A NEW PERCEPTUAL ASSESSMENT METHODOLOGY FOR SELECTIVE HEVC VIDEO ENCRYPTION

Naty Sidaty, Wassim Hamidouche and Olivier Deforges

IETR Lab CNRS 6164, France

ABSTRACT

Video data security is one of the most research topic in the recent years. It is widely used in the multimedia applications such as video-conferencing, Video on Demand and Pay-TV services. Although many video encryption methods and objective measurements have been employed, few real time schemes and no subjective studies have been proposed. In this paper we investigate a set of selective video encryption schemes by encrypting only a few parameters in HEVC video streams. Firstly, we carried out an in-depth subjective study of three proposed selective HEVC video encryption schemes. A panel of observers has participated in this test campaign in order to evaluate the degree of visibility of the encrypted videos at different bitrates. Experimental results are presented and analysed, showing therefore that two proposed selected encryption schemes allow a high perceptual security level by masking the whole details of the video content, while the third scheme achieves a high security level, with a content nearly unidentifiable. In addition, subjective scores can be used as ground truth for assessing selective video encryption methods, instead of classical objective signal-based metrics, which are not correlated with human judgment.

Index Terms— selective video encryption; HEVC standard, subjective visual security assessment.

1. INTRODUCTION

Multimedia services over the Internet have been significantly increased in the last decade due to the tremendous progress in digital image and video technologies. Application service providers have for objective to guarantee that the received content will be of good quality. However, multimedia signal (audio, image, video) undergoes a digital processing, from its production to its restitution, aiming to adapt its content to the transmission channel and end devices. Consequently, ensuring a high multimedia content quality is nowadays a real challenging topic.

The latest video coding standard, named HEVC, for High Efficiency Video Coding, has been gradually adopted in many application systems, since it allows up to 50% bit-rate reduction for equal perceived quality compared with its predecessor H.264/AVC video coding [1].

Despite the coarse consumption of video content over the Internet, a huge work has been devoted to its protection and access control (security and confidentiality). Two main methods, based on security level, can be considered for video content protection: cryptographical security and perceptual security [2]. The first method consists of measurement of security against attacks that try to recovery the plaintext of the encrypted video. Encryption algorithms such as Advanced Encryption Standard (AES) and Rivest Cipher 4 (RC4) are the mostly used standards for this purpose [3]. This method allows the process of the whole video as a unique data without taking into account the structure of the compressed video [4]. Therefore, the encryption and description princesses are becoming more and more cumbersome to realised (time and energy consuming). Consequently, this method is not suitable for applications with real time requirements. The second method, based on selective encryption, has emerged as an effective alternative of the cryptographical security method. It consists of measurement of security against *perceptual* attacks that try to recover as much perceptual information as possible in the video content. The challenge therefore is to determine what data should be selected for the encryption.

Different perceptual encryption methods, of various categories, have been proposed in the literature: DCT (Discrete Cosine transform)-based encryption [5, 6], DWT (Discrete Wavelet Transform)-based encryption [7], Scalability-based encryption [8] and Motion vector scrambling [9]. The majority of these studies have strongly linked the degree of perceptual encryption with the video quality levels. Therefore, Video Quality Assessment (VQA) metrics have been widely used for assessing the visual encryption security. Hence, in [10] authors have employed three objective quality metrics in order to study the degree of visual security: image entropy, structure distortion and spatial correlation. In [11] only Correlation Coefficient (CC) has been used for encryption assessment. Authors in [8] have used, however, two other objective metrics: Peak Signal to Noise ratio (PSNR) and Encryption Quality (EQ).

In order to obtain a best accurate measurement, several encryption-oriented VQA metrics have also been proposed. Thus, few very-specific metrics including Edge Similarity Score (ESS), Luminance Similarity Score (LSS) [12], Similarity of Local Entropy (SLE) [13], Encryption Quality (EQ) [14] and so one, have been developed. The main drawback of these encryption-oriented metrics lies in the fact that they focus on a specific parameter of the video and thus disregarding the multimodal aspect. In addition, no performance against subjective encryption studies have been done. Basically speaking, there were no studies have been carried out in order to asses **subjectively** the degree of an encrypted video as can be perceived by the end user. The **objective** measures, mentioned above, treat the video sequence as a successive series of bits without taking into account the human visual system perception.

