COMPRESSION EFFICIENCY OF THE EMERGING VIDEO CODING TOOLS

Naty Sidaty, Wassim Hamidouche and Olivier Deforges

IETR - INSA Rennes (France)
firstname.lastname@insa-rennes.fr

Pierrick Philippe

Orange & B<>com (France)
Pierrick.Philippe@orange.com

ABSTRACT

With the drastic increasing of multimedia applications and video coarse consumption, video compression and content quality evaluation have become an exciting and challenging topic. Recently, a new coding tool has been developed under the Joint Exploration Model (JEM) software with the main goal to provide a high bit rate saving compared to the HEVC standard. In this paper we present a performance-based comparison between the JEM and HEVC reference software (HM) through an objective measurements and a subjective quality assessments. A set of video sequences, in two spatial resolutions High Definition (HD) and Ultra-High Definition (UHD), have been used in this study. These videos are encoded using both JEM and HM software at different bitrates. Results have shown that the JEM codec enables, subjectively, a quality enhancement up to 40% at similar low bit rates. Objectively, this quality improvement is ranging from 35% to 37% depending on the spatial resolution. However, at high bit rate, the HM reference software enables a high video quality and thus its becomes more difficult to perceive the quality enhancement is about brought by the JEM codec. In addition, some video contents are difficult to encode and, consequently, the JEM enables only slight perceived quality improvement especially at the highest considered bitrates and for 4K resolutions.

***Index Terms*—** Video quality evaluation, HEVC, JEM, subjective assessment, future video coding.

1. INTRODUCTION

The latest video coding standard High Efficiency Video Coding (HEVC) [1], developed jointly by the Motion Picture Experts Group (MPEG) and the Video Coding Expert Group (VCEG) under the Joint Collaborative Team on Video Coding (JCTVC), enables up to 60% [2, 3] bit rate reduction, for the same subjective video quality, with respect to AVC standard [4]. The emerging video applications, such us Virtual Reality (VR 360) combined with High Dynamic Range (HDR), offer an immersive and more natural viewing experience. However, these new services require both higher quality and resolution (4K, 8K) to satisfy the quality of service required by the end users. Moreover, the usage of video services increases exponentially with the proliferation of mobile devices and higher usage of video applications: social networks, live video streaming, TV, network video games, E-learning, etc. To meet with these new challenges, the Joint Video Exploitation Team (JVET) has been recently established by MPEG and ITU to explore a *future video coding technology with a compression capability that significantly exceeds that of the current HEVC standard* [5]. Hence, new coding tools have been integrated within the HEVC reference software (HM) leading to an exploration video codec software platform called Joint Exploration Model (JEM). The main goal of this platform is to provide a bit rate saving between 25-30% [6, 7] compared to HEVC. This coding gain comes at the expense of additional complexity estimated to x10 at both encoder and decoder sides with respect to HEVC [1].

The main purpose of this study is to evaluate, subjectively, the quality enhancements enabled by the JEM with respect to HEVC codec in both resolutions HD and UHD (4K). In fact, these quality improvements have been studied beforehand using objective criteria such as PSNR (Peak Signal to Noise Ratio) and BD-BR (Bjøntegaard-Delta Bit-Rate) metrics [8]. These "signal-based" measurements do not take into account the Human Visual System (HVS) properties and, then, can not really reflect the quality perceived by the end users. Moreover, we would like to confirm, subjectively, the gain that JEM codec enabling 25%-30% DB-rate reduction compared to HEVC standard.
The rest of this paper is organized as follows. A brief description of the most important coding tools introduced in the JEM software is provided in Section 2. Section 3 describes the used subjective quality methodology. The objective and subjective quality assessments are presented and analysed in Section 4. Finally, Section 5 concludes this paper and gives opening for future works.

2. DESCRIPTION OF THE JEM CODING TOOLS

The JEM coding scheme can be seen as an extension of the HEVC coding structure. Especially, the JEM software is built based on the HEVC reference software (called HM) and all the JEM tools can be individually switched off, up to match the reference HEVC software HM [9] when all tools are switched off.
Most of the gains in JEM are provided by a list of approximately 15 tools which is provided in **Table 1** with their individual coding improvement. The coding gain is estimated using the Bjøntegaard (BD)-rate metric [8] measured for each tool when individually activated. The individual coding tool performance was measured in the Random Access configuration as specified in the test conditions [10], in a fashion similar to [6, 7]. As it can be seen, individual tools provide coding gains up to 5 % and the summation of the individual contributions would lead to an hypothetical gain in terms of bit rate reduction in the range of 32%.

