduced is rate control, an algorithm which dynamically adjusts the quantization parameter (QP) in order to achieve a specific target bitrate. To obtain this bitrate, the algorithm allocates a bits budget to each group of pictures, to an individual slice, and/or groups of CTUs within every slice in a video sequence, based on the previous output bits produced by the encoder. The number of bits generated by the encoder can be modified by varying the QP value, the task performed by this rate control algorithm. The algorithm is not a part of the video standard itself, but it is a fundamental algorithm in the encoder.

3. RELATED WORK

Nowadays, HEVC is the cutting edge encoding standard, being the most efficient solution for transmission of Ultra-high-definition television (UHDTV) video content since the consumption of the bistream size is less than 50% in comparison with the previous H264 standard [18] and allows better use of video distribution channels. HEVC have a large complexity [2-3] because of newly introduced coding tools, so several approaches dealing with it can be found in the literature. In contrast to previous video coding standards, HEVC makes use of a flexible quad-tree coding block partitioning structure which enables the use of large and multiple sizes of coding, prediction, and transform blocks. This system is more efficient but also more complex, providing the encoder the capacity to select the proper partition sizes [1][2]. In our previous works, several optimized algorithms [4-7] were proposed to accelerate the encoding process. In [7] an enhanced CU size decision algorithm based on temporally and spatially homogeneous regions detection is proposed. Basically, within the CTU loop, if the image block is classified as spatially or temporally homogeneous, the quad-tree partitioning recursive process is finished since a block belonging to any of these categories should not to be split into smaller partitions. Obviously, temporal homogeneity is only evaluated in inter CUs, while spatial homogeneity is valid for both intra and inter CUs. In this article the optimizations proposed in [7] are used as an example of how to compare an algorithm optimization when other state of the art methods.

4. FRAMEWORK

In order to measure the quality of a flow, different metrics have been typically applied in the video compression literature to characterize novel methods, as can be seen in [19-22]. Some of them are the execution time of the algorithm, the bitstream size and the Peak Signal-to-Noise Ratio (PSNR). As the aim is to compare with a reference software, it is actually measured the difference with respect to this baseline case in terms of percentage, except for PSNR. Equations 1, 2 and 3 model these measurements. Notice that a negative value means a reduction in any of the metrics.

$$\Delta Bitstream(\%) = \frac{Proposal(bytes) - Reference(bytes)}{Reference(bytes)} \times 100 \qquad (1)$$

$$\Delta Time(\%) = \frac{ProposalTime - ReferenceTime}{ReferenceTime} \times 100 \qquad (2)$$

$$\Delta PSNR(dB) = ProposalPSNR - ReferencePSNR \qquad (3)$$

When comparing different video encoders, one of the most important aspects is to stablish a set of common test conditions that allows conducting a fair comparison. It is essential to use an extensive test bench that employs sequences of several resolutions which include different kind of movements and textures. For this purpose, JCT-VT recommends a set of video sequences and common test conditions for HEVC in [23]. This document is very useful for performance comparisons with respect to other works but it cannot be used to characterize properly an algorithm since it only uses four QP values. However, it is important to develop an encoding algorithm which is able to deal with all the possible QP values imposed by the Rate Control [2]. For this reason, we recommend a test bench covering the entire range of quantization values (0-51). Moreover, in order to add to the JCT-VT test bench more HD or 4K sequences we propose the use of repositories such as [24], [25] and [26]. Therefore, the JCT-VT recommendation is useful for comparison with the state of the art methods but it is necessary to complement it with new configurations and sequences for a detailed characterization of our algorithms.

5. ANALYSIS

As has been previously commented, the objective of this study is to find out the best configuration which allows obtaining the best execution time reduction by means of negligible quality loss and a low bitstream size increment. For this purpose, the Pareto analysis [27] has been applied to study those solutions providing the best trade-offs. The best methods can then be identified as those corresponding to the points that belong to the Pareto Front (i.e., those points that are not dominated).

The analysis can be represented as a multivariable problem, using a chart to represent the ΔBitstream /ΔTime space and another chart to represent the ΔPSNR/ΔTime space. Solving the problem in this way, the optimal configurations can be identified as those that belong to the Pareto Front in both spaces. Recently, the Bjontegaard Delta Rate (BD-Rate) [28] metric has been used for HEVC performance comparisons. This metric allows the comparison and analysis of several algorithms using only one chart instead of two since BD-Rate takes into account the bitstream size (bit rate) and the quality (PSNR) in a unique metric [28]. Therefore, BD-Rate helps to considerably simplify the Pareto Analysis.