This paper aims to cover these limitations by proposing an original in-depth study involving subjective and objective measures of an encrypted videos. First, we used an in-house perceptual selective encryption method by applying a new algorithm implemented under the real time HEVC encoder for the encryption process. Then we conducted a set of subjective encryption assessment experiments through which different encryption configurations and quantification parameters have been studied. A statical analysis and objective-based comparison metrics were carried out in parallel.

In the following section, we describe the related work by briefly introducing the HEVC standard and the main proposed selective en-

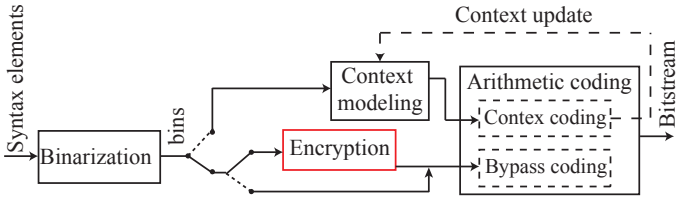

Fig. 1. Three main steps in the CABAC

cryption methods.

2. RELATED WORK

2.1. HEVC standard

The High Efficiency Video Coding (HEVC) [15] is the latest video coding standard finalized by the Joint Collaborative Team on Video Coding (JCT-VC) in January 2013. As mentioned above, HEVC enables a bitrate saving of 50% with respect to the Advanced Video Coding (AVC) [1]. This gain is enabled by the new coding tools introduced in HEVC such as the quad-tree based frame partitioning, more accurate Intra/Inter predictions, Sample Adaptive Offset (SAO) in-loop filter and the enhanced Context Adaptive Binary Arithmetic Coding (CABAC). The selective and compliant encryption is performed in HEVC at the arithmetic coding level. The CABAC engine consists of three main steps: binarization, context modeling and arithmetic coding [16]. First, the binarization step converts syntax elements to binary symbols (bin). Second, the context modeling updates the probabilities of bins, and finally the arithmetic coding compresses the bins into bits according to the estimated probabilities. Five binarization methods are used in HEVC namely Unary (U), Truncated Unary (TU), Fixed Length (FL), Truncated Rice code with an adaptive context p (TRp) and the kth-order Exp-Golomb (EGk) codes. The arithmetic coder can be performed either by an estimated probability of a syntax element (context coded) or by considering equal probability of 0.5 (bypass coded). The three main functions of the CABAC are shown in Figure 1. As illustrated in this figure, the compliant selective encryption is performed between the binarization and the arithmetic coding steps.

2.2. Selective encryption in HEVC

A number of selective encryption solutions have been proposed for HEVC [8, 17, 18, 19]. Authors in [19] studied the impact, in terms of both video quality and bitrate, of encrypting different HEVC parameters (ie. syntax elements). The encryption of a set of parameters including Transform Coefficient (TC), TC signs, Motion Vector (MV) differences, MV difference signs and delta Quantization Parameter (dQP) enables a high degradation of the video quality with a slight increase in bitrate. Shahid et al. [18] proposed a selective encryption solution for the HEVC video at constant bitrate. The proposed solution encrypts a set of HEVC syntax elements including TC, TC signs, MV differences and MV signs. The encryption is performed at the level of the CABAC binstring (i.e., after the binarization process of the CABAC). The binarization of the TC is performed in the HEVC draft 6 [1] through a combination of TRp code and EGk code with $k = 0$ (EG0). Authors in [18] proposed an algorithm to encrypt the suffix of only TC that does not impact the adaptive parameter p after encryption. Thus, this algorithm fulfills constant bitrate and format compliant encryption requirements. Moreover, the AES algorithm is used, in Cipher Feedback (CFB) mode, to encrypt the HEVC syntax elements. Therefore, the authors in [18] proposed an algorithm to transform non-dyadic Encryption Space (ES) to a dyadic ES to prepare the plaintext for AES-CFB encryption. However, in some cases, the last bit suffix can not be encrypted (non-dyadic ES). The encryption of Region of Interest (ROI) has been investigated by the authors in [17] based on the tiles repartition in HEVC [15] through both selective and naive encryption of the tiles within the ROI.

