QUALITY ASSESSMENT OF ASYMMETRIC STEREO PAIR FORMED FROM DECODED AND SYNTHESIZED VIEWS

Philippe Hanhart, Francesca De Simone, Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL), Multimedia Signal Processing Group (MMSPG), 1015 Lausanne, Switzerland

ABSTRACT

When a stereo pair is formed from a decoded view and a synthesized view, it is unclear how the overall quality of the stereo pair should be assessed through objective quality metrics. In this paper, this problem is addressed considering a 3D video represented in the format of multiview video plus depth. The performance of different PSNR-based metrics are analyzed in terms of correlation with subjective perception of video quality. A set of subjective data collected through formal subjective evaluation tests is used as benchmark. Results show that, among the considered metrics, the PSNR of the decoded view has the highest correlation with the perceived quality while the PSNR of the synthesized view has the lowest correlation.

Index Terms— 3D, quality assessment, quality metric, asymmetric stereo, view synthesis.

1. INTRODUCTION

With the rapid growth of three-dimensional (3D) video technologies, the design of objective quality assessment methods, i.e., metrics, that can reliably predict the quality of 3D content as perceived by the end user, is of crucial importance. Subjective tests are time consuming, expensive, and not always feasible. Therefore, objective measurements are needed as well, especially to assess advances in the design of coding technologies. Despite the efforts of the scientific community in recent years, 3D video quality assessment is still an open challenge and there are no metrics which are widely recognized as reliable predictors of human 3D quality perception. The assessment of 3D quality is particularly challenging for stereoscopic video when the stereo pair is formed by two views having different visual quality (mismatched or asymmetric stereo pair). In this case, specific properties of the human visual system, such as binocular suppression (the masking of low-frequency content in one view by the sharp visual content in the other view), should be taken into account when building models that objectively quantify the overall quality of the rendered stereo pair.

Many studies are available in literature dedicated to human perception of mismatched stereoscopic content. Stelmach et al. [1] have shown that for stereo pairs composed of an original view and a low pass filtered view, the perceived quality is closer to that of the original view. A similar behavior has been reported for pairs formed from an original view and a horizontally downsampled view. On the other hand, for stereo pairs composed of an original view and a quantized view, the perceived quality has been proven to be closer to the average quality of the two views. The same observation has been reported for asymmetric coding of stereoscopic video sequences using MPEG-2. Seuntiens et al. [2] have reported similar results in the case of asymmetric JPEG coding of stereo images. Saygili et al. [3] have studied the level of asymmetry needed for maximizing the perceived quality. It has been concluded that asymmetric SNR coding achieves the best perceived quality as long as the view with the lowest quality has a PSNR of at least 32dB.

Recently, the MPEG committee has initiated an effort for the definition of a 3D video coding standard based on the multiview video plus depth (MVD) format [4]. The new MPEG 3D video format will allow advanced processing of stereoscopic content to cope with varying display types and sizes, as well as different viewing scenarios [5]. Particularly, a 2-view stereo coding configuration has been included in the experimental framework defined for the development of the standard [6]. As illustrated in Figure 1, the decoded data, i.e., texture views and corresponding depth maps, is used to synthesize a virtual view at a selected position. One example of view synthesis technique is depth image based rendering (DIBR). In this configuration, the stereo pair displayed on the



Fig. 1. Advanced stereoscopic processing with 2-view configuration.

stereoscopic monitor consists of the decoded left or right view and the synthesized view. Due to the artifacts introduced by the view synthesis algorithm and the compression of the depth maps, the overall quality of the virtual view is lower than that of the decoded view, thus the stereo pair is asymmetric.

In March 2011, a Call for Proposals (CfP) for 3D coding technologies was issued by MPEG [7]. To assess the available technologies and proceed towards the definition of the standard, a formal subjective quality assessment test campaign, named 3DV test campaign, was organized. The test campaign was carried out by the European COST Action QUALINET (European Network on Quality of Experience in Multimedia Systems and Services) and included thirteen test sites [8]. The performance of twenty-two 3D video coding algorithms, submitted by proponents who answered the CfP, were evaluated on coding configurations including the 2-view scenario described above.