Fig. 2 shows an example of comparison of state of the art methods using Pareto Analysis in terms of BD-Rate/Encoding Time Reduction. All of the referenced methods were implemented on the HEVC test model HMv16.2 [2]. HM software allows the assessment of the new technologies included in the HEVC standard [2] but it is a non-optimized software since it does not have support for exploiting Instruction Level Parallelism (ILP) in Very Long Instruction Word (VLIW) architectures. HM also does not support multithreading or any kind of frame parallelism, so encoding with this software requires a lot of time being impossible to achieve real-time execution. Despite this fact, HM is widely used in the literature since this software implements all HEVC tools and provides the upper level of quality. Moreover, HM test model is the software used in the JCT-VT recommendation [23] for comparison purposes.
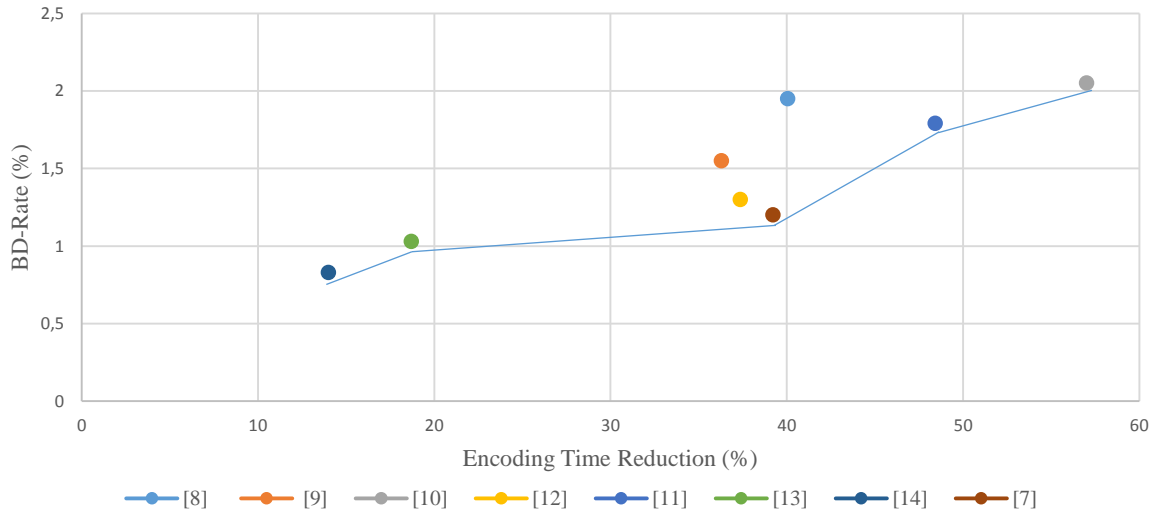
Fig. 2. Pareto analysis for state of the art methods.

In Figure 2 [14], [13], [7], [11] and [10] belong to the Pareto Front so they are the optimal methods. When speed is prioritized over the BD-Rate, [10] is the best method. On the contrary, when BD-Rate is considered as the priority, [14] could be the best solution. The work presented in [7] could be considered as an intermediate solution in which BD-Rate and speed are weighted in an equivalent way. Furthermore, as can be observed in Figure 2, [7] provides the highest quality when the encoding time saving is greater than 35%.

As commented above, BD-Rate is based on PSNR metric. PSNR does not take into account the Human Visual System (HVS) whereas it could be necessary to also consider the human perception when trying to accelerate the encoding flow. There are a lot metrics based on HVS such as SSIM [29], MS-SSIM [30] and VQM [31] so we might want to conduct an analysis using them. If this is the case, the values obtained from the proposed video quality metrics and the related encoding speed should be shown in different charts, one chart for every quality metric. Since the value of the quality depends on the QP value, it is necessary to use a different chart for every QP value. The quality metrics cannot be combined because there is no relation between them and the values obtained for each metric are very different.