In this paper we used the selective encryption solution that we proposed in [8]. This solution encrypts in format compliant a set of HEVC syntax elements including MV differences, MV signs, TC and TC signs. A new algorithm has been proposed to encrypt all TC in format compliant while maximizing the encryption space. This solution is implemented in this paper under the real time HEVC encoder named *Kvazaar* [20]. Therefore, the modified version of the encoder, named *secure Kvazaar*, enables to encode in real time and encrypt in format compliant and constant bitrate HD and UHD HEVC videos.

The remainder of this paper is organized as follows. Section 3 introduces our subjective encryption assessment setup and illustrates the test material, environment and methodologies. Obtained results, statistical analysis and associated discussions are given in section 4. Finally, this paper ends with some conclusions and ideas about future works.

3. PERCEPTUAL ENCRYPTION EXPERIMENT

In order to evaluate, subjectively, the robustness of the used real time selective encryption method, we have performed a set of subjective encryption tests. They consist in evaluating the degree of perceptibility of visual content in the encrypted videos. Different configurations (perceptual schemes) and quality levels (different QPs) have been studied in this tests campaign. We present in this section the global environment and the implementation process to perform a such experience.

3.1. Experimental environment

The subjective evaluations were conducted in IETR laboratory psychovisual room complying with the ITU-R BT.500-13 Recommendation [21]. A display screen Full HD 40 inches Samsung UE40F6640SS was used to visualise the video sequences. Thirty three observers, 24 men and 9 women aged from 20 to 55 years, have participated in this experiment. All the subjects were screened for color blindness and visual acuity using *Ishihara* and *Snellen* charts, respectively, and have a visual acuity of 10/10 in both eyes with or without correction.

3.2. Test video sequences

Since video content can be a real attractive character, a set of video sequences, from different categories, has been selected from MPEG database[2]. The choice of these videos is mainly based on the color content, movement and texture. Five original sequences, of 10 seconds duration each and different resolutions (Classes) have been used in the experiment, as given in Table 1. First, the video sequences are processed with the real time *Kvazaar* HEVC encoder

[1]The binarization of the residual has been changed in the HEVC standard.

[2]This dataset is used by JCTVC MPEG Team in HEVC test conditions for compression

Table 1. The set of benchmark video sequences used in the experiment

Class	Resolution	Framerate	Video
B	1920x1080	50	BasketballDrive,Cactus
D	1280x720	60	FourPeople
B	1920x1080	24	Kimono
E	416x240	60	BQSquare

Table 2. Example of rating scale used in the experiment.

Degree of Content Visibility	Score
Clearly Visible	5
Visible	4
Slightly Visible	3
Barely Visible	2
Completely Invisible	1

using three QP: 22, 27 and 32 in IPPP coding configuration. The *secure Kvazaar* encoder performs both encoding and selective encryption in three encryption schemes: {TC}, {TC signs} and {TC, TC signs, MV, MV signs}. 'TC' means encrypt only the transform coefficients, 'TCs' encrypts only the TC signs and the latest configuration means encrypt the two previous parameters together with MVs and MV signs. In the rest of paper, the latest configuration will be named (All). Finally, these coding configurations, results in 45 encrypted video sequences: 5 originals $\times$ 3 QP $\times$ 3 selective encryption schemes, are then used in the experiment.

3.3. Evaluation procedure

In our subjective quality assessment, the Double Stimulus Continuous Quality Scale (DSCQS) method was used [21]. Each encrypted video was presented twice to participants accompanied by its reference version (original). Participants were asked to numerically quantify the degree of content visibility of the encrypted videos. In other words, each participant must assign a visibility score to each of the 45 test videos, according to a rating scale, given in the Table 2, ranging from '1': video content is completely invisible to '5': video content is clearly visible . At the end of each test condition, a dedicated Graphical User Interface (GUI) is displayed on the screen for about 10 seconds during which the observer gives and then confirms its judgement. At the beginning of the experiment, additional sequences were introduced in order to stabilise the opinion of the observers (these sequences will not be taken into account for the final data processing). To eliminate the memory effect, video sequences were mixed in such a way that two successive sequences must be from different categories, configuration and quality levels.