A detailed description of all the tools is provided in document [11]. A brief description is provided in this section. The main tools can be categorized in the following categories: 1) Frame partitioning (QTBT), 2) Intra Prediction improvement (CCLM, PDPC, RSAF), 3) Inter prediction improvement (PMMVD, BIO, AMVR, OBMC, ATMVP, Affine) and 4) Transform improvement (AMT, NSST) and In-Loop filters (ALF).

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

Table 1: Main tools included in the JEM software

Acronym	Tool name	Gain
AMT	Adaptive Multiple Transforms	5%
PMMVD	Pattern matched motion vector derivation	5%
QTBT	Quad-tree plus Binary Tree	4%
ALF	Adaptive Loop Filters	4%
BIO	Bidirectional Optical Flow	2%
NSST	Secondary Transforms	2%
AMVR	Adaptive Motion Vector Resolution	2 %
OBMC	Overlap Block Motion Compensation	2%
ATMVP	Advanced Temporal Level Motion Vector Prediction	2%
CCLM	Cross-Component Linear Model prediction	1%
PDPC	Position Dependent intra-Prediction Combination	1%
AFFINE	Affine Motion Compensation	1%
RSAF	Reference Sample Adaptive Filtering	1%
Expected total BD-rate gain		32%

2.1. Frame partitioning

In the HEVC standard, the video frame is split into square blocks, using quad-trees, to adapt the local characteristics of the residual signal. This residual signal is either predicted from pixels emanating from the current image (Intra Coding) or from others (Inter Coding). The JEM has further enriched the flexibility of the partitioning through the introduction of binary trees which further divide the square blocks into rectangles. The residual signal is computed on a rectangular division and a subsequent two dimensional rectangular transforms are applied on the same area. Therefore, in JEM the HEVC Prediction Unit (PU), Coding Unit (CU) and Transform Unit (TU) paradigm disappears as the prediction area and transform area become similar. Moreover, JEM enables higher block sizes up to 256x256 instead maximum block of 64x64 in HEVC.

2.2. Intra prediction

Most of the prediction tools of JEM inherits from the HEVC standard. The basic prediction of intra prediction is kept unchanged, although the number of intra modes is increased from 35 in HEVC to 67 in JEM. Position Dependent intra Prediction Combination (PDPC) enables an additional prediction scheme, depending on the prediction mode. It combines filtered and non filtered reference samples. PDPC parameter is signaled in the bitsteam with a flag as since it represents an additional coding choice. When PDPC is not used for a block, a Reference Sample Adaptive Filtering (RSAF) can adaptively be performed to provide a smoother prediction when necessary.

2.3. Inter prediction

In this category, the tool providing the best gain is Pattern matched motion vector derivation (PMVD) which consists in an additional merge mode, enriching the HEVC merges, derived at the decoder side and therefore it does not imply additional signaling. Bi-directional Optical Flow (BIO) [12] refines the motion estimation using a pixel-wise motion compensation in the case of bi-prediction. The motion vector difference between the actual motion vector and its estimated can be adaptively coded using quarter-pel or integer-pel precision. This solutions is called Advanced Motion Vector Resolution (AMVR) where the resolution is signaled at the (CU) level. Overlapped Block Motion Compensation (OBMC) has been also introduced in order to avoid blocking artifacts during motion compensation. The Advanced Temporal Level Motion Vector Prediction (ATMVP) method allows each coding unit to fetch multiple sets of motion information from multiple blocks smaller than the current CU in the collocated reference picture. Finally, motion compensation is improved through the consideration of affine deformation which augments the traditional translational-only movements as considered in previous standards.

