A Testbed for Subjective Evaluation of Omnidirectional Visual Content

Evgeniy Upenik, Martin Řeřábek, Touradj Ebrahimi
Multimedia Signal Processing Group
Ecole Polytechnique Federale de Lausanne, EPFL
Lausanne, Switzerland
firstname.lastname@epfl.ch

Abstract—Omnidirectional visual content is a form of representing graphical and cinematic media content which provides subjects with the ability to freely change their direction of view. Along with virtual reality, omnidirectional imaging is becoming a very important type of the modern media content. This brings new challenges to the omnidirectional visual content processing, especially in the field of compression and quality evaluation. More specifically, the ability to assess quality of omnidirectional images in reliable manner is a crucial step to provide a rich quality of immersive experience. In this paper we introduce a testbed suitable for subjective evaluations of omnidirectional visual contents. We also show the results of a conducted pilot experiment to illustrate the applicability of the proposed testbed.

Index Terms—omnidirectional image and video compression, subjective evaluation, objective evaluation

I. Introduction

Omnidirectional image and video data is a specific type of visual content acquired and represented in such a way that a) the captured field of view covers full sphere, hemi-sphere, or spherical layer, and b) viewers are provided with ability to freely change the direction of their sight to observe different parts of omnidirectional content. An omnidirectional visual content could come with supplementary spatial 3D sound.

Recent advances in display technologies and improvements in computational power have resulted in a shift of general focus towards omnidirectional imaging as witnessed by increased interest in virtual and augmented reality. The initial interest in omnidirectional imaging from computer vision and robotics has now shifted to consumer multimedia. This brings novel requirements in desired features and properties of omnidirectional contents and applications. Presumably, the most important of these requirements is an ability to efficiently store, transmit, and render omnidirectional image data while preserving high visual quality. To improve quality of experience for an omnidirectional visual content, quality metrics and evaluations need to be investigated within applications exploiting such contents. Conventional objective metrics, such as PSNR, SSIM etc., have been developed for regular flat images and might not be suitable for omnidirectional content.

This paper introduces a testbed suitable for subjective evaluations of omnidirectional image and video content. Proposed testbed enables to collect not only the raw scores of participants corresponding to various methodologies and scoring scales but also to record the tracking information, which helps in deeper analysis of omnidirectional content and its consumption. More specifically, obtained data allows for example an accurate estimation and extraction of region of interest or development, validation and verification of novel objective metrics predicting more accurately the visual quality. Description and detailed analysis of results for a pilot subjective experiment to illustrate suitability and usability of the proposed testbed is also presented in the paper.

II. Related work

Several works have been published in the past proposing various approaches for compression of omnidirectional visual content. Main omnidirectional encoding schemes focus either on geometrical transformations [1], partial adaptive content delivery [2], or spherical nature of the picture [3]. Nevertheless, neither objective nor subjective quality evaluations of omnidirectional content have been thoroughly investigated yet.

In [4], Yu et al. proposed a framework for objective evaluation of omnidirectional video, which allows to compare different panoramic projections, such as equirectangular, Lambert cylindrical equal-area, dyadic, and cubic. In this work, authors introduced two methods. One is based on viewport evaluation, another is called spherical PSNR and takes into account relative pixel's positions on the sphere as well as experimental view direction trajectories recorded from subjects wearing a head-mounted display. Zakharchenko et al. introduced a metric which remaps evaluated omnidirectional pictures to Craster parabolic projection in order to compare different geometrical representations. [5]. However, rigorous validation and verification of proposed objective metrics including correlation analysis with subjective scores are missing.

In an earlier work [6], authors proposed no-reference quality assessment methods for omnidirectional video. They considered factors such as temporal and spatial fluctuations in perceived artifacts in addition to blockiness, bluriness, and jerkiness. Nonetheless, exploited dataset was limited only to video-conferencing content with statistically unreliable subjective evaluations.

To the best of our knowledge, there is currently no study considering subjective quality evaluation of omnidirectional visual content with clear description of methodology and test conditions as well as with proper analysis of results.