Peak Signal to Noise Ratio (PSNR) is commonly accepted and used by video coding experts to evaluate the performance of coding algorithms, even if its correlation with human perception of visual quality is known to be limited. Proponents, in their responses to the CfP, have reported the PSNR value of the decoded views, calculated between the original view and its decoded version. Many proponents have also provided the PSNR value of the synthesized view, calculated between the synthesized view generated at the decoder side and the synthesized view generated at the encoder side using the same view synthesis algorithm.

Bosc *et al.* [9] have shown that PSNR has a very low correlation with perceived quality when used for objective quality assessment of synthesized views. However, when the decoded view and the synthesized view in the stereo pair have different quality, it is unclear whether PSNR can be a good estimation of perceived quality. Moreover, when considering the MPEG 2-view coding configuration, it is unclear which views should be taken into account for PSNR computation: the decoded view, the synthesized view, or both views?

To answer to these questions, in this paper, the performance of the PSNR metric for objective quality assessment of asymmetric stereo pairs, formed from a decoded view and a synthesized view, is analyzed. Different PSNR-based stereo quality measures have been computed for the video sequences used in the 3DV test campaign. The subjective results of the 3DV test have been used as benchmark to evaluate the performance of the selected set of objective metrics in predicting human quality perception.

The paper is organized as follows. Section 2 provides an overview of the methodology followed in the 3DV test campaign to collect the subjective results used as benchmark in this study. Different PSNR-based metrics are defined in Section 3. In Section 4, the methodology used to evaluate the performance of the metrics is described. Results are shown and analyzed in Section 5. Conclusions and discussion on future work are presented in Section 6.

2. SUBJECTIVE QUALITY ASSESSMENT

Subjective tests have a significant impact on the validation process of an objective quality metric since their results are used as ground truth to evaluate the metric performance. In the following section, the test material, laboratory environment, rendering conditions, and evaluation methodology used by all the laboratories that performed the subjective tests are described.

2.1. Test material

The 3DV CfP defines some "classes" of test sequences, i.e., sets of spatio-temporal resolutions and target coding bit rates. Among them, Class A, with a frame size of 1920×1088 pixels and a frame rate of 25 frames per second, and Class C, with a frame size of 1024×768 pixels and a frame rate of 30 frames per second. The test material included four different contents in Class A and four different contents in Class C (referred to in the rest of the paper as S01 to S04 for Class A and S05 to S08 for Class C), along with four target coding bit rates. All video sequences were stored as raw video files, progressively scanned, with YUV 4:2:0 color sampling, and 8 bits per sample.

In the 2-view test scenario, as defined in the CfP, the stereo pair consisted of one of the two input views reconstructed at the decoder side and a synthesized view (see Table 1). Readers can refer to the 3DV CfP for more details [7].

2.2. Proponents

By responding to the CfP, twenty-two proponents submitted their codec descriptions, encoded and decoded test sequences at requested target bit rates, and synthesized views. Two anchors, i.e., Multiview Video Coding (MVC) [10], specified in Annex H of H.264/MPEG-4 Part 10, and High Efficiency Video Coding (HEVC) [11], were also included in the set of coding technologies under assessment.

2.3. Laboratory set up and data rendering

The evaluation was performed using a 46" Hyundai S465D polarized stereoscopic monitor with a native resolution of

Table 1. Input views and displayed stereo pair.

| Seq. ID | Test Sequence | Test Class | Input views | Stereo pair | |
|---------|---------------|------------------|-------------|-------------|--|
| S01 | Poznan Hall2 | | 7-6 | 6.5-6 | |
| S02 | Poznan Street | oznan Street 4-3 | | 3.5-3 | |
| S03 | Undo Dancer | ndo Dancer A 2-5 | | 3-5 | |
| S04 | GT Fly | | 5-2 | 4-2 | |
| S05 | Kendo | 3-5 | | 4-5 | |
| S06 | Balloons | C | 3-5 | 4-5 | |
| S07 | Lovebird1 | | 6-8 | 7-8 | |
| S08 | Newspaper | | 4-6 | 5-6 | |

 1920×1080 pixels. Eighteen naive viewers evaluated the quality of each test sequence. Since a maximum of three subjects could be seated in front of the monitor without deteriorating the perception of the 3D rendering, the test room set up included three subjects seating in a row perpendicular to the center of the monitor. The viewers were seated at about four times the height of the active part of the display. The laboratory setup was controlled to produce reliable and repeatable results. The test room was equipped with a controlled lighting system with a 6500K color temperature and an ambient luminance at 15% of maximum screen luminance.