If the rate control algorithm is enabled then the QP value is dynamically adjusted to achieve a target bitrate. The rate control is usually enabled in real-world applications since this is the only way of ensuring a constant data rate in a video transmission. When comparing two different encoders by enabling the rate control it is absolutely necessary to be sure that both encoders are achieving the target bitrate since the number of bits generated for every encoder has to be approximately the same to perform a fair comparison because it could happen that a quality increase was associated to a bitstream size increment.

Power-constrained applications require to include the power consumption as an additional parameter. Another important factor that affects the outcome of a project is the price of the device which is in charge of the encoding process and the related hardware (memory speed and type, peripherals, etc.), so the system cost should also be included.

In this case, we can employ a chart for the "ΔBD-Rate/ΔTime" space and a new chart for the "Power Consumption / System Cost space". The optimal points which represent the desired configuration are those points that belong to the Pareto Front in both spaces. Figure 3 shows an example of this kind of analysis by using the same HEVC real time encoder in different devices. The selected device determinates the visual quality because the encoder includes some techniques that are executed in a different way depending on the number of available threads and the GPU model. The Pareto front is highlighted in blue color. As can be seen in Figure 3, the points 1, 7, and 8 are within the Pareto Front in both charts so they are the optimal configurations
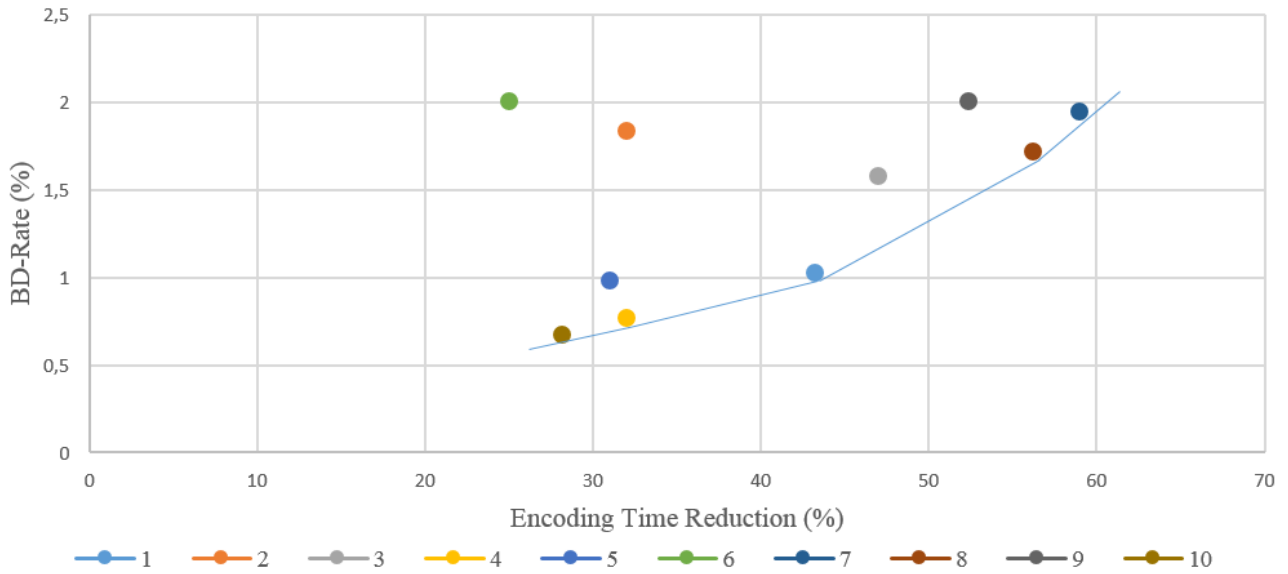
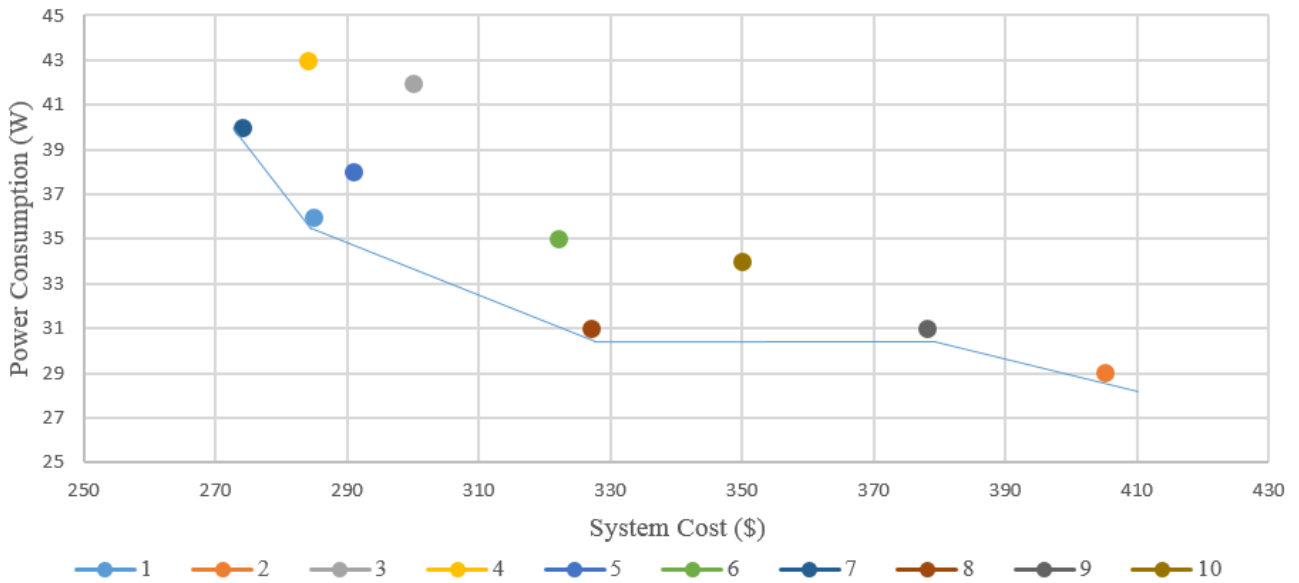Figure 3.a. Pareto Analysis for "ΔBD-Rate/ΔTime space".