III. TESTBED

In this section we present a testbed for subjective evaluation of omnidirectional visual content. The main purpose of this testbed is to provide researchers with a tool to perform subjective assessment of omnidirectional images and video. The testbed consists of a software application for the hand-held and mobile platforms and can be used with head-mounted displays. The application is able to visualize omnidirectional images and video represented in different projections. The set of supported projections or geometrical representations can be easily extended. Moreover, the testbed provides a special customizable graphical user interface and storyboard for subjective quality evaluation experiments.

The application allows viewing images and video using head-mounted displays (HMD), such as Google Cardboard. These mounts do not contain the display itself and a mobile device with a screen must be inserted in order to use them. Additionally, images and video can be displayed by the application without HMD using only hand-held devices, i.e. tablet computers and mobile phones. Omnidirectional visual content is displayed with vertical field of view being set to 90 degrees. Horizontal field of view depends on the device being used and display mode (head-mounted or hand-held). Images and video frames are rendered on the screen via OpenGL ES API which is supported by most mobile devices. The proposed testbed supports visualization of different types of geometrical representations or projections of omnidirectional visual content. The current set includes equirectangular, cubic, half-back cubic, and icosahedral projections. However, because the interfaces are open and scalable, other projections can be easily implemented and added in the future by the authors or by third party contributors from the open source community. More information about different projections can be found in [4][7]. Used with HMD mount or on a hand-held device, the application allows to freely change the direction of sight using motion control. It is implemented by obtaining the attitude data from built-in accelerometer and magnetometer sensors in the mobile device or tablet computer. In the hand-held mode it is also possible to configure the application to use finger gestures to control the direction of view.

The testbed application provides a special storyboard of screens to perform subjective quality evaluation of omnidirectional visual content. The storyboard contains customizable textual instructions displayed to subjects. This text can be easily updated and can contain information such as instructions to the subject on how to proceed during different parts of evaluation. For example, during the training, it can include information on how to score: what part of the content subjects must pay attention to, what type of distortion and artifacts they should consider and what they should exclude from consideration. After a welcome message and basic instructions, participants proceed with the training session. During the training session no special action is required from subjects as the training stimuli are shown with corresponding votes already provided. Subsequently, during the evaluation session, subjects see the test stimuli and, when ready to give a quality score, they can activate a scoring menu. The scoring menu is displayed as an overlay on the screen and contains voting items according to the selected evaluation methodology. Once voting score is selected and recorded, the next stimulus is displayed immediately. Scoring menu as well as the entire storyboard can be customized for different subjective evaluation methodologies, such as single stimulus or double stimulus impairment scale. Data gathered from subjects are anonymized in the application by automatically assigning a unique identifier number to each subject at the beginning of each session.

Prior to the experiment, the stimuli set must be prepared by the experimenter according to a special naming convention and uploaded to the device in order to perform the tests. Additionally, experimenter explicitly marks the images for the training session and dummy stimuli. The proposed testbed application automatically performs randomization of the test stimuli according to the ITU recommendations described in section 2.7 of [8], in such way that the same content is never shown consecutively.

For each evaluation session the proposed testbed application stores: subjects' scores, tracks of direction of view for each stimulus, and time spent by subjects on each stimulus. The direction of subjects' sight during the evaluation of each stimulus is represented by two coordinates: *yaw* and *pitch*, where the former represents horizontal and the latter vertical angles. Attitude coordinates correspond to the center of the current viewport. A separate track is recorded for each test stimulus with configurable sampling rate, e.g. 10 to 60 samples per second. Additionally, during the evaluation session, time spent by the subject on each stimulus is measured. It is also possible to restrict the subjects to spend limited time on each stimuli. All the experimental data, including scores and viewing direction tracks is stored locally on the device in CSV and XML. It can be transfered from the device to a server or a work station for further analysis.

IV. EXPERIMENT

This section describes the pilot experiment consisting in subjective and objective evaluation of omnidirectional images and exploiting the above proposed testbed. It presents details about dataset, objective metrics, and test methodology used for subjective quality assessment, as well as the processing of the collected raw subjective scores and correlation analysis.

