Modeling Peripheral Vision Impact on Perceptual Quality of Immersive Images

Peiyao Guo, Qiu Shen, Mingkai Huang, Rongbing Zhou, Xun Cao, Zhan Ma
School of Electronic Science and Engineering, Nanjing University, Jiangsu, China
Emails: peiyao@smail.nju.edu.cn, {shenqiu, mazhan}@nju.edu.cn

***Abstract*—Conventional images/videos are often rendered within the central area of human visual system (HVS) with uniform quality. Recent virtual reality (VR) device with head mounted display (HMD) extends the field of view (FoV) significantly to include both central and peripheral areas. It exhibits the unequal acuity of the quality sensation because of the non-uniform distribution of photoreceptors in our retina. Hence, we propose to study the impact of image qualities (with respect to the quantization stepsize q or spatial resolution s) in peripheral vision and conclude self-adaptive analytical models that have shown quite impressive accuracy through independent cross validations. These models can further be applied to assign different quality weights at different regions, so as to significantly reduce the transmission data size but without subjective quality loss.**

***Index Terms*—Peripheral vision, image quality, quantization stepsize, spatial resolution**

I. Introduction

Recent advances in display and computing technologies make VR easily accessible to the common users in addition to the trained lab specialists. Top tier companies are investing billions of dollars to further improve the technology to enable more applications for massive market adoption, including Facebook/Oculus, HTC, Google, etc.

A successful VR device should be able to offer the ultra high definition (UHD) scene (i.e., equal or even surpass the resolution of the HVS) and unperceived interaction latency with high frame rate (i.e., faster than neuron stimuli system) through reliable wired or wireless transmission. However, UHD scene at high frame rate requires an enormous network bandwidth for a stable connection. For instance, Netflix suggests 50 Mbps downlink to stably receive its 4K content (encoded around 15 Mpbs) [1]. This is apparently a big challenge for the network infrastructure around the world.

When subjects are viewing the immersive (panoramic) images with HMD, only the content in front of current FoV (Fig. 1) is displayed, but all information inside and outside current FoV have to be delivered in existing systems [2]. On the other hand, even for current FoV, HVS has very different perceptual sensation in different areas according to the existing research works [3] on vision and neuroscience. For instance, macular area, which is about 9° eccentrically (i.e., from the center of our retina), often requires ultra high resolution. While visual perception would have very sharp degradation in our periphery. Ideally, such non-uniform quality sensation of HVS could be leveraged to save the bandwidth consumption by assigning high quality content (at high bit rate) to the content blocks in central area but degraded quality (at much lower bit rate) in peripheral area. Thus, this requires a quantitative model that could explicitly express the peripheral vision impact on image quality shown in HMD.

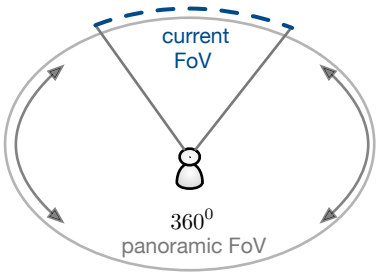


Fig. 1. Illustration of the current FoV and panoramic FoV

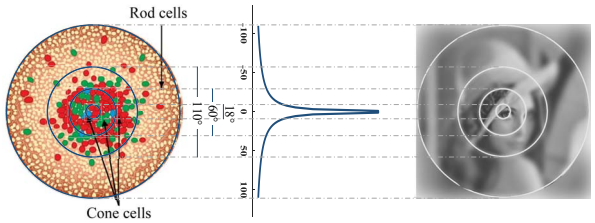


Fig. 2. Distribution of photoreceptors on the human retina and corresponding non-uniform visual perception

We propose to measure the image quality degradation in peripheral area with respect to the degree of eccentricity (θ) from the center of the retina. Here, We use the horizontal eccentricity to simplify the illustration of the 2-D viewing range of a FoV. Note that image quality can be controlled by q[1] or s, so the problem can be rephrased to devise appropriate mathematical forms to describe the variations of q and s, with respect to the θ. Towards this goal, we have invited thirty subjects with normal vision to participate the subjective quality assessment for peripheral vision impact study. Both q and s can be well modeled using a parametric generalized Gaussian model in terms of the θ. Parameters are either fixed or can be estimated using the content features. Proposed models are then utilized to generate images with non-uniform quality (using different q or s in peripheral area) for independent

[1]Images are encoded with the quantization stepsize q via H.264/AVC intra codec. $q = 2^{\frac{QP-4}{6}}$ [4], in which QP represents the quantization parameter.