2.4. Transform improvements

HEVC introduced an additional transform kernel to the well known Discrete Cosine Transform (DCT) of type II: the Discrete Sine Transform (DST) type VII was determined as more efficient for small Intra blocks (4x4). The usage of different trigonometric transforms has been extended in the JEM through the Adaptive Multiple Transform (AMT) tool [13]. A set of kernels belonging to the DCT (type I, II, V and VIII) and the DST (I and VII) families are combined to provide vertical and horizontal transform sets. Those sets depend on the prediction direction in Intra mode and a particular set is provided for Inter coding. A set is composed of five transforms, i.e. five different combinations of trigonometric kernels. The selected transform is signaled to the decoder at the CU level. The kernels cover the sizes from 4x4 to 64x64 blocks to accommodate the higher partitions. As it was shown in [1], Intra prediction modes exhibit oblique patterns, non-separable transforms were introduced to more efficiently handle such residuals. Those transforms are implemented as secondary transforms (hence the name Non-Separable Secondary Transforms, NSST) [14], where the lower frequency coefficients output from the primary transform at the AMT stage are processed with an additional transform stage. This avoids the considerable complexity that would have implied complete non-separable transforms. Finally, Adaptive Loop Filter (ALF) with block based filter adaption are introduced in the JEM. Different filter shapes and coefficient sets are possible.

3. SUBJECTIVE QUALITY ASSESSMENTS

In order to evaluate the coding efficiency of the JEM compared to the HM reference software, we have performed a subjective quality experiment, involving human viewers assessing the perceived quality. Bellow, test methodology and evaluation procedure are described in detail.

3.1. Laboratory conditions

The subjective study has been conducted in the IETR laboratory psychovisual room complying with the ITU-R BT.500-13 Recommendation [15]. A 65” LED-Backlit Ultra High Definition Professional Display (x651UHD) has been used to visualize the video sequences. Seventeen observers, 10 men and 7 women aged from 18 to 44 years, have participated in this experiment. In order to ensure reliable scores, all participants have been gratified.

3.2. Test Video Sequences

In this study, a set of video sequences, of various visual content, has been used. Two categories of spatial resolutions are selected:

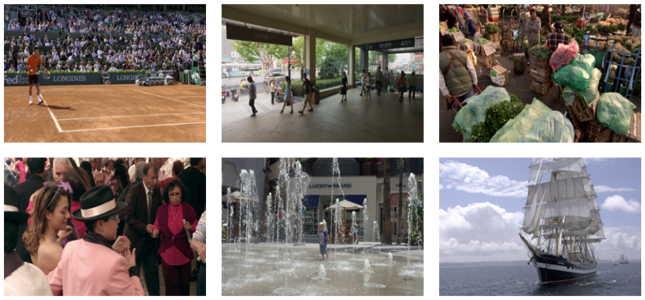


Fig. 1: Snapshots of the used video sequences.

HD (1920 × 1080) and UHD (3840 × 2160), as shown in **Table 2**. Videos are mainly taken from MPEG and 4EVER [1] databases. An example frame of the used video sequences is given in **Figure 1**. 4 bit rates for HD and 4 bit rates for UHD contents are used for creating the subjective experiment dataset. The bit rates were selected to reflect typical operating ranges for a broadcast scenario. The clips, with random access points every 1 seconds, consistent with the JEM Common Test Conditions [6], were coded in the range of 500 kbps-8 Mbps in HD and from 2 to 24 Mbps for the UHD contents. Hence, for each video content and resolution, 4 sequences are generated using HEVC reference software (HM16.7) [9] and 4 sequences using JEM3.0 [11]. Consequently, a total of 96 video sequences have been used in this study (48 for HD and 48 UHD).

Table 2: Test video sequences

Sequence	HD	UHD	Fps
Djoko	1920 x 1080	3840 x 2160	50
FoodMarket2	1920 x 1080	3840 x 2160	60
Metro	1920 x 1080	3840 x 2160	60
Tango	1920 x 1080	3840 x 2160	60
ToddlerFountain	1920 x 1080	3840 x 2160	60
Voiles	1920 x 1080	3840 x 2160	50