A. Dataset preparation and description

Six different 360 image contents were used in the experiment. Original unstitched and uncompressed raw samples from *Ladybug 5* omnidirectional camera were provided by Point

(a) *AtTheMall*

(b) *InsideDepot*

(c) *IntersectionTown*

(d) *Residential*
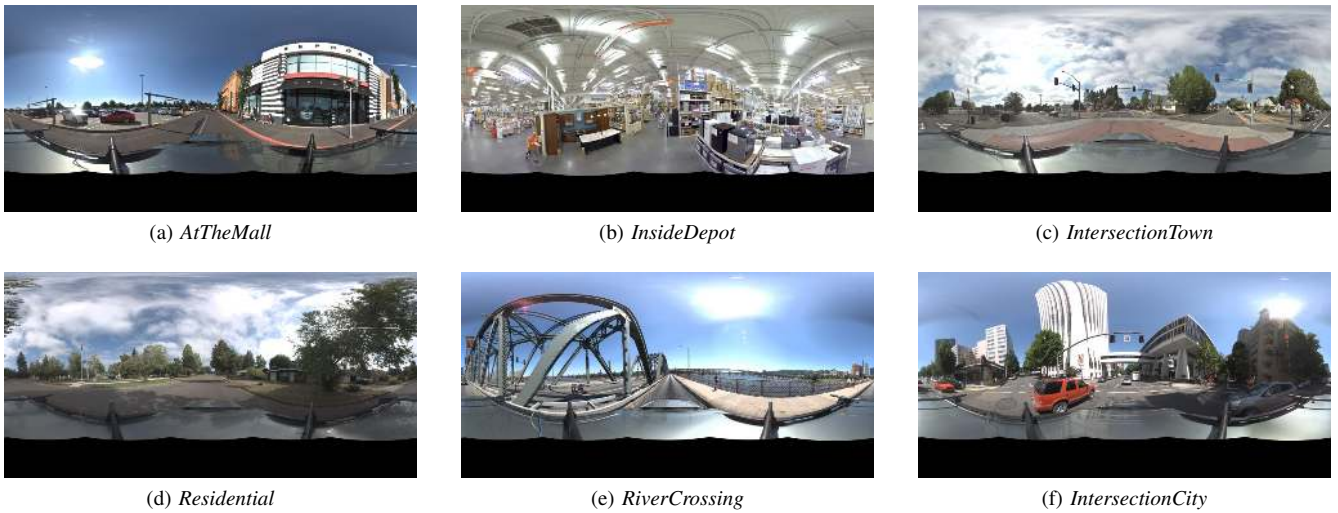
(e) *RiverCrossing*

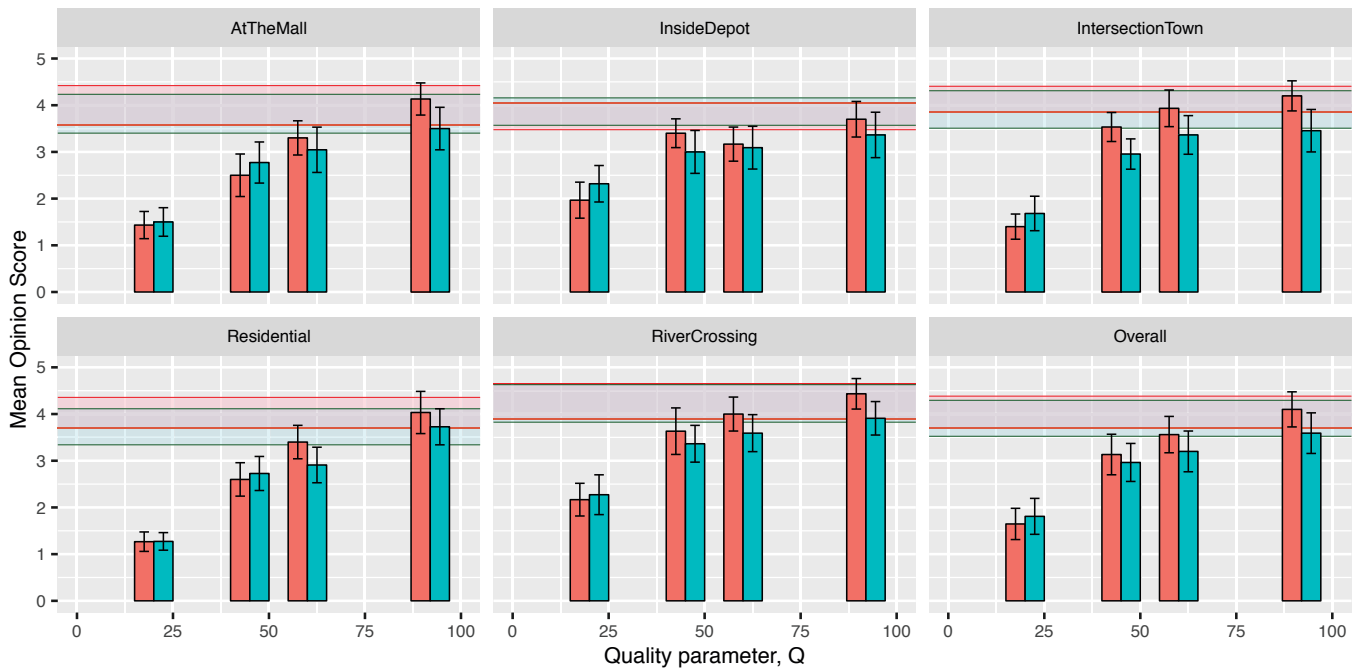(f) *IntersectionCity*

Fig. 1: Images used in the experiments.

Fig. 2: MOSs with CIs obtained using ACR-HR method for JPEG-compressed omnidirectional images. Red bars depict MOSs for equirectangular projection, cyan bars are for cubic. Filled area between two horizontal lines corresponds to the confidence interval of the reference for each projection (red for equirectangular, cyan for cubic). JPEG quality parameters: 20, 45, 60, 92.

Grey Research Inc.[1] Images were processed and stitched to equirectangular projection using LadybugSDK software package. Figure 1 shows examples of equirectangular projections of exploited test samples.

