LETTER Special Section on Enriched Multimedia — Media technologies opening up the future —

An Evaluation of the Impact of Distance on Perceptual Quality of Textured 3D Meshes

Duc NGUYEN[†a)], ***Member***, **Tran THUY HIEN**[††], **Huyen T. T. TRAN**[†††], **Truong THU HUONG**[††], ***and*** **Pham NGOC NAM**[††††], ***Nonmembers***

SUMMARY Distance-aware quality adaptation is a potential approach to reduce the resource requirement for the transmission and rendering of textured 3D meshes. In this paper, we carry out a subjective experiment to investigate the effects of the distance from the camera on the perceptual quality of textured 3D meshes. Besides, we evaluate the effectiveness of eight image-based objective quality metrics in representing the user's perceptual quality. Our study found that the perceptual quality in terms of mean opinion score increases as the distance from the camera increases. In addition, it is shown that normalized mutual information (NMI), a full-reference objective quality metric, is highly correlated with subjective scores.

key words: *textured 3D mesh, distance, quality assessment*

1. Introduction

Textured 3D mesh is a popular format for representing 3D models [1]. A textured 3D mesh is comprised of two parts, namely geometry data and texture data. The geometry data, better known as a polygon mesh [2], is a collection of vertices, edges, and faces that define the shape of a 3D model. The texture data can be simple repetitive patterns or complex images that are mapped onto the surfaces of the 3D model [3]. In order to provide a realistic 3D model, high-quality geometry and texture data are both needed. Thus, the transmission and rendering of textured 3D meshes usually require a significant amount of system resources.

A potential method to reduce the resource requirement of a textured 3D mesh is to dynamically adapt the Level of Detail (LoD) of the mesh based on the distance between the mesh and the virtual camera [4], [5]. For that, 3D meshes are encoded into multiple versions of different LoDs using methods such as geometric sampling, quantization, and smoothing [6]. Low-detail versions are chosen for meshes that are further away, and high-detail versions are chosen for those that are close. Since low-detail versions contain less amount of geometry and texture data, resources required to transmit and render a mesh can be lowered.

Manuscript received February 28, 2023.
Manuscript revised June 15, 2023.
Manuscript publicized September 25, 2023.

[†]The author is with the Faculty of Engineering, Tohoku Institute of Technology, Sendai-shi, 982–8577 Japan.
[††]The authors are with Hanoi University of Science and Technology, Vietnam.
[†††]This work was done when taking part in a project at VinUniversity, Vietnam.
[††††]The author is with VinUniversity, Vietnam.
a) E-mail: ducnguyen@tohtech.ac.jp
DOI: 10.1587/transinf.2023MUL0001

To support optimal generation and selection of LoD versions, it is important to understand how distance from the camera affects the user's perceptual quality of textured 3D meshes. In the literature, the impacts of various factors on the perceptual quality of the textured 3D meshes have been investigated such as types of distortion [6], diffuse colors [7], and light-material interaction [8]. However, to the best of our knowledge, no previous work has explored the effects of distance from the camera on the perceptual quality of textured 3D meshes.

To fill in this gap, in this paper, we conduct a study of subjective and objective quality assessment for textured 3D meshes taking into account the impact of the distance from the camera. The results of the subjective experiment show that the further the distance from the camera is, the higher the Mean Opinion Score (MOS) becomes. On average, the MOS score is increased by approximately 0.3~0.5 as the distance from the camera increases by 4 units. For some textured meshes, the differences in the perceptual quality between LoD versions become negligible at long distances. In addition, we investigate the correlation of the subjective scores with eight popular objective quality metrics and found that the NMI metric [9] has the highest correlation with the subjective scores.

The remainder of the paper is as follows. Related work is given in Sect. 2. The subjective experiment is presented in Sect. 3. The evaluation of objective quality metrics is given in Sect. 4. Finally, the paper is concluded in Sect. 5.

2. Related Work

In the literature, there are several works on quality assessment of textured 3D mesh. The authors in [10] investigate the impacts of geometry and texture resolution on the quality of textured meshes. It is found that viewers are more sensitive to the distortion of the texture than that of the geometry. The impact of the masking effects caused by compression artifacts of the texture is evaluated in [11]. In [6], a subjective study is conducted to investigate the influence of five types of distortion on both geometry and texture data (i.e., compression, simplification, smoothing, JPEG encoding, and subsampling) on the perceptual quality of the textured mesh. It is found that meshes with complex textures are very sensitive to simplification, whereas highly curved models are sensitive to smoothing. The influence of source models, animations, and viewpoints with diffuse colors is explored in [7]. In

[8], the authors study the effects of light-material interaction on the quality of textured 3D meshes. Related to objective quality assessment, a no-reference metric for colored mesh is proposed in [12]. However, no previous work has examined the influence of the distance from the camera on the perceptual quality of textured meshes.