978-1-5386-0462-5/17/$31.00 ©2017 IEEE.

cross-validations, which demonstrate the effective results by applying the unequal quality to the content blocks in the central and peripheral regions of an immersive image shown in HMD. Thus, these models can be devised in VR streaming to further reduce the transmission bandwidth without noticeable quality degradation.

The reminder parts of this work are organized as follows: Section II explains the details of how to measure the peripheral vision impact on an immersive image shown in HMD, and propose analytical models to quantify the q and s with respect to θ. The proposed peripheral vision model is cross-validated in Section III. Finally, conclusion is drawn in Section IV.

II. Peripheral Vision Model

It is known that human eye has different spatial resolution distinguishability between central and peripheral vision area because of the highly non-uniform distribution of photoreceptors on the human retina [6], shown in Fig. 2. Tyler [7] has proposed a power function to quantify the density of cones (e.g., measured by the number of cones per mm^2) from 1° to 20° eccentrically, i.e.,

$$\rho(\theta) = 50000 \cdot \left(\frac{\theta}{300}\right)^{-\frac{2}{3}}, \quad \theta \in [1^\circ, 20^\circ], \tag{1}$$

while $\rho(\theta)$ is linearly decreased until 4000 cones/mm^2 with $\theta > 20^\circ$. Mathematically, Tyler's $\rho(\theta)$ can be seen as an approximation of the generalized Gaussian distribution.

Intuitively, users have sensitive perception in the area with higher density cones, but may not tell the difference of image degradation in the area with less cones. We hypothesize that the ability to distinguish the image perceptual quality variation should follow the density distribution of the photoreceptors $\rho(\theta)$ in our retina. As image quality can be represented by its signal fidelity (that is often controlled using q or s) [8], the overall problem is to model q or s with respect to the degree of eccentricity θ without noticing the image quality degradation perceptually. Towards this goal, we have carefully designed the subjective tests to collect users' opinion scores so as to develop the analytical models.

A. Subjective Experimental Setup

Fourteen immersive images from the SUN360 database [5] downsampled to spatial resolution at 4096×2160 are chosen as our test materials, as shown in Fig. 3. These images cover wide range of spatial complexity (expressed by spatial information index (SI) [9]), and their saliency regions locate in the central region of users' FoV. Eight of them are selected to model the peripheral vision impact on perceptual quality, and the other six images are picked up for cross validation (marked with star). Immersive images are rendered using the HTC Vive system with its HMD. Participants are restricted to stay steady by focusing on a green cross overlaid in the FoV center without head and body movement (and possible eye movement) when performing the tests. This is to avoid viewing noise when the user randomly shifts their attention in a large area. Since VR is more prevalent to the young people, we recruit thirty students (aging from 20 to 25) from Nanjing University to participate this assessment. All viewers have normal vision (or after correction) and color perception.

We perform the central vision impact study first with uniform quality level for each test sample. We switch the test samples from its highest quality to the lowest one step by step, and ask the subjects whether they can tell the difference. If positive, we record the associated q or s respectively, which are defined as q_c and s_c.

Given that HTC HMD offers a 110°-wide FoV and peripheral area actually consists of the near and far ranges [3], current FoV is divided into three regions as shown in Fig. 4, i.e., central region with θ from 0° to 9° in one side, near peripheral region with $\theta \in (9^\circ, 30^\circ]$ and the rest from 30° to 55° for far peripheral region. The regions outside of current FoV are not discussed in this work. During the peripheral vision impact study, we fix the content quality at central region of each test sample via using the q_c or s_c, and degrade the quality of near and far peripheral region together until the subject notices the distortion. We record the corresponding q or s, noted as q_{p_n} or s_{p_n}, respectively. Then, we fix the quality of near peripheral region with the q_{p_n} or s_{p_n}, and continue to degrade the quality of the far peripheral region separately till subjects feel the difference perceptually. Similarly, associated q_{p_f} or s_{p_f} are marked. Each test sample is displayed for about 3 seconds with 1 second in between. There is a 1-minute interval for subjects to rest their eyes between two different images.

