A perception-based quality model and quality metric for omnidirectional images

Phu Nguyen Minh[1], Huyen T. T. Tran[2], Nguyen Phung Dinh[1], Nam Pham Ngoc[1], Truong Cong Thang[2]
[1]Hanoi University of Science and Technology, Hanoi, Vietnam
[2]The University of Aizu, Aizuwakamatsu, Japan

***Abstract*—Virtual Reality (VR) is a technology that has been gaining popularity in the last few years and is capable of applying in many fields. However, current quality models and quality metrics which can be used to measure omnidirectional images are still limited. In this work, we propose a new quality model based on human perception for omnidirectional images, and from that model, we build a quality metric taking into account user's perception, called foveated peak signal to ratio (FPSNR). We also compare it with PSNR, a traditional quality metric. Based on subjective test results, it can be seen that FPSNR has better correlation with user's experience to omnidirectional images than PSNR.**

***Index Terms*—virtual reality, omnidirectional image, quality metric, quality model**

I. Introduction

Virtual Reality (VR) is a technology that has been gaining popularity in the last few years and is capable of applying in many fields such as education, tourism, military [2–4]. However, hardware resources (CPU/GPU) and network bandwidth required to transmit omnidirectional images to viewers are too high. There are many studies on optimizing the use of resources while still ensuring the quality of the experience for users [5] [6]. However, quality metrics that measure the quality of omnidirectional images are still limited.

Most recently studies use PSNR as a quality metric to evaluate the quality of omnidirectional images [7] [8]. Besides, there are some proposed PSNR variants like weighted to spherically uniform PSNR (WS-PSNR) [9], spherical PSNR without interpolation (S-PSNR-NN) [10], spherical PSNR with interpolation (S-PSNR-I) [11], and PSNR in Crasters Parabolic Projection (CPP-PSNR) [12]. A comparison between eight quality metrics has been conducted in [13]. The finding is that PSNR is the most appropriate metric for evaluating omnidirectional images.

Recently, eye tracking technology has been integrated into VR headset. It allows a system to detect the position where viewer's eyes focus on. Therefore, a new technique (called foveated rendering) which is based on an important biological property of the human eye has been developed to optimize resource usage by reducing resolution of the area where viewer's eyes do not focus on [14]. Specifically, there are two types of photoreceptor cells in the human eyes: rods and cones. Rods are responsible for vision at low-light conditions, while cones are active at higher light levels and they can capture fast motions, colors, details of acuity. The fovea is a circular region with diameter of 1.2 mm at the center of retina [15].

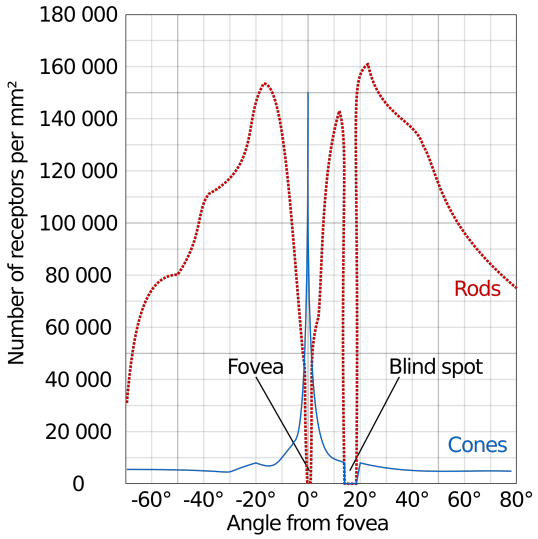

Fig. 1. The distribution of rod and cone photorceptors across the human retina [1].

The straight line from the pupil to the fovea and the other regions in the retina make an angle called retinal eccentricity [16]. This line can be adjusted to intersect to any point in sight due to the eye's movement. This intersection point is called the fixation point, where the viewer's visual attention is fixed. As we can see in Fig. 1, the fovea region (where eccentricity is close to zero degrees) is rod-free and has a very high density of cones or ganglion cells. So the fovea allows the human eyes to capture high-resolution images when a person is gazing at somewhere. However, when the retinal eccentricity increases (at about 2°), i.e. away from the fixation point, the brain captures less details (or has lower resolution), which means the vision is blurred.