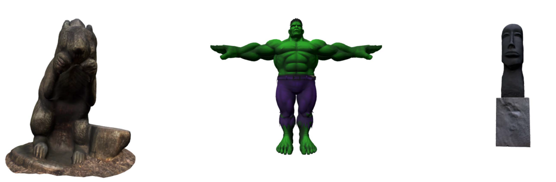
3. Subjective Experiment

In this section, we first present the settings of the subjective experiment. Then, we perform an analysis of the experiment results.

3.1 Experiment Settings

For the subjective experiment, we use three textured 3D meshes of *Hulk*, *Squirrel*, and *Statue* from a public dataset constructed in [6]. The snapshots of the three meshes are shown in Fig. 1. The *Hulk* is an artificial model created using modeling software. It has structured texture content and smooth texture seams. The *Squirrel* and *Statue* are reconstructed from multiple images of actual objects. Thus, the texture images of these meshes are noisier and contain more complex texture seams.

There are many ways to generate LoD versions of textured meshes such as simplification, quantization, or smoothing of the geometry data, and downsampling or compression of the texture images [6]. Among that, simplification is the most widely used method, being supported by popular 3D software. Thus, in this paper, we apply simplification to the geometry data to generate LoD versions of the considered meshes. Other LoD generation methods will be considered in our future work. In particular, for each of the meshes, three LoD versions are generated. The number of vertices of each LoD version of the three meshes is given in Table 1. For the distance from the camera, we consider five distance values of $d = \{4, 8, 12, 16, 20\}$. Figure 2 show the snapshots of Version 1 at different distances from the camera. Totally, there are 45 stimuli in the experiment.



(a) Squirrel (b) Hulk (c) Statue

Fig. 1 Snapshots of three textured 3D meshes.

Table 1 Number of vertices of nine LoD versions of three textured meshes.

LoD Version	Squirrel	Hulk	Statue
Version 1	18417	40497	311820
Version 2	9204	28926	78024
Version 3	4599	17355	31206

To obtain the subjective score of each stimulus, we employ the double-stimulus impairment scale (DSIS) method [13]. In particular, the subject is presented with a series of pairs of meshes. The first mesh in a pair is the original one and the second one is one of the stimuli. Each mesh is displayed for 8 seconds with two seconds of the grey screen in between. After viewing the two meshes of a pair, the subject is asked to give his/her opinion score of the quality of the second mesh on a five-grade impairment scale as follows: 5 (imperceptible), 4 (perceptible, but not annoying), 3 (slightly annoying), 2 (annoying), 1 (very annoying). The process is repeated until all the stimuli are evaluated. The stimuli are displayed randomly on a laptop screen with a resolution of 1366x768 using A-Frame [14] which is a web-based framework for AR/VR applications.

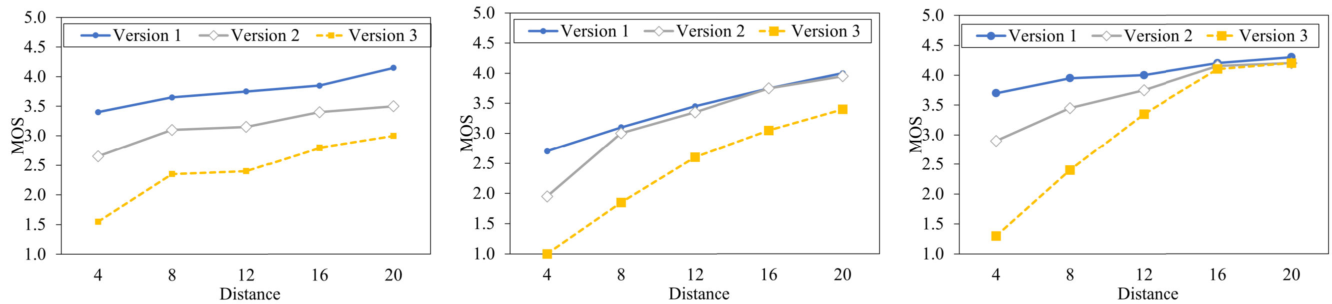
We recruit students and staff from our institutions to participate in the subjective experiment. Totally, there are 20 people took part in the experiment, aged between 20~32, all with normal or corrected normal vision. On average, it took approximately 30 minutes for one subject to complete rating all the stimuli. Screening analysis is performed on the obtained scores according to [15], and one person is rejected. The Mean Opinion Score (MOS) of each stimulus is calculated as the average score of all the valid participants.