3.3. Subjective Test Procedure

In this subjective quality assessment, the Degradation Category Rating (DCR) method has been used, according to the ITU-T Recommendation P.910. Each compressed video was presented to participants accompanied by its reference version (original). Participants were asked to numerically quantify the quality of the compressed video. In other words, each participant must assign a quality score to each of the 96 test videos, according to a rating scale from 1 (lowest quality) to 10 (highest quality), as recommended by the ITU P.910 [16]. The experiment is divided into two parts: HD and UHD. Each part has a duration of about 22 minutes. For an optimal visual comfort, the two parts are separated by a break of 3 minutes, during which, viewers leave the psychovisual room and have a rest. Before starting the experiment, participants have received clear and deep explanations about the evaluation procedures. Moreover, a training test, with example video excerpts, was organised in order to stabilize the opinion of the observers. Finally, the observers scores are collected and analyzed in order to calculate the Mean Opinion Score (MOS).

[1] For Enhanced Video ExpeRience 2 project at www.4ever-2.com

4. RESULTS AND ANALYSIS

First of all, a well known objective metrics, as (weighted Peak Signal to Noise Ratio) wPSNR and BD-BR (Bjøntegaard-Delta Bit-Rate) [8] measurements, are used for the objective-based comparisons. **Figure 3** shows the wPSNR-based results for the whole dataset, (HD and UHD), used in the experiment. As we can notice, JEM obtains a higher wPSNR values compared to the HM reference software (HEVC) and, *a priori*, a better video quality. This behavior is consistently observed for different video sequences at different bit rates, for both test resolutions (HD and UHD).

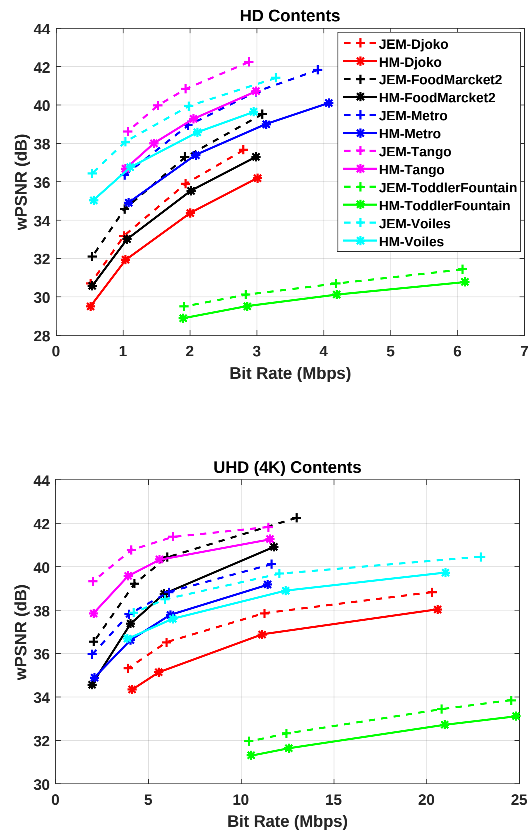
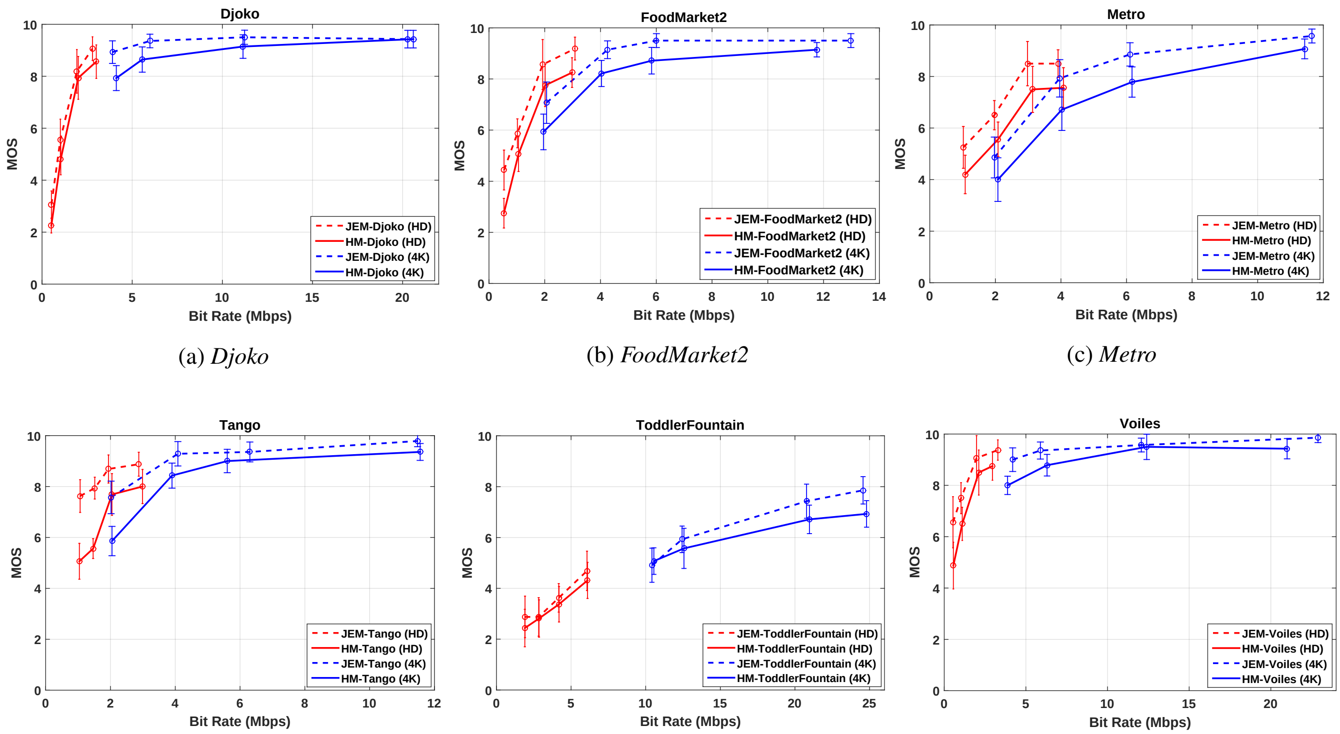