B. Analytical Models

q and s are normalized using $q_{\max}$ = 228, and $s_{\max}$ = 4096x2160, repsectively, and the averaged $\hat{q}$ and $\hat{s}$ in central and peripheral areas with respect to the θ are plotted in Fig. 5. It is shown that the $\hat{q}$ is increasing but the $\hat{s}$ is decreasing with the increase of θ. Since each region is divided with respect to θ, we then could derive analytical $\hat{q}$ or $\hat{s}$ in terms of the θ. We make the hypothesis that the ability of image quality differentiation should follow the density distribution of photoreceptors $\rho(\theta)$. Therefore, we propose to use an unified parametric generalized Gaussian function to model $\hat{q}$ and $\hat{s}$ as follows:

$$\frac{1}{c\sqrt{2\pi}} \times e^{-\frac{|(b\theta)^a|}{2c^2}} + d, \tag{2}$$

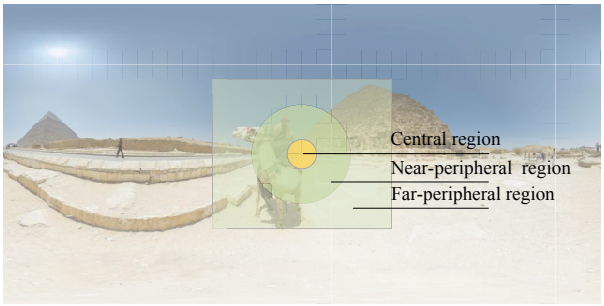
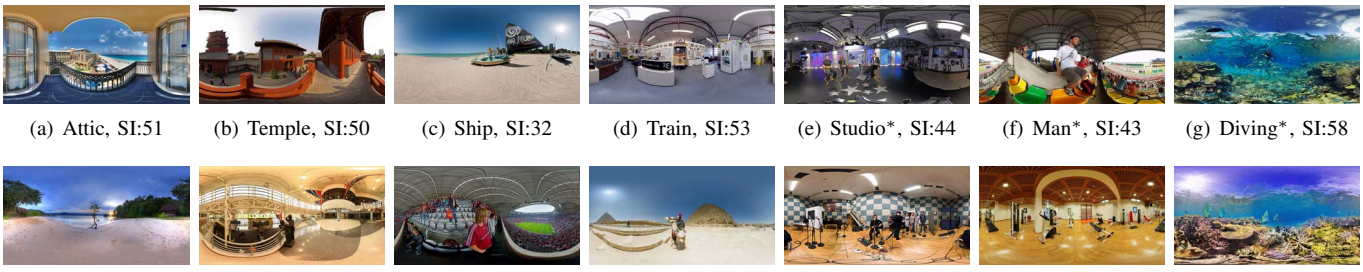


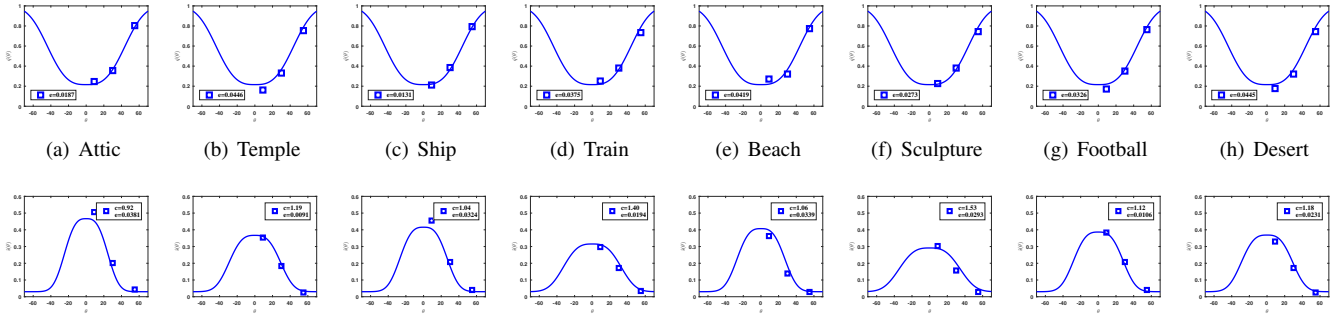
Fig. 4. Illustration of various perceptual regions of a FoV



(a) Attic, SI:51 (b) Temple, SI:50 (c) Ship, SI:32 (d) Train, SI:53 (e) Studio*, SI:44 (f) Man*, SI:43 (g) Diving*, SI:58

(h) Beach, SI:36 (i) Sculpture, SI:62 (j) Football, SI:74 (k) Desert, SI:38 (l) Singer*, SI:57 (m) Gym*, SI:46 (n) Fish*, SI:88