Figure 3.b. Pareto Analysis for "Power Consumption / System Cost space".

6. CONCLUSION

In this paper a method to compare different video encoder implementations is explained. The Pareto analysis has been employed to obtain the optimal configurations in other kind of applications but it is not widely used in the case of video compression algorithms. The main goal of this work is to improve the reader's knowledge about how to compare different HEVC encoders to find out the optimal encoder configuration which provides the desired trade-off in terms of quality, speed, bitstream size, power consumption and system cost.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] G. Sullivan, J. 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.

[2] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "High Efficiency Video Coding (HEVC) encoder description 0v16 (HM16)," JCT-VC High Efficiency Video Coding N14 703, 2014.

[3] H. Koumaras, M. Kourtis, and D. Martakos, "Benchmarking the encoding efficiency of h.265/hevc and h.264/avc," in Future Network Mobile Summit (FutureNetw), 2012, July 2012, pp.

[4] D. G. Fernandez, A. A. Del Barrio, G. Botella and C. Garcia, "4K-based intra and inter prediction techniques for HEVC", Proc. SPIE, Real-Time Image and Video Processing 2016, vol. 98970B, April 2016.

[5] D. G. Fernández, A. A. Del Barrio, G. Botella and C. García, U. Meyer-Baese, A. Meyer-Baese, "HEVC optimizations for medical environments", in Proc. SPIE 9871, Sensing and Analysis Technologies for Biomedical and Cognitive Applications 2016, 98710B, May 19, 2016.

[6] D. G. Fernández, A. A. Del Barrio, G Botella and C. García. "Fast CU size decision based on temporal homogeneity detection", in Design of Circuits and Integrated Systems (DCIS), 2016 Conference on, 1-6.

[7] D. G. Fernández, A. A. Del Barrio, G Botella and C. García. "Fast and effective CU size decision based on spatial and temporal homogeneity detection". Multimedia Tools and Applications, 1-21.