3.4. Experiment data processing

The first step in the results analysis is to calculate the average score of Mean Opinion Score (MOS) for each video used in the experience. This average is given by equation 1.

$$MOS_{jk} = \frac{1}{N}\sum_{i=1}^{N} s_{ijk} \tag{1}$$

where s_{ijk} is the score of participant i for degree of visibility j of the sequence k and N is the number of observers.
In order to better evaluate the reliability of the obtained results, it is advisable to associate for each MOS score a confidence interval, usually at 95%. This is given by equation 2. Scores respecting the experiment conditions must be contained in the interval $[MOS_{jk} - IC_{jk}, MOS_{jk} + IC_{jk}]$.

$$IC_{jk} = 1.95\frac{\delta_{jk}}{\sqrt{N}}, \quad \delta_{jk} = \sqrt{\sum_{i=1}^{N}\frac{(s_{ijk} - MOS_{jk})}{N}} \tag{2}$$

4. RESULTS AND ANALYSIS

4.1. Selective Encryption-based subjective scores

The subjective scores of all participants, collected through the dedicated GUI, have been used for the perceptual encryption measurement. Figure 2, represents the perceived degree of visibility (expressed by MOS) of different selective encryption schemes for QP=22, QP=27 and QP=32. It shows that subjects scores range generally between 3 (slightly visible) and 1 (completely invisible) point, depending on the selective encryption schemes. This implies that the human visibility is very significantly reduced when encrypt only the TC by the proposed secure *Kvazaar* encoder. In fact, '3' point in the used rating scale means that the video content is slightly visible. Otherwise, subjects can only recognize the global context (sport, film, nature) without seeing any detail of the presented video. The same results can also be noticed when encrypting only the TCs with a slight MOS variation depending on the video content and the used QP.

The behaviour changes completely by applying the third encryption scheme (All). Indeed, the subjects scores vary between 'Bearly Visible', '2' and 'Completely Invisible', '1'. The '2' point in the rating scale means that subjects can scarcely see a few things of the video (without being able to recognize the global context of the presented video). These latest results depend strongly of the video classes (resolutions). As we can see in the figure 2, video content 'BQSquare' (Classe E) is completely invisible by the whole subjects, with $MOS \simeq 1$, with very few variations depending on the used QP. In addition, 'BasketballDrive' shows high visibility scores in both TC and TCs selective encryption schemes due to its strong movement character. Curves of this video are dramatically decreased when encrypting the motion vector MV in the latest configuration (All).

4.2. Selective Encryption-based objective metrics

Figure 3 illustrates the average performance in terms of $SSIM$ (Structural SIMilarity) metric [22] of the three proposed selective encryption schemes at QP=32 (same behaviour for the other QPs). According to this figure, $SSIM$ is decreased according to the used selective encryption scheme. Using the (ALL) configuration, the $SSIM$ value is less than 0.4 depending on video content and resolution. In terms of content quality, this value indicates that the quality of video content is very degraded. Unlike results given in Figure 2, selective encryption schemes TC and TCs showed results significantly different. In fact, the curves decrease in a significant manner, thereby obtaining a gain greater than 20%, depending on video resolution. These results remain quite different from the obtained subjective findings (there are no significant differences between schemes TCs and All).

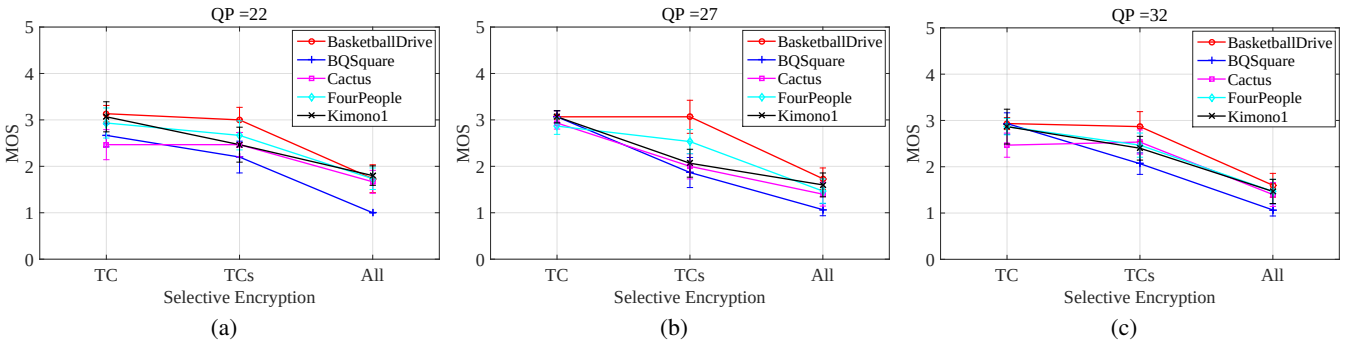

Fig. 2. Subjects visibility scores including 95% confidence intervals for (a) QP=22, (b) QP=27 and (c) QP=32.