There are some studies about quality model based on human perception for flat images [17] [18]. However, to our knowledge, there is no study on quality model based on human perception for omnidirectional images. It should be noted that an omnidirectional image is mapped into a virtual sphere with the center is human eye in a virtual environment. And in fact, only a small part of the omnidirectional image (called viewport) is watched by viewers. For flat images, the positions of pixels are determined by using rectangular coordinate system. Meanwhile, for omnidirectional images, spherical coordinate system is used to compute the positions

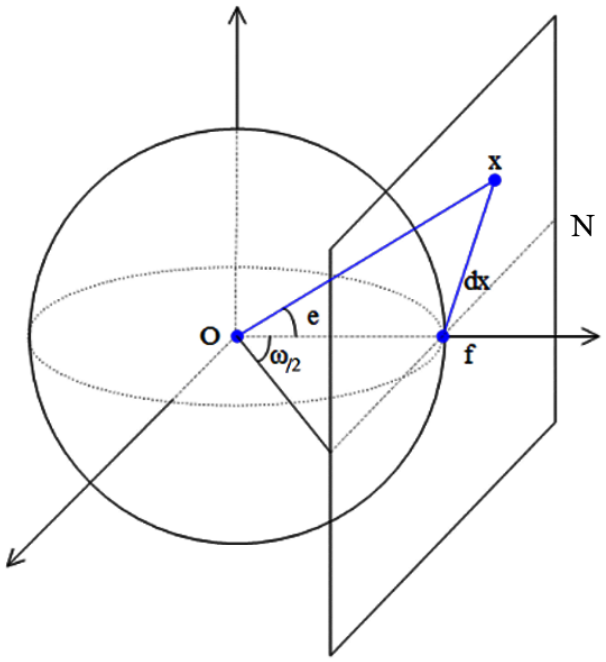

Fig. 2. Foveated Visual Sensitivity Model for a viewport of omnidirectional image.

of pixels. Therefore, the models proposed for flat images can not be used for omnidirectional images.

In this paper, we firstly propose a new quality model for omnidirectional images based on the human eyes' theory above. Then, a perception-based quality metric called FPSNR (foveated peak signal to ratio) for omnidirectional images is built. A subjective test is conducted to evaluate the performance of FPSNR. Based on the evaluation results, it is found that the performance of FPSNR is higher than that of PSNR.

The remainder of the paper is organized as follows. In Section 2, our proposed model and quality metric are presented. Section 3 describes the details of the subjective experiment. Section 4 discusses the obtained subjective quality scores and relationship between quality metrics. Finally, Section 5 concludes the paper and provides an outlook on future work.

II. Proposed Model and Quality Metric

A. Foveated Visual Sensitivity Model

Psychological experiments have been conducted to measure the contrast sensitivity of human visual system as a function of retinal eccentricity [19–21]. In [22], a model that fits the experimental data is given by

$$CT(f,e) = CT_0 \exp\left(\alpha f \frac{e+e_2}{e_2}\right), \tag{1}$$

where f is the spatial frequency (cycles/degree), e is the retinal eccentricity (degrees), CT_0 is a constant minimal contrast threshold, α is the spatial frequency decay constant, e_2 is the half-resolution eccentricity, and $CT(f,e)$ is the visible contrast threshold. Following [22], for calculating $CT(f,e)$, we set α = 0.106, e_2 = 2.3 and CT_0 = 1/64. For a given eccentricity e, Eqs. (1) can be used to find its cut off frequency f_c in the sense that any higher frequency component beyond it is imperceivable. Here, f_c can be calculated by setting $CT(f,e)$ to 1 (the maximum possible contrast) and solving for f:

$$f_c = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha(e_{(v,x)}+e_2)}\left(\frac{cycles}{degree}\right). \tag{2}$$

Assume that focus point is at the center of the image. The eccentricity is given by

$$e(v,x) = tan^{-1}\left(\frac{d_x}{vN}\right)(degree), \tag{3}$$

where d_x is Euclidean from point x to fovea point (pixel), N is image size (pixel), and v is the ratio of distance from eyes to image per image size.