[8] Jian Xiong, Hongliang Li, Qingbo Wu, and Fanman Meng. "A fast HEVC inter CU selection method based on pyramid motion divergence". IEEE Transactions on Multimedia, vol. 16, no. 2, Feb. 2014

[9] Liquan Shen, Zhi Liu, Xinpeng Zhang, Wenqiang Zhao, and Zhaoyang Zhang. "An effective CU size decision method for HEVC encoders". IEEE Transactions on Multimedia, vol. 15, no. 2, February 2013

[10] J. Xiong, H. Li, F. Meng, S. Zhu, Q. Wu, and B. Zeng, "MRF-based fast HEVC inter CU decision with the variance of absolute differences," IEEE Trans. Multimedia, vol. 16, no. 8, pp. 2141–2153, Dec. 2014.

[11] Jian Xiong, Hongliang Li, Fanman Meng, Qingbo Wu, King Ngi Ngan. "Fast HEVC inter CU decision based on latent SAD estimation". IEEE Transactions on Multimedia vol. 17, no. 12, pp. 2147-2159, December 2015.

[12] S. Ahn, B. Lee, and M. Kim, "A novel fast CU encoding scheme based on spatio-temporal encoding parameters for HEVC inter coding," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 3, pp. 422–435, Mar. 2015.

[13] X. Liu, Y. Liu, P. Wang, C. Lai, and H. Chao "An adaptive mode Decision algorithm based on video texture characteristics for HEVC intra prediction". IEEE Transactions on Circuits Systems for Video Technology, vol.PP, iss.99, Apr. 2016.

[14] H. Zhang and Z. Ma, "Fast Intra mode decision for high efficiency video coding (HEVC)," IEEE Transactions on Circuits Systems for Video Technology, vol.24, iss.4, pp.660-668, Apr. 2014

[15] H. Zhang, andd Z. Ma, "Fast intra mode and CU size decision for HEVC". IEEE Transactions on Circuits Systems for Video Technology, vol.PP, iss.99, Apr. 2016.

[16] x265 project, http://x265.org/, accessed January 2017.

[17] Sze, Vivienne and Budagavi, Madhukar and Sullivan Gary J., High Efficiency Video Coding (HEVC), 1st ed. Springer International Publishing, 2014. [Online]. Available: http://www.springer.com/gp/book/9783319068947f

[18] "Objective Video Quality Assessment: Using H264 and H265 Codecs for 4K Transmissions" in Scientific Bulletin of the Electrical Engineering Faculty – 2015 No.1 (29)

[19] D. Wu, F. Pan, K. Lim, S. Wu, Z. Li, X. Lin, S. Rahardja, and C. Ko, "Fast intermode decision in h.264/avc video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no.7, pp. 953–958, July 2005.

[20] Liquan Shen, Zhi Liu, Zhaoyang Zhang, and Xuli Shi. "Fast Inter Mode Decision Using Spatial Property of Motion Field", IEEE Transactions on Multimedia, vol 10, no. 6, October 2008

[21] Liquan Shen, Zhi Liu, Suxing Liu, Zhaoyang Zhang, and Ping An. "Selective Disparity Estimation and Variable Size Motion Estimation Based on Motion Homogeneity for Multi-View Coding", IEEE Transactions on Broadcasting, vol. 55, no. 4, December 2009.

[22] Liquan Shen, Zhi Liu, Tao Yan, Zhaoyang Zhang, and Ping An. "View-Adaptive Motion Estimation and Disparity Estimation for Low Complexity Multiview Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 6, June 2010.

[23] F. Bossen, 'Common test conditions and software reference configurations,' JCT-VC Document, JCTVC-K1100, October 2012.

[24] xiph.org, Derf's test media collection, https://media.xiph.org/video/derf/, 2017.

[25] Test sequences, Ultra Video group, Tampere University of Technology, http://ultravideo.cs.tut.fi/#testsequences , 2017.

[26] Elemental technologies 4k test sequences, Elemental technologies, http://www.elementaltechnologies.com/resources/4k-test-sequences, 2017.

[27] S. W. Smith, 'The Scientist and Engineer's Guide to Digital Signal Processing', (California Technical Publishing, 1997, second edition)

[28] G. Bjontegarrd, 'Calculation of average PSNR differences between RD curves,' ITU-T SC16/Q6 13th VCEG meeting, Austin, Apr. 2001.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[30] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-Scale Structural Similarity for Image Quality Assessment," in Proceedings of the 37th IEEE Asiloma Conference on Signal, Systems and Computers, Pacific Grove, CA, 2003.

[31] Feng Xiao, "DCT-based Video Quality Evaluation---Final Project for EE392J".