To render correctly the test material on the stereoscopic monitor, a set of processing steps was performed on the video files received from the proponents. First, the two views of the stereo pair were cropped and horizontally shifted to obtain a pre-defined depth for each content (different shift parameters were set for different content). Then, the resulting sequences were interlaced (right view on top) and padded to produce a full HD resolution video.

2.4. Evaluation methodology

The Double Stimulus Impairment Scale (DSIS) evaluation methodology was selected to perform the tests. Subjects were presented with pairs of video sequences (i.e., stimuli), where the first sequence was always a reference video (stimulus A) and the second, the video to be evaluated (stimulus B). Subjects were asked to rate the quality of each stimulus B, when compared to stimulus A. After the presentation of each pair of sequences, five seconds voting time followed. The test subjects were asked to enter their quality score for the stimulus B in paper scoring sheets during these five seconds.

An 11-grade numerical categorical scale was used [12]. The rating scale ranged from 0 to 10, with 10 indicating the highest quality, i.e., the test sequence is indistinguishable from the reference, and 0 indicating the lowest quality.

All subjects taking part in the evaluations underwent a screening to examine their visual acuity, using the Snellen chart, and color vision, using the Ishihara chart. Their stereo vision was also tested using the Randot test.

Before each test session, written instructions and a short explanation by a test operator were provided to the subjects. Also, a training session was run to show the graphical user interface, the rating sheets, and examples of processed video sequences. These training videos were produced using two contents not included in the test material, with coding conditions similar to those used to produce the actual test material.

A basic test session of DSIS methodology including twenty-four test pairs, three dummy stimuli pairs, and one reference versus reference pair, was designed. Thus, the test material resulted in a total of sixteen sessions for each of the two classes of data. In each session, the stimulus pairs were presented in random orders, but never with the same video content in consecutive pairs.

2.5. Data processing

The subjective results have been processed by first detecting and removing subjects whose scores appeared to deviate strongly from the other scores in each test session. A score by subject *j* and test condition *i*, s_{ij} , was considered as outlier if $s_{ij} > q_3 + 1.5(q_3 - q_1) \lor s_{ij} < q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles of the scores distribution for test condition *i*, respectively [13]. A subject was considered as an outlier, and thus all his/her scores were removed from the results of the session, if more than 20% of his/her scores over the session were outliers [13]. Then, the mean opinion score (MOS_i) was computed for each test sequence as the mean across the rates of the valid subjects, as well as associated standard deviation (σ_i) and 95% confidence interval, assuming a Student's t-distribution of the scores.

3. OBJECTIVE QUALITY ASSESSMENT

In the 2-view configuration, as considered in the CfP, a pair of cameras is used to produce the input views at the encoder side. At the decoder side, the displayed stereo pair is formed from the decoded right view and a synthesized view, located inbetween the input views, as specified in Table 1 and depicted in Figure 1. For one Class A content and all Class C contents, the location of the synthesized view matched the location of a real view, called intermediate view, available in the original data (but not used by the encoder).

Thus, the following objective video quality models (VQR) were included in our study:

- a) PSNR of the decoded view, calculated between the decoded view and the original view: $PSNR(V'_R, V_R)$
- b) PSNR of the intermediate view, calculated between the synthesized view at the decoder side and the intermediate view from the original data (when available): PSNR (V'_S, V_I)
- c) PSNR of the synthesized view, calculated between the synthesized view at the decoder side and the synthesized view at the encoder side: $PSNR(V'_S, V_S)$
- d) Average PSNR of the decoded view and the intermediate view, computed as the mean value of a) and b)
- e) Average PSNR of the decoded view and the synthesized view, computed as the mean value of a) and c)

The PSNR of a specific view was calculated by computing the PSNR for each frame of the corresponding video sequence and averaging the PSNR values across the frames, as follow:

$$\begin{split} \text{MSE}(P,R,t) &= \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} \left[R(x,y,t) - P(x,y,t) \right]^2 \\ \text{PSNR}(P,R) &= \frac{1}{F} \sum_{t=0}^{F-1} 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}(P,R,t)} \right) \end{split}$$

where R(x, y, t) is the luma component of the t-th frame of the reference view R, P(x, y, t) is the luma component of the t-th frame of the processed view P, W, H, and F are the width, height, and number of frames of the video sequence, respectively, and MAX is the maximum possible pixel value (MAX = 255).

