Visual Quality Analysis of Judder Effect on Head Mounted Displays

Saeed Mahmoudpour and Peter Schelkens
Vrije Universiteit Brussel (VUB), Dept. of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium
imec, Kapeldreef 75, B-3001 Leuven, Belgium
{smahmoud, pshelke}@etrovub.be

***Abstract*—The extended field of view (FoV) offered by head mounted displays (HMD) increases the immersive experience, but it also introduces new visual quality challenges to be addressed. The judder artefact is a quality degradation factor that appears during object tracking and it is caused by eye movement relative to the display. As a first attempt to investigate the negative effect of judder in wide FoV applications, we built a new dataset of omnidirectional videos at different bitrates and judder severity levels. Two subjective tests were conducted to assess the quality in terms of perceived severity of judder in compressed video sequences. The outcomes provide new findings about the effect of judder on human perception. The results also give further understanding about the interaction between presentation quality and judder that can be utilized for developing objective models to predict quality degradation in presence of judder.**

***Index Terms*—omnidirectional video, visual quality, judder, field of view, object tracking**

I. Introduction

In recent years, noticeable effort has been devoted towards development of wide field of view (FoV) displays. With the advent of omnidirectional (360°) capturing devices and advanced head mounted displays (HMDs), we have witnessed a significant rise of 360° multimedia content. The display FoV for offering an immersive experience can cover a full 360° sphere or it may be limited to a spherical segment.

The growth of omnidirectional video in consumer market introduces various challenges of which measuring the visual quality of experience (QoE) of end-users is a critical one. Therefore, it is of importance to study the effect of various degradation factors influencing the QoE of omnidirectional videos. Earlier studies have focused on factors such as compression [1], stalling [2], and resolution [3].

Judder is a visual artefact caused by eye movement relative to the display and it is considered as a critical quality issue that impacts the sense of immersion and viewer engagement. When watching a moving object in a video, the image of the object slips over the retina for a duration of frame and falls on incorrect spatial locations which results in temporal aliasing. This QoE factor can be observed as a non-smooth choppy motion blur. Johnson et al. [4] analyzed the judder effect based on object speed using window of visibility (WoV) theory. Oh et al. [5] proposed an objective metric to measure the judder artefact in 2D video sequences according to spatial and temporal features.

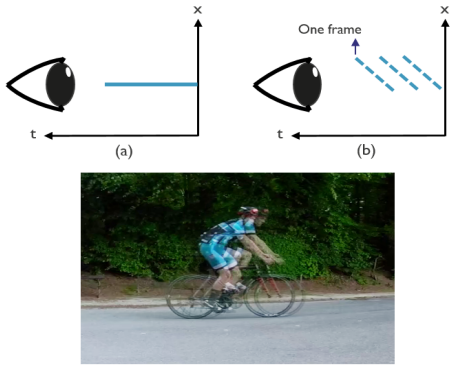

Fig. 1. Demonstration of the judder effect. (a) an ideal (real-world) object tracking. (b) object tracking on a digital display. (c) Visualization of the judder effect during object tracking.

Although, judder is manifesting itself on all displays, particularly at lower frame rates, it is more prominent on wide FoV displays (such as HMDs) and contributes to motion sickness. Because of the extended FoV, scene motions can be observed and tracked for longer duration than in traditional 2D displays. Moreover, the user can turn his head at much higher speed, which increases the slip rate between eye and object and hence causes more severe presence of judder effects. Figure 1 illustrates the perceptual effect of judder during object tracking. The space-time diagram in Fig. 1a shows an ideal object tracking in real-world where the relative position of object with eye is always constant. Note that the x axis indicates the displacement of object relative to the eye and not to the real world. Figure 1b presents object tracking on a display with a given refresh rate in which the virtual position of an object slips over the eye and falls on incorrect positions during each frame. The relative object-eye displacement induces noisy sensory motions to the eyes and produces a judder effect that is simulated in Fig. 1c for visualization purpose.