Follow Fig. 2, with omnidirectional image, we can replace v value by viewport's field of view (FOV) angle. The eccentricity is changed to

$$e(x) = tan^{-1}\left(\frac{2d_x \tan(\omega/2)}{N}\right)(degree), \tag{4}$$

where ω is viewport's FOV angle, and N is viewport's size (pixel). The angle width of a pixel at position x is

$$\beta_x = tan^{-1}\left(\frac{2(d_x+0.5)\tan(\omega/2)}{N}\right) - tan^{-1}\left(\frac{2(d_x-0.5)\tan(\omega/2)}{N}\right)\left(\frac{degree}{pixel}\right). \tag{5}$$

And display resolution at position x is

$$r_x = \frac{1}{\beta_x}\left(\frac{pixels}{degree}\right). \tag{6}$$

In [23], according to the Nyquist-Shannon sampling theorem, the highest spatial frequency f_d that can be represented without aliasing by the display is half of the display resolution:

$$f_d = \frac{r_x}{2} = \frac{1}{2\beta_x}\left(\frac{cycles}{degree}\right). \tag{7}$$

Finally, the local frequency of foveated visual sensitivity model for omnidirectional images is

$$f_{pn}(x) = min\left(\frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha(e_2+e_x)}, f_d\right) \times \beta_x \left(\frac{cycles}{pixel}\right). \tag{8}$$

B. Quality Metrics

It is well-known that PSNR is a popular objective quality metric used for both images and videos [24]. Let $v(x_n)$ and $g(x_n)$ be respectively the values of pixel x_n in the original and distorted images, MAX_I is the maximum possible pixel value of the image. The PSNR metric is given by

$$PSNR = 10\log_{10}\frac{MAX_I^2}{\sum_{n=1}^{N}[v(x_n)-g(x_n)]^2}(dB). \tag{9}$$

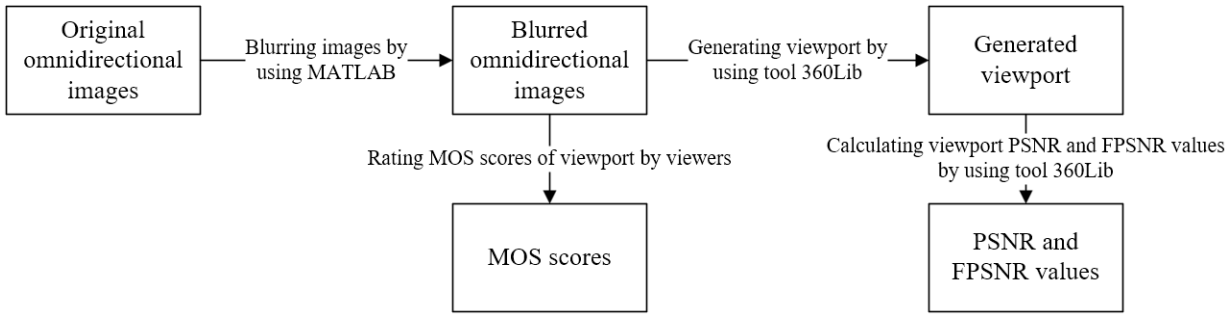

Fig. 3. Experiment process.

TABLE I
FEATURES OF SOURCE IMAGES

Image	Description
Image #1	day, low spatial complexity, a wheat field on a sunny day
Image #2	day, high spatial complexity, a market in Japan
Image #3	night, low spatial complexity, a night view of the harbor
Image #4	night, high spatial complexity, an event at Times Square in New York City

Follow the proposed model for omnidirectional images, local frequency f_{pn} can be used as a weight of pixels. So the new quality metric of FPSNR is given by

$$FPSNR = 10\log_{10}\frac{\left(\sum_{n=1}^{N} f_{pn}^2\right) MAX_I^2}{\sum_{n=1}^{N}[v(x_n) - g(x_n)]^2 f_{pn}^2}(dB). \quad (10)$$

From Eqs. (9) and (10), we can see that PSNR treats pixels on the image equally. Meanwhile, with FPSNR, the closer pixels are to focus point, the higher weights become.

III. Experiment Description

For the experiment, four omnidirectional images with different conditions of photo time and spatial complexity are chosen. The characteristics of these images are shown in Table I. The resolution of these images is 3840x1920. For each image, we select a specific viewport that is used to calculate quality metric values. Note that, in the subjective test, viewers only watch and score these specific viewports.

Fig. 3 shows the experiment process conducted in this study. Firstly, an original omnidirectional image has been partly blurred by using MATLAB. Specifically, we divide the targeted viewport area to three zones namely center zone, transition zone, and outside zone by two circles whose centers are the targeted viewport's center as shown in Fig. 4. There are two types of blurred images used in this experiment including 1) blurring in center zone and 2) blurring in outside zone. Each type has two settings of the circle angular diameters as listed in Table II. Center zone or outside zone have been blurred by using Gassian blur filters with the standard deviation of 50 and the five filter sizes of 5, 10, 20, 40, and 60 [26]. With transition zone, we use a quadratic function for masking to make the blur smoothly. Totally, there are 80 blurred images (i.e., 4x2x2x5) generated from the four original images.