4. PERFORMANCE INDEXES

The results of the subjective tests can be used as ground truth to evaluate how well the objective PSNR-based measures estimate perceptual quality. The following properties of the VQR estimation of MOSs are considered in this study: accuracy, monotonicity, and consistency [14].

First, a linear least squares regression is fitted to each [VQR, MOS] data set, as compliant to the standard procedure for evaluating the performance of objective metrics [14]. The linear regression aligns the VQR range to the MOS range and allows removing any systematic offset, which may be present in the relationship between the objective and subjective data and is irrelevant for the goal of metric performance comparison. At the same time, it avoids the risk of data over fitting, which may occur when considering non-linear regression.

The linear regression is of the form:

$$MOS_p(VQR) = a \cdot VQR + b$$

Then, the Pearson linear correlation coefficient (PCC) and the root-mean-square error (RMSE) are computed between MOS_p and MOS to estimate the accuracy of the VQR. To estimate monotonicity and consistency, the Spearman rank order correlation coefficient (SCC) and the outlier ratio (OR), are computed between MOS_p and MOS, respectively [14].

The root-mean-square error (RMSE) and the outlier ratio (OR) are defined as follow:

$$RMSE = \sqrt{\frac{1}{(N-D)} \sum_{i=1}^{N} (MOS_i - MOS_{pi})^2}$$
$$OR = \frac{\text{total number of outliers}}{N}$$

outlier: point for which
$$|MOS_i - MOS_{ni}| > 2\sigma_i$$

N

where N is the total number of points, D is the degree of freedom for the curve fitting (linear: D = 2), and σ_i is the standard deviation corresponding to MOS_i .



Fig. 2. Subjective versus objective results. First row: Class A, second row: Class C.

5. RESULTS

Scatter plots of subjective versus objective results are presented in Figure 2 for Class A and Class C sequences. Significant differences in terms of minimum and maximum PSNR values can be noticed between the different contents, especially for Class A. This effect is due to the PSNR dependency on the spatio-temporal features of the content. Due to this strong content dependency, the performance analysis has been carried for each content separately. It can be noticed that results for each content are more scattered for the synthesized view and the intermediate view, while they are more clustered for the decoded view and the average of the decoded and synthesized views, especially for content S01.

For some contents (S03, S05, and S06), a few sequences have a low value for the PSNR of the synthesized (intermediate) view while the corresponding stereo pair has a high MOS. These data points are significantly distant from the trend in the scatter plot of the synthesized (intermediate) view. It is known that one proponent used a different view synthesis algorithm. Our hypothesis is that those results are from this specific proponent and that PSNR cannot be used to estimate the perceived quality in this case, which is coherent with the observations from Bosc *et al.* [9].

The accuracy, monotonicity, and consistency indexes of the objective video quality models, as defined in Section 4, are reported for each content separately in Table 2. The resulting values for the Pearson linear correlation coefficient confirm the graphical analysis. In general, the PSNR of the decoded view has the highest correlation with perceived quality. Two extreme cases can be noticed. First, content S01. In this case, the PSNR of the synthesized view (PCC=0.4489) has a significantly lower correlation than the average PSNR of the decoded and synthesized views (PCC=0.8502) or the PSNR of the decoded view (PCC=0.9216). Second, content S07¹. For this content, the PSNR of the decoded view (PCC=0.9219) expresses perceptual quality significantly better than the PSNR of the synthesized view (PCC=0.3788) or the PSNR of the intermediate view (PCC=0.1119). One content (S08) stands out of the trend, but the difference is insignificant. For this content, the PSNR of the synthesized view (PCC=0.9043) is more correlated with the perceived quality than the PSNR of the decoded view (PCC=0.8813).

Taking into account both views increases correlation with perceived quality as opposed to considering the synthesized (intermediate) view only.