A straightforward solution to reduce the judder is to increase the displays refresh rate. Unfortunately, refresh rates of current displays are still not sufficient to compensate for this artefact. In case of HMDs, consumer devices (such as Oculus Rift and HTC Vive) are operating at refresh rates of 90Hz, which is still way below the desired refresh rates to completely surpass the judder. As mentioned earlier, the judder is more significant

978-9-0827-9703-9/19/$31.00

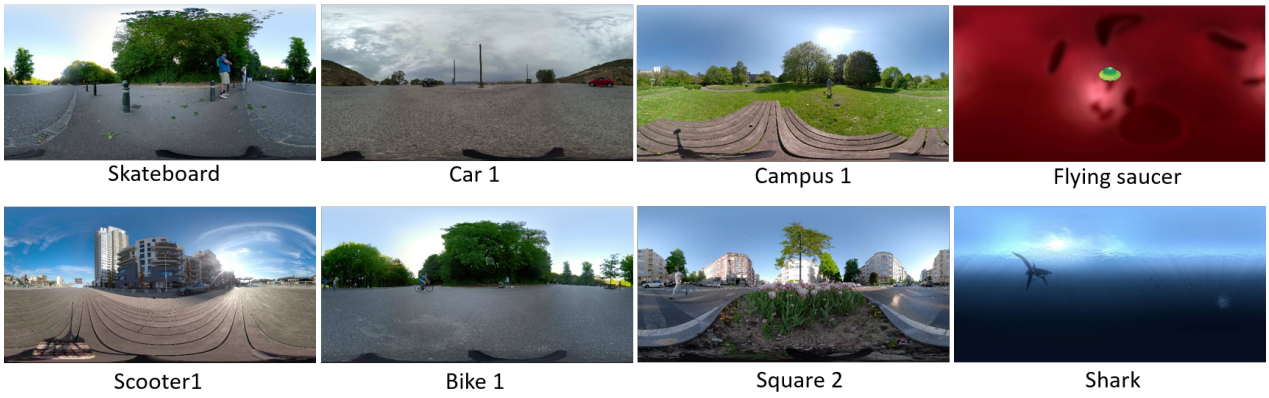

Fig. 2. Screenshots of some video sequences used in the database

in wide FoV displays due to high degree of freedom and faster head speeds, thus the required refresh rate for judder suppression in such displays is even higher than the one needed for 2D displays. This needs a significant improvement in capturing and display technologies which may not be achieved in near future to completely solve the judder impact.

Another major requirement for any type of video content is the ability to compress the data for efficient storage and transmission. Most of the existing objective quality metrics (such as Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM) [6] index, and Visual Information Fidelity (VIF) [7]) index are generally useful to quantify the quality loss caused by compression artefacts, whereas such metrics lack features to consider other quality factors such as judder. Since compression artefacts are a prevalent type of distortion observed in all video types, we decided to investigate the impact of judder on videos compressed at various bitrates. Zeng et al. [8] studied the effect of video compression and playback stalling factors for quality assessment. The joint effect of quantization and frame rate artefacts is explored in [9]. Bitrate reduction for video compression leads to degradation of the presentation quality [8] of decoded video. On the other hand, users experience a negative judder impact when tracking objects in videos. In such cases, judder may interact with presentation quality causing impacting the overall QoE. Therefore, we conducted a subjective experiment to clarify the relation between judder and presentation quality. More specifically, we established a new database[1] of omnidirectional videos to assess the judder effect, which have been compressed at three bitrate levels to modulate the video presentation quality. Then, subjective tests are conducted to evaluate the QoE of videos. The reported results reveal a number of outcomes that are helpful for future development of new objective QoE models that can measure the impact of judder on video quality.

II. Subjective Test Methodology

The existing publicly available omnidirectional video databases are mostly focused on the presentation quality affected by compression artefacts. To the best of our knowledge, the impact of judder has never been studied for omnidirectional video content in which object tracking can produce significant judder artefacts. Moreover, this work is the first attempt to investigate the dependency between the presentation quality and judder. To this end, we established a new omnidirectional video database with certain features to study the effect of judder and videos of different bitrates.

The database encompasses fifteen high-quality video sequences with resolution 3840x1920 @30 fps, including twelve natural and three synthetic videos. The natural video sequences are captured with a GoPro Fusion 360° camera. Two of the synthetic sequences (Airplane and Flying saucer) are generated using a video animation software and one video (Shark) is downloaded from Youtube with getting permission from the video owner. To induce the judder effect, each video scene presents a target object/person that is moving in a certain direction; tracking the scene objects by the viewer can cause judder as a visual artefact. Figure 2 shows screen shots for some of the source video sequences used in the experiment.