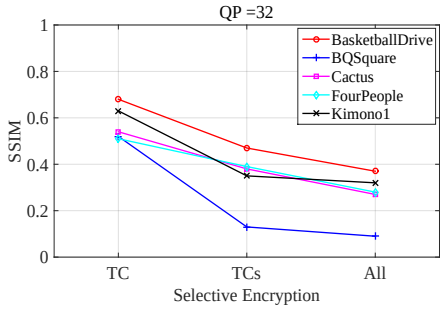

Fig. 3. Average $SSIM$ for the whole dataset, at QP=32.

Another security measure, based on $SSIM$ metric, is generally used to evaluate the encrypted content security: the Structure Distortion (SD). This value, given by $SD = 1 - SSIM$, indicates the degree of visual security [10]. In other words, the bigger the value of SD is, the more disorderly cipher bitstreams is, and consequently the video has more visual security. As an example, SD value is bigger than 0.6, when using the (ALL) selective encryption scheme.

Table 3 shows the average performance in terms of PSNR metric at QP=22 (similar results for the other QPs). In general, the PSNR values decrease, in a significant manner, whatever the used selective encryption schemes (PSNR of original videos is about $40dB$). However, results do not show a significant difference according to the employed selective encryption scheme, contrary to the subjective obtained results. The major drawback of this objective metric lies in the fact that it does not taking into account the perceptual aspect of the video by treating the whole stream as a simple text data.

The above findings have shown that $SSIM$ and $PSNR$ are not suitable to evaluate the robustness of a selective encryption scheme, since they are uncorrelated with obtained subjective assessment scores. Hence the interest, in terms of comparison, of the proposed methodology as it provides a subjective encryption-oriented ground truth.

4.3. Statistical Analysis

A statistical study was performed using the Analysis Of Variance (ANOVA) approach [23] . Indeed, ANOVA allows studying whether the variation in visibility scores is a result of the intended variation of experimental variables (i.e. QP, Resolution, Encrypted Scheme and Content), or simply due to chance. Table 4 indicates that only 'Encrypted Scheme' parameter has a significant influence on the subjects scores with *p-value* < 0.0001 [3]. Subjectively speaking, the three proposed schemes provide a good performance measurements regardless the used QP, classes and video content.

Table 3. Mean PSNR (Y) values in dB of the different video classes used in the experiment, at QP=22.

Classes	TC	TCs	All
B	12.46	13.96	10.45
D	9.49	9.07	8.77
E	13.66	8.91	9.78

Table 4. Analyseof Variance ANOVA on the whole dataset, **Df**: number of degrees-of-freedom and **F-value**: Fisher test

Source	**Df**	**F-value**	**p-value**
Class	2	1.0158	0.3708
Content	4	0.9801	0.4293
QP	2	0.1031	0.113
Encrypted Scheme	2	97.133	< 0.0001

5. CONCLUSION

In this paper we investigated a three selective encryption schemes in HEVC *Kvazaar* encoder by carrying out a set of subjective experiment studies. HEVC video sequences, from different categories and classes, have been used in this test campaign. Experiment results have demonstrated that two proposed selective encryption schemes allow a high perceptual security by decreasing the degree of visibility of the video content. The third encryption scheme achieves a high security level with a video content almost completely invisible. An objective-based comparison metrics and statistical analysis have been performed on the whole dataset showing, therefore, that $PSNR$ and $SSIM$ are not appropriated metrics for selective encryption method measurements, despite their widely used. Consequently, the MOS obtained in this experiment can be used as ground truth for assessing selective video encryption algorithms.

[3] a factor is considered influencing if *p-value* < 0.05

6. REFERENCES

[1] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparaison of the Coding Efficiency of Video Coding standards including High Efficiency Video coding (HEVC)," *IEEE TCSVT*, vol. 22, pp. 1858–1870, December 2012.