Reference images in the prepared stimuli set were in lossless compressed PNG format, sRGB 8-bit color space, and resolution of 3000x1500 pixels for equirectangular format and 2250x1500 for cubic. Test stimuli were prepared by compressing the reference images with JPEG encoder using four different quality parameters for each image: 20, 45, 60, and 92. The latter was set to achieve transparent quality. The stimuli used in the experiment contain a training set of five stimuli obtained from the same content representing all the quality levels including original and an evaluation set consisting of 25 images for each of two projections obtained from five different contents.

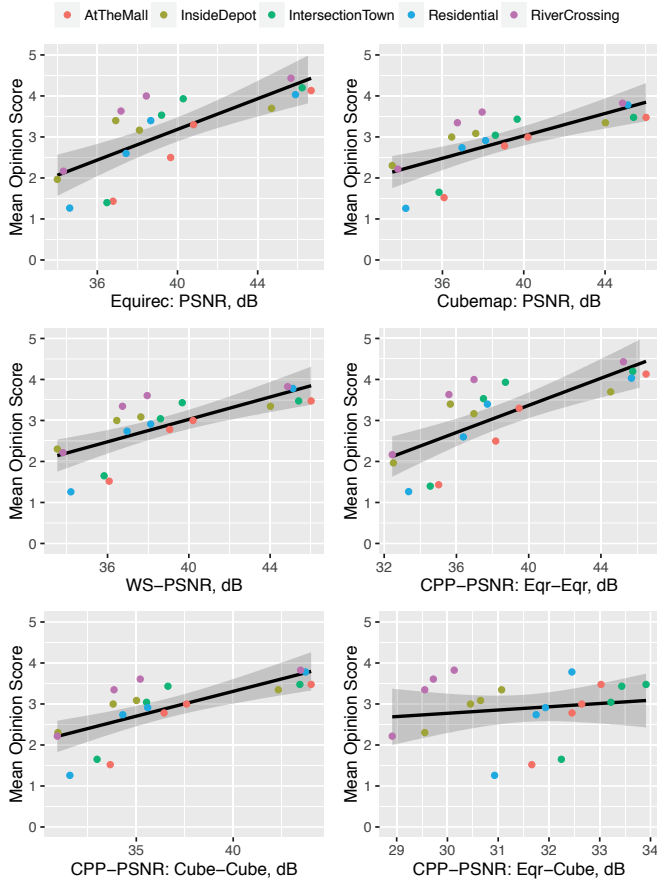To select the lower and upper quality bounds, an expert

Fig. 3: Mapping of objective scores to subjective ratings.

screening session was conducted for each content separately trying to cover the full quality scale for each content.