Since judder is a function of tracking speed, the video sequences cover three velocity levels for the tracked objects referred as low, medium and high velocity. The average angular velocity of a target object in each video sequence is approximately obtained by computing the displacement of the target between successive frames. The initial position of an object is manually located in the first frame and then tracked in next frames.

The distance between the centre points of the bounding boxes covering a target object in two frames is measured as a displacement value. Since an omnidirectional video frame is presented in form of an equi-rectangular projected image, the captured 360° content is a projection to spherical image plane. Hence, the displacement value should actually be computed between two points on the sphere [10] and we first need to convert the points from 2D Cartesian to spherical coordinates.

A pixel point on a spherical plane is defined with spherical coordinates radius r, polar angle θ, and azimuthal angle ϕ. Having a sphere with unit radius, a point is translated from

[1] http://data.etrovub.be/qualitydb/judder-vqa

TABLE I
THE PROPERTIES OF REFERENCE VIDEOS

	Video name	Bitrate (Mbps)	Velocity (deg/s)
Low Velocity	Campus1	493.59	23
	Skateboard	499.37	22
	Scooter1	358.21	22
	Car1	403.08	21
	Shark	109.69	20
Med Velocity	Campus2	595.67	55
	Bike1	460.44	41
	Square1	440.90	45
	Car2	401.13	50
	Airplane	30.98	40
High Velocity	Scooter2	350.02	64
	Square2	441.71	70
	Bike2	446.44	66
	Car3	401.97	72
	Flying saucer	30.25	63

Cartesian to the spherical coordinates as follows:

$$\theta = \frac{2\pi}{W}x, \qquad \phi = \frac{\pi}{H}y \tag{1}$$

where x, y are the Cartesian coordinates and W, H denote the width and length of the equi-rectangular image. Having two points on the spherical plane (θ_1, ϕ_1), (θ_2, ϕ_2), the angle at center of the sphere separating two points is computed using the haversine distance $\Delta\lambda$ [11].

$$\Delta\lambda = 2\arcsin\sqrt{(\sin^2\frac{\Delta\phi}{2} + \cos\phi_1.\cos\phi_2.\sin^2\frac{\Delta\theta}{2})} \tag{2}$$

Hence, the displacement value is expressed by the angular velocity of the object. The average angular velocity for all frames is used to measure the object velocity.

The source videos are divided into three categories based on the object velocity levels and listed in Table 1 with detailed information.

Each source video sequence is encoded at three bitrate levels using H.265/HEVC. The selected bitrates for natural sequences are 4 Mbps, 12 Mbps, and 40 Mbps. The animations are encoded at 800 Kbps, 1.3 Mbps and 2 Mbps. In total, we obtained 60 test video sequences including 15 source and 45 compressed videos. The test videos are viewed using Oculus Rift HMD with 2160x1200 resolution @ 90 Hz and FoV of 110 degrees. A single stimulus quality scoring strategy is used for the test to which a total number of 20 naive test subjects (12 males and 8 females, age range 20 to 35 year) participated. All test subjects passed the visual acuity and colour vision Ishihara tests before participating the subjective experiment.

Two experiments were designed for quality assessment of the generated videos. The first test only focused on the presentation quality (in terms of compression artefacts) for reference purposes, while the second test considered both compression and judder effects in quality evaluation. A dedicated training session was conducted at the beginning of each experiment to provide sufficient information for quality rating while avoiding subjects to become over-educated or biased.

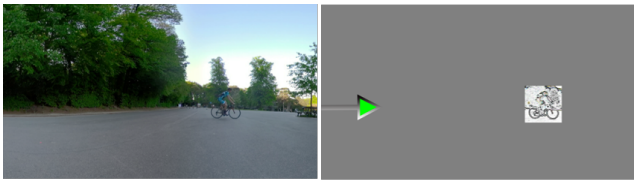

Fig. 3. (a) a segment of the first frame of equi-rectangular video (bike1) (b) The guide-mask to fix the view direction of users prior to tracking.