3.2 Result Analysis

Figure 3 shows the MOSs of the LoD versions of three meshes at five distance values. It can be seen that the MOS increases as the distance increases for all three meshes. The average increase in MOS for every four distance units is 0.3 for the *Squirrel* mesh, 0.5 for the *Hulk* mesh, and 0.4 for the *Statue* mesh. These results imply that the further the distance is, the more difficult for the participants to recognize the distortion level of an LoD version. This can be explained that the further the distance is, the smaller the size of the mesh on the screen would become.

(a) $d = 4$ (b) $d = 8$ (c) $d = 12$ (d) $d = 16$ (e) $d = 20$

Fig. 2 Snapshots of Version 1 of the *Hulk* mesh at different distances.



(a) Squirrel (b) Hulk (c) Statue

Fig. 3 MOSs of the LoD versions of three meshes at five distance values.

We can also see that the impact of the distance across different LoD versions of a mesh is varying. For the *Squirrel* mesh, the difference in MOS between the three versions is quite stable across all distances. In particular, the MOSs of Version 1 are 0.45~0.75 higher than that of Version 2, and the MOSs of Version 2 are 0.5~1.1 higher than that of Version 3. This result indicates that for the *Squirrel* model, selecting a low-detail LoD version comes at a cost of reducing the perceptual quality even if the distance from the camera is relatively far.

For the *Hulk* mesh, the MOS of Version 2 is 0.75 lower than that of Version 1 at $d = 4$. But, the differences in MOS between the two versions are less than 0.1 for $d \geq 8$. These results indicate that the perceptual quality of Version 1 and Version 2 of the *Hulk* mesh is very similar for $d \geq 8$ even though the number of vertices of Version 2 is approximate 30% less than that of Version 1. With this insight, it is possible to significantly reduce the bandwidth required for the transmission of the mesh by choosing Version 2 instead of Version 1 without causing a negative impact on user experience. The difference in the MOSs between Version 3 and Version 1 also becomes smaller as the distance increases. Yet, unlike Version 2, the MOS of Version 3 at $d = 20$ is still 0.7 lower than that of Version 1.

In the case of the *Statue* mesh, the difference in MOS of the three versions becomes smaller as the distance increases. At $d = 12$, the difference between Version 1 and Version 2 is 0.25 MOS, and between Version 2 and Version 3 is 0.4 MOS. At $d = \{16, 20\}$, the MOSs of the three versions are almost the same. Given that the number of vertices of Version 3 is approximately 10% of that of Version 1, the resource required to render the mesh can be reduced by a large margin by choosing Version 3 instead of Version 1 for $d = \{16, 20\}$. These results show that understanding the impacts of distance from the camera on perceptual quality can help optimize the transmission and processing of textured meshes.

4. Objective Quality Metric Evaluation

In this section, we investigate the correlation of objective quality metrics with MOS scores obtained in Sect. 3. In particular, we consider eight image-based quality metrics, namely PSNR [16], SSIM [17], MS-SSIM [18], MSE [19], NMI [9], BRISQUE [20], NIQUE [21], and PIQE [22]. The first five metrics are full-reference metrics and the last three metrics are non-reference ones. Besides image-based quality metrics, geometry-based quality metrics have also been developed for textured meshes [12]. However, the geometry-based metrics are independent of the distance from the camera. Thus, we do not consider geometry-based metrics in this study.

We briefly explain the characteristics of individual metrics here. For more detailed descriptions, please refer to the original publications. The MSE metric measures the average squared difference in pixel values of the original and distorted images. The PSNR metric is derived from the MSE metric and indicates the maximum pixel intensity to the power of the distortion. The SSIM metric combines local image structure, luminance, and contrast into a single quality score. The MS-SSIM metric is an improved version of the SSIM metric in which luminance information at the highest resolution is combined with the structure and contrast information at several downsampled resolutions. The NMI metric measures the normalized mutual information between input images. The BRISQUE metric is a non-reference metric trained on a database with known distortions and thus is limited to evaluating images with the same type of distortion. Despite the fact that the NIQUE metric is trained on a database of pristine images, it can measure the quality of images with arbitrary distortions. The PIQE metric measures the local variance of distorted image blocks to compute the quality score.