Fig. 3. Immersive images used for modeling the peripheral vision and cross validation (marked with star) [5]



(a) Attic (b) Temple (c) Ship (d) Train (e) Beach (f) Sculpture (g) Football (h) Desert

(i) Attic (j) Temple (k) Ship (l) Train (m) Beach (n) Sculpture (o) Football (p) Desert

Fig. 5. Normalized data responding to different images, (a) - (h) for $\hat{q}(\theta)$; (i) - (p) for $\hat{s}(\theta)$. e represents the root mean square error (RMSE).

where a, b, c, d are control parameters derived using the least-squared criteria by fitting the measured points and hypothesized analytical model in Eq. 2. Fitted paramters are shown in Table I. Note that parameters are different for q and s due to the reason that adapting the q infers the high frequency information loss of compression while varying the s implies the loss of some low frequency content.

Parameter a is identical as it reflects the decay speed of the quality perception with respect to the increase of the θ. This is mainly because it is determined by the density of the cones of human retina. b is correlated with the quality degradation factors. As aforementioned, q would introduce compress-induced high frequency information loss while s brings the loss of some low frequency content due to the spatial upsampling. Parameter c is content-dependent. But for $\hat{q}$, we can still use the fixed c for all images. It is suspected that current Vive display does not offer sufficient pixel density (e.g., pixel per inch or PPI) to accurately reflect the compression-induced amplitude variations per pixel. But for spatial upsampling, pixels are restored with predefined filers, which significantly differ from the original pixel in native 4096x2160 resolution. Moreover, d indicates the model limits that must fit our intuition. For example, we could not perceive any difference when θ goes to infinity (i.e., the number of the photoreceptor goes to zero). For q, it means that we could apply the $q_{\max}$, implying d = 1; but for s, it is impractical to have s = 0, thus we set d = 0.03 with the least model prediction error.

TABLE I
PARAMETERS IN PERIPHERAL VISION MODEL FOR $\hat{q}$ OR $\hat{s}$

	a	b	c	d
$\hat{q}$	3	0.016	-0.51	1
$\hat{s}$	3	0.042	c(x)	0.03

Following the previous discussion, we introduce how to predict parameter c for model $\hat{s}(\theta)$. Intuitively, image quality is mainly determined by its spatial complexity, color distribution, and local orientation. Through careful examination, we have found that c could be predicted by the linear combination of the $\rho_{c_{\mathrm{SI}}}$, ρ_{μ_I} and $\rho_{\mu_{\gamma_v}}$, i.e.,

$$c = -0.004 \cdot \rho_{c_{\mathrm{SI}}} + 1.894 \cdot \rho_{\mu_I} + 13.869 \cdot \rho_{\mu_{\gamma_v}} - 0.849 \quad (3)$$

where $\rho_{c_{\mathrm{SI}}}$ is the SI in the central region of an image. (This is because we constrain the saliency region in the central vision.) ρ_{μ_I} is the averaged intensity of the image in HSI color space [10]. $\rho_{\mu_{\gamma_v}}$ refers to the intensity of the vertical orientation which is calculated using a 3×3 Gabor filter [11].

III. Cross Validation

To ensure that our model is generally applicable, we invite another sixteen subjects to participate independent cross validation assessment. Six images (shown in Fig. 3) are selected, and each of them is processed into two copies: one with uniform quality using q_c or s_c, and the other one with non-uniform quality where q or s are assigned according to our models in Sec. II. For non-uniform quality, the picture is divided into seven parts to approximate smooth degradation

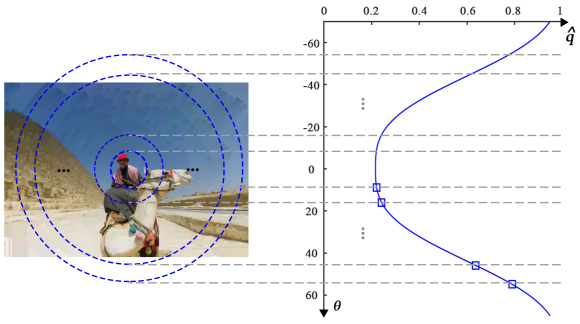
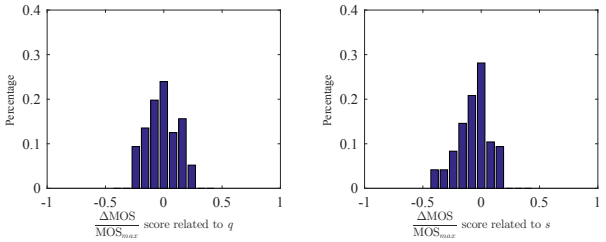


Fig. 6. An image with non-uniform quality with q calculated via the peripheral vision model (2)



(a) (b)

Fig. 7. The distribution of deltaMOS of the content using the uniform quality and non-uniform quality; non-uniform quality is applied using the peripheral vision model (2).

as shown in Fig. 6. Note that except those fixed parameters, content features are extracted from the images to derive the corresponding c and model itself explicitly.