For experiment 1, the compression artefacts are introduced in the training session such that test subjects understand the type of expected distortions. Thereafter during the test session, each video sequence was displayed two times so that the subjects had enough time to freely explore different parts of the scene. The sequences were presented in random order to each subject and a scoring panel was displayed after each video sequence to record the quality score. As a result, all subjects evaluated the presentation quality of the videos by considering the compression distortions. In this part of the experiment, the naive subjects were not informed about the goal of second test (judder effect) and they were instructed to only focus on the compression as the effective distortion type.

In experiment 2, additional instructions were given during the training session to understanding the judder effect that appears during object tracking. An 8-second guide-mask (Fig. 3) is inserted before each video playback so that the subjects can locate the starting position of a target moving object. The guide-mask helps users to have prior knowledge about the location and direction of the tracking and fixate the eyes on the object of interest. This provides smooth tracking experience and avoids potential tracking failures by subjects. After the guiding period, the video is played and the subject starts tracking a target object in the scene. Finally, the overall quality score is rated by considering both judder effect and compression artefacts observed during tracking. Unlike the first experiment, each video is displayed only once.

The whole test takes one hour and 15 minutes per subject which is divided into four sessions with a 5-minutes break in between the sessions to avoid visual fatigue and motion sickness. Figure 4 illustrates the structure of the two test sessions and the scoring strategy for rating the quality. The scores range from 0 to 100 respectively from poorest to best video quality.

III. RESULTS ANALYSIS

The results of 20 subjects were gathered from two experiments. A rejection analysis was performed to detect outlier subjects. One outlier subject was detected and removed, and hence 19 valid test subjects remained. The subjective rates were converted to Z-scores [12] using mean and standard deviation of quality rates. Then for each video sequence, mean opinion score (MOS) is computed by averaging the Z-scores of all subjects. Thus, two MOS scores were obtained for each video based on the two experiments; (a) a MOS

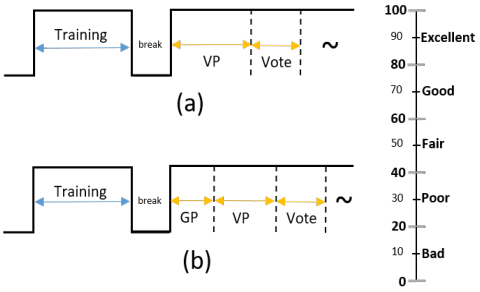

Fig. 4. The scoring method and subjective test protocols for experiments 1 and 2. VP: video playback, GP: guiding period.

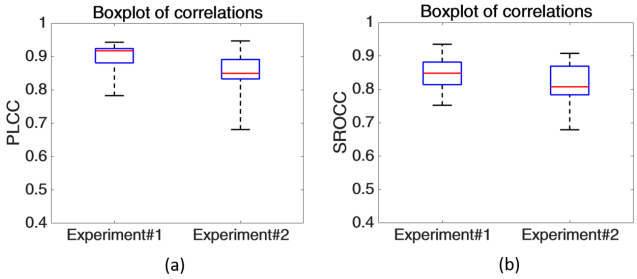

Fig. 5. Boxplots indicating the correlation between the scores of subjects and MOS in experiments 1 and 2; (a) Pearson linear correlation coefficient (PLCC) and (b) Spearman rank-order correlation coefficient (SROCC)

representing the presentation quality (compression) (MOS_C) and (b) a judder-compression overall quality (MOS_{JC}).

To evaluate the performance of the test subjects in the two experiments, we computed the correlation of scores obtained from each subject versus MOS as ground-truth data. Figure 5 presents the boxplots of correlation coefficients for all subjects in terms of Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC). As it can be observed, the test subjects are generally performing well in predicting the quality in both experiments.

To observe the quality variations in presence of judder, a scatter plot of presentation quality MOS_C versus MOS_{JC} (experiment2 scores) for 60 video sequences at three velocity levels is depicted in Fig. 6a. The data points below the diagonal (black) line imply the quality drops due to judder while the data points above the line indicate quality improvements. Fig. 6b shows a logistic curve fitted on scores at different velocity levels which helps to better observe the quality changes at three velocity levels. The relation between two groups of MOS values provides a number of important observations described.