Similar to [23], to measure the correlation of objective quality metrics with MOS, we first use a logistic function to predict the MOS from the objective quality metric:

$$S^p = d + \frac{a - d}{1 + (\frac{x}{c})^b} \tag{1}$$

where x is the objective quality metric and S^p is the predicted MOS. a, b, c, and d are learnable parameters; then, two correlation coefficients of Pearson Correlation Coefficient (PCC) and Root Mean Square Error (RMSE) are calculated for each metric. PCC quantifies the linear relationship between the predicted MOSs and the actual MOSs. Let N be the number of stimuli, S_i and S_i^p respectively be the actual MOS and predicted MOS of the ith stimuli, S_{mean} and S_{mean}^p respectively

Table 2 PCC and RMSE of objective quality metrics. Bold numbers indicate the highest correlations with the MOS.

Mesh	PCC								RMSE							
	PSNR	SSIM	MS-SSIM	MSE	NMI	BRISQUE	NIQUE	PIQE	PSNR	SSIM	MS-SSIM	MSE	NMI	BRISQUE	NIQUE	PIQE
Squirrel	0.79	0.67	0.66	0.80	**0.81**	0.44	0.50	0.40	0.46	0.49	0.50	0.39	**0.38**	0.64	0.57	0.60
Hulk	0.85	0.76	0.75	0.87	**0.90**	0.65	0.81	0.32	0.49	0.54	0.54	0.40	**0.35**	0.81	0.49	0.81
Statue	0.88	0.87	0.85	0.89	**0.93**	0.69	0.70	0.73	0.59	0.71	0.77	0.37	**0.30**	0.80	0.58	0.57
All	0.78	0.67	0.63	0.80	**0.81**	0.40	0.67	0.31	0.51	0.60	0.63	0.49	**0.47**	0.80	0.60	0.77

be the mean values of the actual and predicted MOSs, PCC is defined as follows [24].

$$PCC = \frac{\sum_{i=1}^{N}(S_i - S_{mean})(S_i^p - S_{mean}^p)}{\sqrt{\sum_{i=1}^{N}(S_i - S_{mean}) \sum_{i=1}^{N}(S_i^p - S_{mean}^p)}} \tag{2}$$

The larger the PCC, the better the correlation. The range of PCC values typically falls between −1 and 1, where −1 represents a perfect negative correlation, 1 represents a perfect positive correlation, and 0 indicates no linear correlation. The RMSE measures the difference between the predicted and actual MOSs, and is defined as follows [25].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(S_i - S_i^p)^2}{N}} \tag{3}$$

A lower RMSE indicates better performance.

The PCC and RMSE of each metric are shown in Table 2. Here, we consider two cases of individual and cross-meshes. The cross-mesh case, denoted by *All*, refers to the case where we use all the data of the three meshes to calculate the PCC and RMSE of individual metrics. It can be seen that, for both of the cases, the NMI metric achieves the highest PCC values and the lowest RMSE values (i.e., PCC ≥ 0.81 and RMSE ≤ 0.47) in all cases. This implies that the NMI metric can be used to evaluate the impact of distance on the quality of not only the same mesh but also different meshes. The MSE metric has a slightly lower performance than the NMI metric. Interestingly, the PSNR metric is found as a simple but good metric (i.e., PCC ≥ 0.78 and RMSE ≤ 0.59). Both the SSIM and MS-SSIM metrics are worse than the PSNR metric in terms of both PCC and RMSE. Although full-reference quality metrics have high correlations with the MOS scores, they require access to the original textured meshes which is might not always possible.

Among the three no-reference quality metrics, it can be seen that the correlation of the NIQUE metric with MOSs is the highest but unstable (i.e., $0.50 \leq$ PCC ≤ 0.81 and $0.49 \leq$ RMSE ≤ 0.60). Meanwhile, the BRISQUE and PIQE metrics have the smallest PCC and RMSE values. For the cases of *Squirrel* and *All*, the PCC values with the BRISQUE metric are smaller than 0.5. The PCC value of the PIQUE metric across all meshes is only 0.31. Therefore, it is still necessary to improve non-reference metrics for effective evaluations.