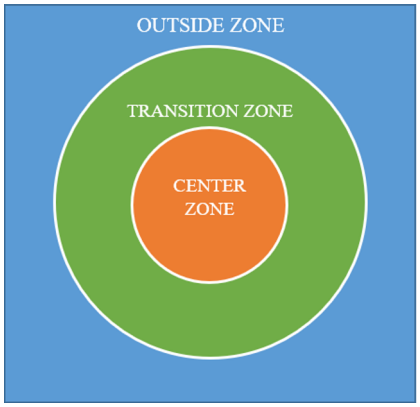

Fig. 4. Three zones in a viewport of experiment images

Secondly, the blurred images are scored by viewers. Specifically, each of the 80 images is randomly displayed during the experiment by a device set consisting of a Samsung Galaxy S6 smartphone and a Samsung Gear VR headset. The field of view of Samsung Gear VR is limited to 90°x90° by using our self-developed software [27]. This software is built based on WebVR [28], and can be opened by Samsung Internet VR. The Samsung Galaxy S6 has the screen resolution of 2880x1440 and the display size of 5.1 inches. For each image, the corresponding targeted viewport is static during the rating process, and the center of the targeted viewport is marked by a small red dot. Viewers have to keep focus on the center of viewport, which is considered as focus point to measure FPSNR, then gives a rating score for each image with the score ranging from 1 (bad) to 5 (excellent). Before doing actual subjective tests, the viewers are trained to get accustomed to the devices and the rating procedure. During the experiment, every 20 minutes, there is a break for the viewers. There are totally 18 people taking part in this experiment. The participants have ages between 20 and 37 with an average age of 23. The Absolute Category Rating method is used in our experiment [29]. A screening analysis of the subjective test results is performed following [29] and no subject is rejected.

Finally, to evaluate the performance of FPSNR, a comparison of the correlations between two objective quality metrics of PSNR and FPSNR and a subjective quality metric of

TABLE II
EXPERIMENT IMAGES'S SETTINGS

Setting	Outside-blurred image		Setting	Inside-blurred image	
	Angular diameter of inside circle	*Angular diameter of outside circle*		*Angular diameter of inside circle*	*Angular diameter of outside circle*
Setting #1	10 °	20 °	Setting #3	5 °	10 °
Setting #2	30 °	60 °	Setting #4	5 °	60 °

TABLE III
CORRELATION COEFFICIENTS OF PSNR AND FPSNR

Metric	Image #1		Image #2		Image #3		Image #4		All Image	
	PCC	*RMSE*	*PCC*	*RMSE*	*PCC*	*RMSE*	*PCC*	*RMSE*	*PCC*	*RMSE*
PSNR	0.85	0.48	0.76	0.60	0.85	0.46	0.74	0.60	0.69	0.64
FPSNR	0.93	0.41	0.91	0.39	0.90	0.39	0.88	0.42	0.73	0.60

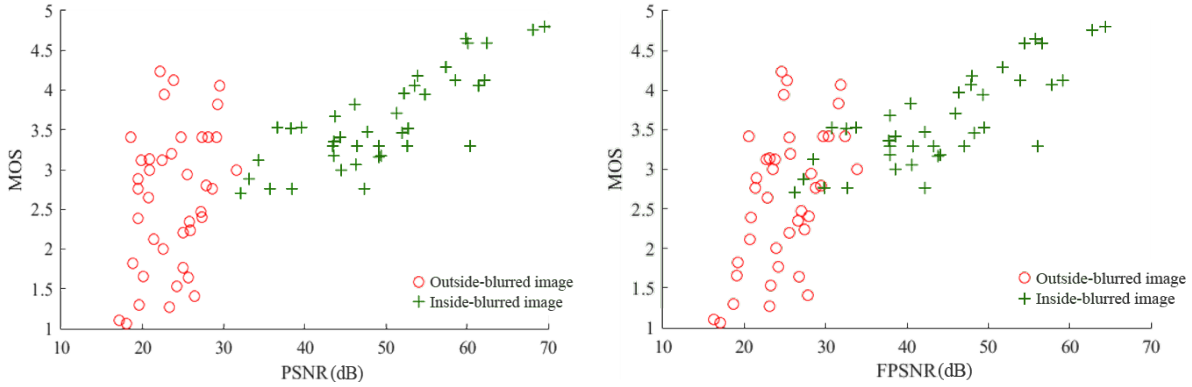

Fig. 5. PSNR and FPSNR with different blur position

MOS is conducted. In particular, the 360Lib tool [25] is employed to generate a targeted viewport for each blurred omnidirectional images, and then compute the corresponding FPSNR and PSNR values. Note that to calculate FPSNR, we consider that focus point is in the center of each image's viewport. Each viewport's resolution is 960x960 with field of view (FOV) is 90°x90°. The 360Lib tool is modified by us to integrate FPSNR. Then, based on the subjective test results, the relationships between two objective quality metrics of PSNR and FPSNR and MOS are analyzed.