B. Test methodology

The Absolute Category Rating with Hidden Reference (ACR-HR) methodology [9] was selected for evaluations. A five-grade quality scale (1: Bad; 2: Poor; 3: Fair; 4: Good; 5: Excellent) was used. The subjects were asked to judge the overall quality of the evaluated omnidirectional images. To reduce contextual effects, the order of rendered stimuli was randomized so that the same content was never shown consecutively. This is done automatically by the proposed testbed software.

Half of 48 subjects participating in the subjective assessment study evaluated test stimuli represented in equirectangular format, another half evaluated cubic projection. Subjects (36 males and 12 females) were between 19 and 36 years old with the corresponding age average and median of 25.1 and 24.7 years respectively.

In order to minimize visual fatigue effects, each test session was designed to take no more that 15 minutes. Before the experiments, short written training instructions were provided to subjects to explain their tasks. Subsequently, more detailed instructions were shown on the screen during the training

session, where five training samples, representing all quality levels were displayed to familiarize subjects with the assessment procedure. The training instructions and samples as well as the entire test were presented to subjects using the proposed evaluation testbed described above.

Subjects were asked to watch the images using special HMD mount ("MergeVR"[2]) with a mobile device (iPhone 6S) placed inside as a screen. Resolution of the device was 1334x750 pixels overall, or 667x750 per eye. The HMD mount used featured two buttons on the top to select and scroll within a menu in order to rate the quality of presented stimuli. All subjects were asked to sit on a rolling chair with ability to easily turn in all directions.

C. Data processing

Several PSNR-based objective measures of luma component for each of the test stimuli (original omnidirectional images represented in equirectangular projection encoded with various compression ratios) were used for objective quality evaluation. The metrics were PSNR, WS-PSNR and CPP-PSNR [5]. PSNR was computed between images of the same projection (for equirectangular and cubic). WS-PSNR is only suitable for equirectangular projection. The only objective metric which allows comparing different projections is CPP-PSNR. Stimuli represented in cubic projection were compared to uncompressed equirectangular images. Samsung 360 Tools[3] was used to calculate WS-PSNR and CPP-PSNR. The results of objective evaluations were further used for correlation analysis.

To evaluate perceived quality, standard statistical indicators describing distribution of scores across subjects for each test condition were computed. First, outlier detection was performed according to the guidelines described in Section 2.3.1 of Annex 2 in [8] to remove subjects whose scores deviated strongly from others. In our experiments, two subjects were detected as outliers for all test sessions. Then, the mean opinion score (MOS) and corresponding 95% confidence intervals (CI), assuming a Student's t-distribution of the scores, were computed for each test condition. (See Figure 2). Finally, a correlation analysis indicating how well the objective metrics estimates the perceived quality of an omnidirectional image content was performed. A regression was fitted to the [Objective metric, MOS] data sets to map the objective scores to the subjective ratings. (See Figure 3). Three standard performance indexes were computed. Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation (SROCC) estimate linearity and monotonicity respectively. Accuracy and consistency were assessed using the root-mean-square error (RMSE).

D. Results and discussion

Figure 2 shows the resulting MOS and CI plots for equirectangular and cubic projections obtained from subjective evaluation experiment performed using ACR-HR method. Areas
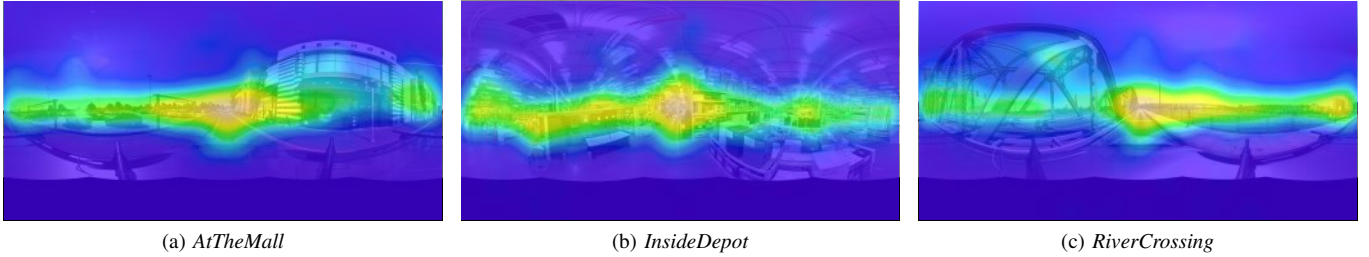
| (a) *AtTheMall* | (b) *InsideDepot* | (c) *RiverCrossing* |

Fig. 4: Experimentally obtained saliency maps for some test images.