5. Conclusions

In this paper, we have studied the impact of distance from the camera on the perceptual quality of textured 3D meshes. Specifically, we have conducted a subjective experiment to assess the quality of nine distorted versions of three different meshes at five different distances from the camera. It is found that the MOS increases as the distance from the camera increases. However, the impact of the distance is variable for different meshes. Also, we have measured the correlations of objective quality metrics with the obtained subjective quality scores. The results showed that the NMI metric can be used to evaluate the impact of distance on the perceptual quality of not only the versions of the same mesh but also across different meshes. Also, PSNR is found as a simple but good metric. In future work, we will expand our evaluation to other types of meshes and distortion. In addition, we will evaluate the performance of geometry-based metrics in comparison to the image-based metrics. Also, we plan to develop a distance-aware adaptive transmission method for textured 3D meshes using the finding in this paper.

References

[1] "3D modeling." https://en.wikipedia.org/wiki/3D_modeling.
[2] "Polygon Mesh." https://en.wikipedia.org/wiki/Polygon_mesh.
[3] "Texture Mapping." https://en.wikipedia.org/wiki/Texture_mapping.
[4] D. Luebke, M. Reddy, J.D. Cohen, A. Varshney, B. Watson, and R. Huebner, Level of detail for 3D graphics, Morgan Kaufmann, 2003.
[5] F. Strugar, "Continuous distance-dependent level of detail for rendering heightmaps," Journal of Graphics, GPU, and Game Tools, vol.14, no.4, pp.57–74, 2009.
[6] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoue, "Subjective and Objective Visual Quality Assessment of Textured 3D Meshes," ACM Trans. Appl. Percept., vol.14, no.2, pp.1–20, Oct. 2016.
[7] Y. Nehmé, F. Dupont, J.P. Farrugia, P. Le Callet, and G. Lavoué, "Visual quality of 3d meshes with diffuse colors in virtual reality: Subjective and objective evaluation," IEEE Trans. Vis. & Comp. Graph., vol.27, no.3, pp.2202–2219, 2021.
[8] K. Vanhoey, B. Sauvage, P. Kraemer, and G. Lavoué, "Visual quality assessment of 3d models: On the influence of light-material interaction," ACM Trans. Appl. Percept., vol.15, no.1, pp.1–18, Oct. 2017.
[9] C. Studholme, D.L.G. Hill, and D.J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," Pattern recognition, vol.32, no.1, pp.71–86, 1999.
[10] Y. Pan, I. Cheng, and A. Basu, "Quality metric for approximating subjective evaluation of 3-d objects," IEEE Trans. Multimedia, vol.7, no.2, pp.269–279, 2005.
[11] W. Griffin and M. Olano, "Evaluating texture compression masking effects using objective image quality assessment metrics," IEEE Trans. Vis. Comput. Graphics, vol.21, no.8, pp.970–979, 2015.
[12] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3d colored point cloud and mesh models," IEEE Trans. Cir. & Sys. Video Tech., vol.32, no.11, pp.7618–7631, 2022.
[13] I. BT, "Methodologies for the subjective assessment of the quality

of television images, document recommendation itu-r bt. 500-14 (10/2019)," ITU, Geneva, Switzerland, 2020.

[14] "A-frame." https://aframe.io/.

[15] I.T. Recommendation, "P. 913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment."

[16] "Peak signal to noise ratio." https://en.wikipedia.org/wiki/Peak-signal-to-noise-ratio.

[17] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.

[18] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, pp.1398–1402, Ieee, 2003.

[19] "Mean Squared Error." https://en.wikipedia.org/wiki/Mean_squared_error.

[20] A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process., vol.21, no.12, pp.4695–4708, 2012.

[21] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a "completely blind" image quality analyzer," IEEE Signal Process. Lett., vol.20, no.3, pp.209–212, 2012.

[22] N. Venkatanath, D. Praneeth, M.C. Bh, S.S. Channappayya, and S.S. Medasani, "Blind image quality evaluation using perception based features," 2015 twenty first national conference on communications (NCC), pp.1–6, IEEE, 2015.

[23] H.T. Tran, C.T. Pham, N.P. Ngoc, A.T. Pham, and T.C. Thang, "A study on quality metrics for 360 video communications," IEICE Trans. Inf. & Sys., vol.101, no.1, pp.28–36, 2018.

[24] K. Pearson, "vii. Note on regression and inheritance in the case of two parents," Proc. Royal Society of London, vol.58, no.347-352, pp.240–242, 1895.

[25] L. Lebart, A. Morineau, and K.M. Warwick, "Multivariate descriptive statistical analysis; correspondence analysis and related techniques for large matrices," JMR, vol.22, no.2, p.225–226, 1985.