IV. RESULT ANALYSIS

Fig. 5 shows the relationships between two objective quality metrics PSNR and FPSNR and MOS. Each marker shows one blurred image type. It can be seen that, the relationships between two objective quality metrics and MOS can be modeled by logistic functions. Therefore, a logistic function of the form $f(x) = d + \frac{a-d}{1+(x/c)^b}$ is used to fit the measurement results.

We use the correlation coefficients including Pearson Correlation Coefficient (PCC) and Root Mean Square Error (RMSE) to quantify how well the objective quality metric and MOS correlate are presented in Table III. Although the four original images used in this study are different in some characteristics, the behaviors of these coefficients in all four images are consistent. Specifically, the correlation coefficients are high when fitting for individual images (0.73 - 0.85 for PSNR, 0.87 - 0.93 for FPSNR). Meanwhile, when fitting for all images, the correlation coefficients are low (0.69 for PSNR, 0.73 for FPSNR). So, the considered objective quality metrics are only used to compare image quality of the same content, but not across different content types. This observation is in line with conclusions in [30] [31].

For PSNR, the images with low complexity (Image #1 and #3) have higher PCC values (0.85) and lower RMSE values (0.46 - 0.48) than that of the images having high complexity (Image #2 and #4) (0.74 - 0.76 with PCC and 0.60 with RMSE). But with FPSNR, PCC and RMSE values do not change too much across different complexity levels. Low complexity images have PCC values (0.90 - 0.93) and RMSE values (0.39 - 0.41) not too much different than that of high complexity images (0.88 - 0.91 with PCC, 0.39 - 0.42 with RMSE). It seems that FPSNR doesn't depend too much on the complexity level of content. Note that with the images captured at daytime (Image #1 and #2), the correlation coefficients of FPNSR are slightly higher than that of the images captured at night (Image #3 and #4). This is due to photo time, Image #1 and #2 are daytime shots, so the luminance value of pixels over the images are very close, while Image #3 and #4 are taken at night with the light

from light bulbs, making big differences in luminance value among image zones. Blurring the image makes huge change in luminance value over the continuous area between light and dark, in which the human eyes are very sensitive with. Therefore, viewers feel uncomfortable with this, so it's harder to rate night images than daytime images.

From Fig. 5, we can see that most of PSNR values are higher than FPSNR in the inside-blurred image, and FPSNR values are higher than PSNR in the outside-blurred images. This is because with FPSNR, pixels on the position where viewer's visual attention is fixed are weighted higher than other pixels around. We can tell that with PSNR, inside-blurred image are overrated and outside-blurred images are underrated, compared with viewer's experience.

In any case of the experiment, the objective quality metric FPSNR has higher PPC value and lower RMSE value than PSNR. It means that FPSNR is closer to user's perception than PSNR.

V. Conclusions

In this paper, we have have proposed a foveated visual sensitivity model and an objective quality metrics for omnidirectional images, called FPSNR. Through experimental results, it was shown that thanks to taking into account perception information, FPSNR is closer to user's experience than PSNR. For future work, we will apply our proposed model to some objective quality metrics such as SSIM and MSSSIM, then evaluate and compare their performance.

References

[1] "File:Human photoreceptor distribution.svg," [Online]. Available: https://commons.wikimedia.org/wiki/File:Human_photoreceptor_distribution.svg. [Accessed: 7 Mar. 2018].

[2] Forbes.com, "Virtual Reality: The Next Generation Of Education, Learning and Training". [online] Available at: https://www.forbes.com/sites/forbesagencycouncil/2017/12/13/virtual-reality-the-next-generation-of-education-learning-and-training/#4bb6bed8733f [Accessed 4 Mar. 2018].

[3] BBC News, "Virtual reality: Tourism firms use VR to attract visitors". [online] Available at: http://www.bbc.com/news/uk-wales-41635746 [Accessed 4 Mar. 2018].