Fig. 3: wPSNR-based comparison values in: (top) HD and (bottom) UHD (4K) resolution.

Table 3 presents the Bjøntegaard measurement (BD-BR) for HD and UHD contents. On average, the JEM codec enables bit rate savings of about 35% and 37% for HD and UHD (4K) video sequences, respectively.

Subjectively, **Figures 4** show results based on the obtained Mean Opinion Score (MOS), associated with their confidence intervals for HD and UHD video content resolutions.
The JEM codec enables to improve the subjective video quality for different videos at constant bitrates compared to the HM codec (HEVC). Participants have judged the sequences encoded with JEM to have a better quality than those compressed using the HM reference software, often within low confidence intervals. These improvements depend slightly of the considered sequence. For video sequence *Djoko*, the JEM improvement is less than some other se-



(a) *Djoko* (b) *FoodMarket2* (c) *Metro*

(d) *Tango* (e) *ToddlerFountain* (f) *Voiles*

Fig. 4: MOS-based comparison, with associated 95% confidence intervals, for the used video dataset.

Table 3: BD-BR between JEM and HM video codecs

Sequence	BD-BR (HD)	BD-BR (UHD)
Djoko	-30.94%	-37.17%
FoodMarket2	-35.52%	-36.31%
Metro	-35.59%	-37.21%
Tango	-38.28%	-42.18%
ToddlerFountain	-31.34%	-28.45%
Voiles	-42.34%	-40.83%
Mean	**-35.66%**	**-37.02%**

quences, especially for *Metro* and *Tango* sequences.

These results confirm those of objective measurement where JEM codec enables a significant improvement of objective video quality at the same bit rate. However, at high cosidered bitrate, the HM codec enables a high video quality and reaches the quality of the video coded with JEM. In fact, at these bit rates the quality of the two codecs are often perceived both as close to the reference.

Nevertheless, for particular video sequence content, as *ToddlerFountain* sequence in HD resolution, neither HM nor JEM enable a good quality at the considered bit rates. In fact, this video has a noisy content (water drops), which is difficult to encoder. Despite the fact that, objectively, the JEM has demonstrated a significant quality improvement compared to the HM reference software, observers do not notice any quality improvement. It is worth noting here that we are compelled to select a high bit rates, ranging from 10 to 25 Mbps, for UHD resolution, in order to have a distinctive content quality for this particular sequence.