The different performance indexes have similar behavior across the different contents. Considering overall performance indexes, averaged on all contents, the best performance is obtained using the PSNR of the decoded view (PCC=0.9200, SCC=0.9114), followed by the average PSNR of the decoded and intermediate views (PCC=0.9028, SCC=0.8892) and the average PSNR of the decoded and synthesized views (PCC=0.8834, SCC=0.8762). These objective video quality models outperform in general the PSNR of the synthesized view (PCC=0.6668, SCC=0.6797) and the PSNR of the intermediate view (PCC=0.5858, SCC=0.6234).

| | Decoded | ecoded Intermediate | Synthesized | Decoded and | Decoded and | Decoded | Intermediate | Synthesized | Decoded and | Decoded and |
|------|----------------------------------------|---------------------|-------------|--------------|-------------|---------------------------------------------|--------------|-------------|--------------|-------------|
| | Decoucu | | | intermediate | synthesized | | | | intermediate | synthesized |
| | Pearson linear correlation coefficient | | | | | Spearman rank order correlation coefficient | | | | |
| S01 | 0.9216 | - | 0.4489 | - | 0.8502 | 0.9452 | - | 0.4211 | - | 0.8509 |
| S02 | 0.9469 | - | 0.8158 | - | 0.9406 | 0.9534 | - | 0.7958 | - | 0.9373 |
| S03 | 0.9366 | 0.8130 | 0.8370 | 0.9199 | 0.9240 | 0.9362 | 0.8504 | 0.8674 | 0.9219 | 0.9266 |
| S04 | 0.8889 | - | 0.8476 | - | 0.8843 | 0.8372 | - | 0.8055 | - | 0.8412 |
| S05 | 0.9073 | 0.4592 | 0.4877 | 0.8498 | 0.8309 | 0.9078 | 0.5910 | 0.5831 | 0.8182 | 0.8019 |
| S06 | 0.9558 | 0.6477 | 0.6145 | 0.9235 | 0.9039 | 0.9773 | 0.6370 | 0.5934 | 0.9288 | 0.9049 |
| S07 | 0.9219 | 0.1119 | 0.3788 | 0.9186 | 0.8284 | 0.9158 | 0.1935 | 0.4704 | 0.9278 | 0.8723 |
| S08 | 0.8813 | 0.8972 | 0.9043 | 0.9024 | 0.9044 | 0.8185 | 0.8449 | 0.9006 | 0.8492 | 0.8747 |
| Mean | 0.9200 | 0.5858 | 0.6668 | 0.9028 | 0.8834 | 0.9114 | 0.6234 | 0.6797 | 0.8892 | 0.8762 |
| | Root-mean-square error | | | | | Outliers ratio | | | | |
| S01 | 0.9476 | - | 2.1812 | - | 1.2850 | 0.0220 | - | 0.2088 | - | 0.0659 |
| S02 | 0.7882 | - | 1.4169 | - | 0.8316 | 0.0000 | - | 0.0978 | - | 0.0217 |
| S03 | 0.7861 | 1.3062 | 1.2277 | 0.8795 | 0.8575 | 0.0000 | 0.0312 | 0.0208 | 0.0104 | 0.0104 |
| S04 | 1.1044 | - | 1.2790 | - | 1.1255 | 0.0217 | - | 0.0543 | - | 0.0217 |
| S05 | 1.1532 | 2.4362 | 2.3941 | 1.4455 | 1.5260 | 0.0172 | 0.2241 | 0.2241 | 0.0862 | 0.0862 |
| S06 | 0.8395 | 2.1743 | 2.2515 | 1.0945 | 1.2205 | 0.0000 | 0.1429 | 0.1429 | 0.0238 | 0.0476 |
| S07 | 0.6844 | 1.7550 | 1.6345 | 0.6981 | 0.9892 | 0.0000 | 0.0526 | 0.0526 | 0.0000 | 0.0000 |
| S08 | 1.1639 | 1.0878 | 1.0513 | 1.0615 | 1.0510 | 0.0185 | 0.0556 | 0.0370 | 0.0185 | 0.0185 |
| Mean | 0.9334 | 1.7519 | 1.6795 | 1.0358 | 1.1108 | 0.0099 | 0.1013 | 0.1048 | 0.0278 | 0.0340 |

Table 2. Accuracy, monotonicity, and consistency indexes of the objective video quality models under consideration.