[2] R. Lukac, Ed., *Perceptual Digital Imaging, Methods and Applications*, CRC Press, Broken Sound Parkway, NY, 2012.

[3] I. Agi and Gong L., "An Empirical Study of Secure MPEG Video Transmissions," *Network and Distributed System Security*, pp. 201–210, February 1996.

[4] L. Qiao and K. Nahrstedt, "Comparison of MPEG Encryption Algorithms," *Data Security in Image Communication and Networking*, vol. 22, pp. 437–448, December 1998.

[5] L. Weng and B. Preneel, "On Encryption and Authentication of the dc dct Coefficient," in *Proceedings SIGMAP*, 2007, pp. 375–379.

[6] G. Liu, T. Ikenaga, S. Goto, and T. Baba, "A Selective Video Encryption Scheme for MPEG Sompression Standard," *IEICE Transaction on Fundamentals of Electronics, Communications and Computer*, vol. E89-A, pp. 194–202, 2006.

[7] S. Lian, J. Sun, and Z. Wang, "Perceptual Cryptography on SPIHT Compressed Images or Videos," in *Proceedings IEEE International Conference on Multimedia and Expo (ICME)*, 2004, pp. 2195–2198.

[8] W. Hamidouche, M. Farajallah, M. Raulet, O. Dforges, and S. El Assad, "Selective Video Encryption Using Chaotic System in the SHVC Extension," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 1762–1766.

[9] Y. Bodo, N. Laurent, and J.L. Dugelay, "A scrambling method based on disturbance of motion vector," in *Proceedings of the 10 ACM International Conference on Multimedia*, December 2002, pp. 89–90.

[10] Y. Yao, Z. Xu, and W. Li, "Visual security assessment for video encryption," in *Communications and Networking in China, 2008. ChinaCom 2008. Third International Conference on*, 2008, pp. 1317–1322.

[11] J. Prabhudev, P. Nagabhushan, and R.P. Kumar, "A novel perceptual image encryption scheme using geometric objects," *International Journal of Computer Science & Information Technology*, vol. 5 (4), pp. 165–173, 2013.

[12] Y. Mao and M. Wu, "Security evaluation for communication-friendly encryption of multimedia," in *in Proc. of the IEEE Int. Conf. on Image Processing (ICIP04). Singapore: IEEE Signal Processing Society*, 2004.

[13] J. Sun, Z. Xu, J. Liu, and Y. Yao, "An objective visual security assessment for cipher-images based on local entropy," *Multimedia Tools and Applications*, vol. 53, no. 1, pp. 75–95, 2011.

[14] H.H. Ahmed, H.M. Kalash, and O.S. Farag Allah, "Encryption quality analysis of rc5 block cipher algorithm for digital images," *Optical Engineering*, vol. 45, no. 10, 2006.

[15] G. J. Sullivan, Ohm J. R., 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, pp. 1649–1668, January 2012.

[16] Vivienne Sze and Madhukar Budagavi, "High Throughput CABAC Entropy Coding in HEVC," *IEEE TCSVT*, vol. 22, pp. 1778–1791, December 2012.

[17] M. Farajallah, W. Hamidouche, O. Dforges, and S. El Assad, "Roi encryption for the hevc coded video contents," in *IEEE International Conference on A Image Processing (ICIP)*, September 2015.

[18] Zafar Shahid and William Puech, "Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings," *IEEE Transactions on Multimedia*, vol. 16, pp. 24 – 36, January 2014.

[19] Glenn Van Wallendael, Andras Boho, Jan De Cock, Adrian Munteanu, and Rik Van de Walle, "Encryption for High Efficiency Video Coding with Video Adaptation Capabilities," *IEEE Transactions on Consumer Electronics*, vol. 59, pp. 634 – 642, August 2013.

[20] Marko Viitanen, Ari Koivula, Ari Lemmetti, Jarno Vanne, and Timo D. Hamalainen, "Kvazaar HEVC encoder for efficient intra coding," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*. 2015, pp. 1662–1665, IEEE.

[21] ITU-R BT.500-13 Recommendation, "Methodology for the subjective assessment of the quality of television picture," *Geneva*, 2012.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE TIP*, May 2003.

[23] G. Gamst, L. Meyers, and A. Guarino, "Analysis of variance designs: A conceptual and computational approach with spss and sas," in *Cambridge University Press, New York, USA*, 2008.