Finally, a statistical analysis, with Multivariate ANOVA [17], has been performed on the collected MOS using Bit rate, Content, Resolution and Codec as independent variables. Results showing that only codec parameter (JEM, HM) has a significant influence on the subjects scores, with *p-value* < 0.0001 [2].

5. CONCLUSION

In this paper, we carried out a subjective video quality assessment comparing JEM and HEVC coding performance, through its reference implementation (HM). A brief overview of the new video coding tools introduced in the JEM software is presented. Objective and subjective-based quality evaluations have been made on a set of video sequences into two spacial resolutions HD and UHD (4k). Four main observations can be made. First, the JEM codec enables a significant subjective video quality improvement for different videos at same bit rate compared to the HM reference software. Second, at high bitrates, the HM codec enables a high video quality and reaches the quality of the video coded with the JEM. Third, for a particular video sequence, as *ToddlerFountain* in HD resolutions, neither HM nor JEM enable a good quality at the considered bit rates and the JEM does not show a valuable quality improvement compared to the HM. It is therefore imperative to imagine a sequencing coding solution by which salient region (child in the image, for example) have more bit rate allocation compared to the background and other no salient regions. Finally, subjective tests have showed bit rate savings of about 40% can be achieved by the JEM codec for the same perceived video quality and seem to confirm the main goal that the JEM bit rate saving is up to 30%.

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

6. REFERENCES

[1] J-S. Gary, 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. 1649–1668, Dec. 2012.

[2] J-R. Ohm, G. Sullivan, J. Schwarz, T-K. Tan, and 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.

[3] T-K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J-R. Ohm, and J-S. Gary, “Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 76–90, Jan. 2016.

[4] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, 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, July 2003.

[5] “Jvet terms of reference,” in *http://wftp3.itu.int/av-arch/jvet-site/jvet_tor.htm*.

[6] H. Schwarz, C. Rudat, M. Siekmann, B. Bross, D. Marpe, and T. Wiegand, “Coding Efficiency / Complexity Analysis of JEM 1.0 coding tools for the Random Access Configuration,” in *Document JVET-B0044 3rd 2nd JVET Meeting: San Diego, CA, USA*, February 2016.

[7] E. Alshina, A. Alshin, K. Choi, and M. Park, “Performance of JEM 1 tools analysis,” in *Document JVET-B0044 3rd 2nd JVET Meeting: San Diego, CA, USA*, February 2016.

[8] G. Bjøntegaard, “Calculation of Average PSNR Differences Between RD-curves,” in *VCEG-M33 ITU-T Q6/16, Austin, TX, USA*, April 2001.

[9] C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, “High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description,” in *Document JCTVC-X1002 JCTVC Meeting: Geneva, CH*, May 2016.

[10] K. Suehring and X. Li, “JVET common test conditions and software reference configurations,” in *Document JVET-B0044 3rd 2nd JVET Meeting: San Diego, CA, USA*, February 2016.

[11] J. Chen, E. Alshina, G. J. Sullivan, J-R. Ohm, and J. Boyce, “Algorithm description of Joint Exploration Test Model 3 (JEM3),” in *Document JVET-C1001 3rd JVET Meeting - Geneva, CH*, May 2016.

[12] A. Alshin and E. Alshina, “Bi-directional Optical Flow for Future Video Codec,” in *2016 Data Compression Conference (DCC)*, March 2016, pp. 83–90.

[13] X. Zhao, J. Chen, M. Karczewicz, L. Zhang, X. Li, and W-J. Chien, “Enhanced Multiple Transform for Video Coding,” in *2016 Data Compression Conference (DCC)*, March 2016, pp. 73–82.

[14] X. Zhao, J-C-V. Seregin, H. Egilmez, and M. Karczewicz, “NSST: Non-Separable Secondary Transforms for Next Generation Video Coding,” in *2016 Picture Coding Symposium (PCS)*, December 2016.

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

[16] ITU-T P.910 Recommendation, “Subjective video quality assessment methods for multimedia applications,” *September*, 1999.

[17] 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.