¹It should be noticed that the MPEG committee has decided not to use this content anymore for further evaluation.

6. CONCLUSION AND FUTURE WORK

In this paper, the correlation between different PSNR-based metrics and the perceived quality of an asymmetric stereo pair has been investigated. The stereo pairs considered in this study were formed from a decoded view and a synthesized view having different visual quality. Results show that the PSNR of the decoded view has the highest correlation in terms of the Pearson correlation coefficient with perceived quality. Similar performance can be achieved when considering the average PSNR value of both views. On the other hand, the PSNR of the synthesized view has a significantly lower correlation with the subjective results. These conclusions are similar to that of previous studies on the perceived quality of asymmetric stereo pairs. Therefore, we suggest that the objective quality assessment for this scenario should be predicted by the PSNR of the decoded view and/or the average PSNR of both views rather than by the PSNR of the synthesized view.

To extend our work, state of the art 2D metrics, such as SSIM and VQM, and, even more important, 3D metrics should also be evaluated for the same target application in future investigations. The correlation between objective metrics and perceived quality of stereo pairs formed from two synthesized views should also be studied. Indeed, in this case, the two views contain artifacts due to the rendering process. Therefore, it can be expected that some properties of the human visual system, such as binocular suppression, will not have as much impact.

7. ACKNOWLEDGEMENT

We would like to thank Panos Nasiopoulos and Mahsa T. Pourazad from University of British Columbia (UBC), as well as Kjell Brunnström, Kun Wang, and Börje Andrén from Acreo AB for providing their subjective results for the 2view case scenario. This work was partially supported by the COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services - QUALINET.

8. REFERENCES

- L.B. Stelmach, W.J. Tam, D.V. Meegan, A. Vincent, and P. Corriveau, "Human Perception of Mismatched Stereoscopic 3D Inputs," in *Int. Conf. on Image Processing*, September 2000, vol. 1, pp. 5–8.
- [2] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," ACM Transactions on Applied Perception, vol. 3, pp. 95–109, April 2006.
- [3] G. Saygili, C.G. Gurler, and A.M. Tekalp, "Evaluation of Asymmetric Stereo Video Coding and Rate Scaling

for Adaptive 3D Video Streaming," *IEEE Transactions* on *Broadcasting*, vol. 57, no. 2, pp. 593–601, June 2011.

- [4] A. Smolic, K. Mueller, P. Merkle, and A. Vetro, "Development of a new MPEG standard for advanced 3D video applications," in *Proc. of 6th Int. Symp. on Image and Signal Processing and Analysis*, September 2009, pp. 400–407.
- [5] ISO/IEC JTC1/SC29/WG11, "Applications and Requirements on 3D Video Coding," Doc. N12035, Geneva, Switzerland, November 2011.
- [6] ISO/IEC JTC1/SC29/WG11, "Report on Experimental Framework for 3D Video Coding," Doc. N11631, Guangzhou, China, October 2010.
- [7] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on 3D Video Coding Technology," Doc. N12036, Geneva, Switzerland, November 2011.
- [8] ISO/IEC JTC1/SC29/WG11, "Report of Subjective Test Results from the Call for Proposals on 3D Video Coding Technology," Doc. N12347, Geneva, Switzerland, November 2011.
- [9] E. Bosc, M. Köppel, R. Pépion, M. Pressigout, L. Morin, P. Ndjiki-Nya, and P. Le Callet, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *Int. Conf.* on *Image Processing*, 2011, pp. 2597–2600.
- [10] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The Emerging MVC Standard for 3D Video Services," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, 2009.
- [11] G. J. Sullivan and J.-R. Ohm, "Recent developments in standardization of high efficiency video coding (HEVC)," in *Proc. SPIE* 7798, 2010.
- [12] ITU-R BT. 510-11, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, 2002.
- [13] F. De Simone, L. Goldmann, J.-S. Lee, and T. Ebrahimi, "Towards high efficiency video coding: Subjective evaluation of potential coding technologies," *Journal of Visual Communication and Image Representation*, vol. 22, no. 8, pp. 734–748, 2011.
- [14] ITU-T Tutorial, "Objective perceptual assessment of video quality: Full reference television," ITU-T Telecommunication Standardization Bureau, 2004.