Each image repeats 3 times and is arranged in random order. Subjects give their mean opinion score (MOS) from 1 to 5 (where 1 represents the worst and 5 represents the best) for each displayed image successively. We analyze the mean MOS of the image with uniform quality as well as corresponding non-uniform quality. When the absolute value of the difference is less than 1, we think the model works well. The distributions of the delta of MOS (deltaMOS) are shown in Fig. 7. It can be observed from the figure that almost all of deltaMOS are located between -1 and 1, especially some of them are equal to 0. This means most of the subjects can not tell the difference between uniform and non-uniform compression. Although the others catch some difference, they are not always in favor of uniform compression, which can be considered as subjective noise.

Based on these results, it is proved that the non-uniform perception of human eyes can actually be modeled mathematically and utilized to guide the image compression. Furthermore, with parameters that can be well estimated using the content features, our models are self-adaptive and generally applicable.

IV. Conclusion

Non-uniform distribution of photoreceptors on human retina infers that visual perception on natural image is less sensitive in peripheral area than in central area. We study such peripheral vision impact on perceptual quality of immersive images rendered in VR HMD, and reach close-form theoretical models that explicitly describe the control factors (i.e., quantization stepsize and spatial resolution) of image quality with respect to the degree of the eccentricity in peripheral vision. Models are well explained by an unified parametric generalized Gaussian function where parameters are either fixed or can be well estimated by the features extracted from the native content. We randomly select another set of images, extract their features and form the models to guide the non-uniform quality assignment in different regions. Independent cross validation is conducted to demonstrate that users could not tell the difference between uniform and non-uniform quality assignments in peripheral vision. This evident the efficiency of our proposed models.

As the future work, we will extend this static spatial vision study to consider the temporal variation that is often happened when users navigate the content inside the VR environment. Note that impacts of quantization and spatial resolution are examined separately in this work. Their joint impact would be another interesting topic.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) projects (Grant #s: 61371166, 61422107 and 61571215), in part by the National Science Foundation for Young Scholar of Jiangsu Province, China (Grant # BK20140610). Dr. Z. Ma is the corresponding author of this paper.

References

[1] [Online]. Available: http://bgr.com/2013/09/26/netflix-4k-streaming/
[2] X. Corbillon, A. Devlic, G. Simon, and J. Chakareski, "Viewport-Adaptive Navigable 360-Degree Video Delivery," *arXiv preprint arXiv:1609.08042*, 2016.
[3] H. Strasburger, I. Rentschler, and M. Juettner, "Peripheral vision and pattern recognition: A review," *Journal of Vision*, vol. 11, no. 13, pp. 1–82, May 2011.
[4] S. Ma, W. Gao, D. Zhao, and Y. Lu, "A study on the quantization scheme in h. 264/avc and its application to rate control," *Advances in Multimedia Information Processing-PCM 2004*, pp. 192–199, 2005.
[5] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba, "Recognizing scene viewpoint using panoramic place representation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2695–2702.
[6] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *Journal of comparative neurology*, vol. 292, no. 4, pp. 497–523, 1990.
[7] C. W. Tyler, "Analysis of human receptor density," in *Basic and clinical applications of vision science*. Springer, 1997, pp. 63–71.
[8] Y. Xue, Y.-F. Ou, Z. Ma, and Y. Wang, "Perceptual video quality assessment on a mobile platform considering both spatial resolution and quantization artifacts," in *Proc. of PacketVideo*, 2010.
[9] H. Yu and S. Winkler, "Image complexity and spatial information," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*. IEEE, 2013, pp. 12–17.
[10] [Online]. Available: https://en.wikipedia.org/wiki/HSL_and_HSV
[11] Wikipedia, "Gabor filter — wikipedia, the free encyclopedia," 2017. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Gabor_filter&oldid=781673347