TABLE I: Standard performance indexes

Objective Metric	Pearson	Spearman	RMSE
Equirec: PSNR	0.7636	0.8281	37.5939
Cubemap: PSNR	0.7362	0.7911	37.2267
WS-PSNR	0.7535	0.8229	35.1002
CPP-PSNR: Eqr-Eqr	0.7584	0.8206	36.4556
CPP-PSNR: Cube-Cube	0.7183	0.7971	34.7583
CPP-PSNR: Eqr-Cube	0.1558	0.1950	29.3769

between two horizontal lines correspond to the confidence intervals of the reference stimuli for each projection. One can see from the plots that some different contents stimuli compressed with quality parameters 60 and 92 are perceived as transparent to the original, and for some contents stimuli compressed with quality parameter 45 are perceived as transparent as well. However, the MOSs for the stimuli with the lowest quality parameter 20 have statistically significant differentiation from the neighbors for all the contents.

There exist a tendency among subjects to under-rate the stimuli. It can be possibly explained by the fact that many of participants reported themselves evaluating existing stitching artifacts, despite the instructions.

According to the recorded values of the time spent to evaluate each test stimulus by the participants of the experiment with combined number of 48 volunteers, in average subjects looked at an omnidirectional image for a duration of 30 seconds. This data can be used in designing future subjective quality evaluation tests for omnidirectional images and video.

In Figure 3 one can find scatter plots of MOSs values against objective metrics (PSNR, WS-PSNR, and CPP-PSNR). Linear regression fitting was applied to the data, and its results are represented by black solid lines. Gray area near the lines corresponds to 95% confidence intervals for predictions from linear model. From the plots in Figure 3 and Table I it can be seen that PSNR-based metrics do not show hight correlation with the MOSs obtained from the subjective evaluation experiment. However, because the main purpose of this study is to validate the proposed testbed, the amount of obtained experimental data is not enough to make statistically significant conclusions about performance of objective metrics, and further investigations are needed.

Figure 4 depicts statistical frequencies of looking at a par-

ticular part of the omnidirectional image or empirical saliency maps, obtained from the recorded tracks of subjects' direction of sight, described in Section III. These saliency maps were constructed assuming that the focus of interest is concentrated in the center of a given viewport. Experimental saliency maps data can be further analyzed and used for accurate estimation and extraction of regions of interest in omnidirectional images and video.

V. CONCLUSION

In this paper we demonstrated a testbed for subjective quality evaluation of omnidirectional visual content. The proposed testbed allows to perform experiments using different methods. The software application has customizable storyboard and grading menus. Experimental data, which can be obtained with the testbed, includes subjective scores, time spent on stimulus, and view direction tracks.

We performed a pilot experiment to validate the proposed testbed and presented the analysis of obtained data, including MOSs and correlation analysis of objective metrics and perceived quality. Additionally, we computed saliency maps for each content using experimentally obtained direction of sight tracks.

REFERENCES

[1] I. Bauermann, M. Mielke, and E. Steinbach, "H.264 based coding of omnidirectional video," in *Computer Vision and Graphics*, ser. Computational Imaging and Vision, 2006, no. 32, pp. 209–215.

[2] P. Rondao Alface, J.-F. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," vol. 16, no. 4, pp. 135–147, 2012.

[3] I. Tosic and P. Frossard, "Low bit-rate compression of omnidirectional images," in *Picture Coding Symposium, 2009. PCS 2009*, pp. 1–4.

[4] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2015, pp. 31–36.

[5] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *SPIE Optics + Photonics*, vol. 9970, 2016.

[6] S. Leorin, L. Lucchese, and R. G. Cutler, "Quality assessment of panorama video for videoconferencing applications," in *2005 IEEE 7th Workshop on Multimedia Signal Processing*, 2005, pp. 1–4.

[7] C.-W. Fu, L. Wan, T.-T. Wong, and C.-S. Leung, "The rhombic dodecahedron map: An efficient scheme for encoding panoramic video," vol. 11, no. 4, pp. 634–644, 2009.

[8] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, January 2012.

[9] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, April 2008.