For videos with low velocity objects, the tracking experience does not induce noticeable judder and the overall quality is not dropped. Besides, visual masking effect on compression artefacts occurs resulting in decrease of the visibility of artefacts and quality rising during tracking. Such improvement is slightly stronger at the middle-range quality than low quality because the artefacts are still highly visible at low quality while they are better masked at middle-range quality. At high quality, where the compression artefacts are minor, the

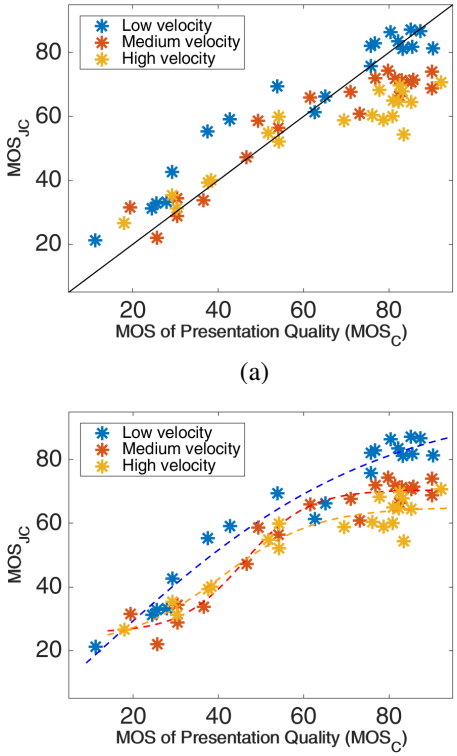

Fig. 6. Scatter plots of the subjective scores from two experiments.

TABLE II
THE AVERAGE MOS IN TWO SUBJECTIVE EXPERIMENTS OVER THREE BITRATES AND VELOCITY LEVELS.

	Bitrates			Velocity levels		
	R1	**R2**	**R3**	**Low**	**Medium**	**High**
MOS_C	28.18	58.32	80.48	60.12	62.64	64.31
MOS_{JC}	33.15	59.82	70.92	65.06	57.38	55.13

MOS_{JC} is very close to MOS of presentation quality.

For videos with medium and high object velocity, the judder effect is more critical. At higher speeds, less artefacts may be visible due to the motion masking effect but the judder on the other hand induces a negative quality impact. At lower quality levels, the masking and judder interact each other and there is not a noticeable change in MOS_{JC} compared to the presentation quality. By contrast, at higher quality levels (MOS>60) the judder impact is stronger causing a significant drop of MOS_{JC} compared to MOS_C. The quality drop is more severe at high velocity compared to medium velocity level. Table 2 compares the average MOS_C and MOS_{JC} at different bitrates and velocity levels, which confirms our observations. In general, the MOS_{JC} is slightly improved than MOS_C at low bitrates while it significantly drops at high bitrates and high speeds due to the impact of judder.

The videos are evaluated using several full-reference (FR) and no-reference (NR) quality metrics and the correlation is computed between objective scores and subjective differential MOS (DMOS). The DMOS of the compression

TABLE III
PERFORMANCE OF OBJECTIVE VQMS ON TEST VIDEOS.

Metrics	$DMOS_C$		$DMOS_{JC}$	
	PLCC	SROCC	PLCC	SROCC
PSNR (FR)	0.687	0.688	0.398	0.473
SSIM [6] (FR)	0.602	0.625	0.373	0.380
HDR-VDP [13] (FR)	0.725	0.739	0.478	0.480
VIF [7] (FR)	0.930	0.931	0.682	0.695
VQM [14] (FR)	0.718	0.720	0.436	0.458
BRISQUE [15] (NR)	0.600	0.621	0.512	0.526
dipIQ [16] (NR)	0.793	0.802	0.488	0.549
HOSA [17] (NR)	0.676	0.614	0.345	0.305