[4] Stone, V, "How Virtual Reality Is Changing Military Training". [online] Samsung Business Insights. Available at: https://insights.samsung.com/2017/07/13/how-virtual-reality-is-changing-military-training/ [Accessed 4 Mar. 2018].

[5] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "A New Adaptation Approach for Viewport-adaptive 360-degree Video Streaming," 2017 IEEE International Symposium on Multimedia (ISM), 2017.

[6] C. Zhou, Z. Li, and Y. Liu, "A Measurement Study of Oculus 360 Degree Video Streaming," Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys17, 2017.

[7] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," 2017 IEEE International Conference on Communications (ICC), 2017.

[8] C. Ozcinar, A. D. Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," 2017 IEEE International Conference on Image Processing (ICIP), 2017.

[9] Y. Sun, A. Lu, and L. Yu, "AHG8: WS-PSNR for 360 video objective quality evaluation," Joint Video Exploration Team of ITUT SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0040, 4th Meeting, Chengdu, 2016.

[10] Y. He, B. Vishwanath, X. Xiu, and Y. Ye, "AHG8: InterDigitals projection format conversion tool," Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0021, 4th Meeting, Chengdu, 2016.

[11] M. Yu, H. Lakshman, and B. Girod, "A Framework to Evaluate Omnidirectional Video Coding Schemes," 2015 IEEE International Symposium on Mixed and Augmented Reality, Fukuoka, pp.3136, Sept. 2015.

[12] V. Zakharchenko, E. Alshina, A. Singh, and A. Dsouza, "AHG8: Suggested testing procedure for 360-degree video," Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0027, Chengdu, 2016.

[13] Tran, H., T. Pham, C., Pham Ngoc, N., T. Pham, A. and Thang, T, "A Study on Quality Metrics for 360 Video Communications," IEICE Transactions on Information and Systems, E101.D(1), pp.28-36.

[14] "Understanding Foveated Rendering," Sensics, 28-Aug-2016. [Online]. Available at : http://sensics.com/understanding-foveated-rendering/. [Accessed: 04 Mar. 2018].

[15] D. Purves, "Anatomical Distribution of Rods and Cones," Neuroscience. 2nd edition., 1AD. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK10848/. [Accessed: 4 Mar. 2018].

[16] Mahsa Mohammadkhani, "Foveation-Based Video Coding". Available from https://era.library.ualberta.ca/files/08612p44v [accessed 14 Jan. 2018].

[17] Andre Luiz N. Targino Da Costa and M. N. Do, "A Retina-Based Perceptually Lossless Limit and a Gaussian Foveation Scheme With Loss Control," IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 3, pp. 438453, 2014.

[18] Sanghoon Lee, Alan C. Bovik and Brian L. Evans, "Efficient Implementation of Foveation Filtering"

[19] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual fleld," Vision Research, vol. 21, pp. 409-418, 1981.

[20] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling," Journal of the Optical Society of America, vol. 8, pp. 1775-1787, 1991.

[21] T. L. Arnow and W. S. Geisler, "Visual detection following retinal damage: Prediction of an inhomogeneous retino-cortical model," in Human Vision and Electronic Imaging, Proc. SPIE, vol. 2674, pp. 119-130, 1996.

[22] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in Proc. SPIE, vol. 3299, pp. 294-305, July 1998.

[23] Wang, Zhou. "Rate Scalable Foveated Image and Video Communications". 2001.

[24] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 660668, 2008.

[25] Joint Video Exploration Team, "360Lib," [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-2.0.1/.

[26] "2-D Gaussian filtering of images - MATLAB imgaussfilt" [Online]. Available: www.mathworks.com/help/images/ref/imgaussfilt.html.

[27] "phunm211/WebVR_image: 360 VR Image viewer based on WebVR" [Online]. Available: https://github.com/phunm211/WebVR_image.

[28] "WebVR," Bringing Virtual Reality to the Web. [Online]. Available: https://webvr.info/. [Accessed: 04 Mar. 2018].

[29] P.913, Recommendation ITU-T, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," 2014.

[30] Huynh-Thu, Q., and M. Ghanbari. "Scope of validity of PSNR in image/Video quality assessment," Electronics Letters, vol. 44, no. 13, 2008, p. 800.

[31] H. T. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham and T. C. Thang, "An evaluation of quality metrics for 360 videos," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, 2017, pp. 7-11.