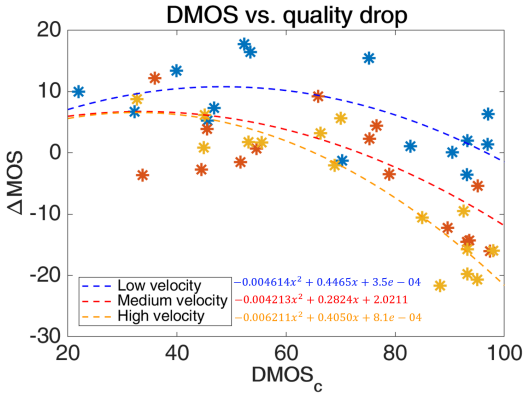

Fig. 7. The scatter plot of $DMOS_C$ versus quality drop caused by judder and the fitted curves and fitting parameters

quality $DMOS_C$ is obtained by computing the difference of MOS_C in the reference and compressed videos. The DMOS of the videos considering judder $DMOS_{JC}$ is computed as ($DMOS_C$+ΔMOS) where ΔMOS denotes the quality difference (MOS_{JC} - MOS_{JC}) caused by judder. Table 3 shows the consistency of objective metrics with $DMOS_C$ and $DMOS_{JC}$. Most of the existing metrics perform reasonably well when only the compression artefact is considered, whereas all metrics fail to accurately predict the quality when considering the judder degradation impact. Therefore, it is necessary to design new QoE models and tune current metrics for better performance. As an initial effort, we provide an empirical model of MOS changes (ΔMOS) based on the current subjective tests that can be used as a correction function to better predict the quality in presence of judder. The fitted model is given as $f(x) = b_1x^2+b_2x+b_3$. Figure 7 presents the quality difference versus the presentation quality DMOS together with the fitted curves and fitting parameters computed at three velocity levels. The velocity is required to be known to choose one of these three models. In future, we will further explore the velocity estimation methods and more sophisticated QoE models for objective quality assessment of videos with judder.

IV. CONCLUSION

A database of 60 omnidirectional video sequences is generated to study the joint effect of judder and compression artefacts. The dataset is used to perform subjective experiments to model the human QoE responses in presence of judder. The test results help to understand the relation between judder and tracking speed and the impact of the overall content quality. An interesting observation is that the quality drop due to judder is more significant at high quality levels. In addition to the subjective testing methodology, we presented simple models to describe the quality variations caused by the judder artefact. The reported results provide guidelines to design more suitable video QoE models that can consider the judder impact in quality prediction.

ACKNOWLEDGMENT

This research work is funded by the research project imec ICON ILLUMINATE HBC.2018.0201. We thank Curiscope company for granting us the permission to publish the video 'Shark' in our dataset.

REFERENCES

[1] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 1–1, 2018.
[2] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *Int. Conf. Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.
[3] R. Zhou, M. Huang, S. Tan, L. Zhang, D. Chen, J. Wu, T. Yue, X. Cao, and Z. Ma, "Modeling the impact of spatial resolutions on perceptual quality of immersive image/video," in *Int. Conf. on 3D Imaging (IC3D)*. IEEE, 2016.
[4] P. V. Johnson, J. Kim, D. M. Hoffman, A. D. Vargas, and M. S. Banks, "Motion artifacts on 240-Hz OLED stereoscopic 3D displays," *Journal of the Society for Information Display*, vol. 22, no. 8, pp. 393–403, 2014.
[5] S. R. Oh, S. Jeong, P. G. Heo, D. Kim, H. Y. Kim, and H. W. Park, "A new no-reference method for judder artifact assessment," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 1–1, 2018.
[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[7] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
[8] K. Zeng, H. Yeganeh, and Z. Wang, "Quality-of-experience of streaming video: Interactions between presentation quality and playback stalling," in *IEEE Int. Conf. Image Processing (ICIP)*. IEEE, 2016.
[9] Y. F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Trans. Circuits Sys. Vid. Technology*, vol. 21, no. 3, pp. 286–298, 2011.
[10] B. Xu, S. Pathak, H. Fujii, A. Yamashita, and H. Asama, "Optical flow-based video completion in spherical image sequences," in *IEEE Int. Conf. Robotics and Biomimetics (ROBIO)*. IEEE, 2016.
[11] L. M. Kells, W. F. Kern, and J. R. Bland, ," in *Plane And Spherical Trigonometry*. McGraw Hill Book Company, Inc., 1940, pp. 324–326.
[12] A. M. van Dijk, J. B. Martens, and A. B. Watson, "Quality asessment of coded images using numerical category scaling," in *Advanced Image and Video Communications and Storage Technologies*. Proc. SPIE, 1995, vol. 2451, pp. 803–806.
[13] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 010501, 2015.
[14] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
[15] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
[16] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.
[17] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Processing*, vol. 25, no. 9, pp. 4444–4457, 2016.