

Subjective quality assessment of textured human full-body 3D-reconstructions

Alexandros Doumanoglou, Nikolaos Zioulis, Emmanouil Christakis, Dimitrios Zarpalas, Petros Daras

Centre for Research and Technology - Hellas

Information Technologies Institute

{aldoum, nzioulis, manchr, zarpalas, daras}@iti.gr

Abstract—Geometry and texture resolution are two common system parameters of any modern volumetric 3D reconstruction pipeline. In tele-immersive applications, besides their apparent impact on the visual quality of the output 3D mesh, their absolute values implicitly influence the computational load of the whole tele-immersion pipeline from acquisition to 3D reconstruction, compression and transmission. Thus, tuning those parameters to an optimal combination has evident benefits. In this paper, we conduct a subjective experiment to assess the visual quality of textured human 3D-reconstructed meshes that are produced by a volumetric 3D reconstruction algorithm as a joint function of the geometry and texture resolution production parameters. The experiment is based on the forced choice pairwise comparison methodology on pre-rendered views of the real-time reconstructed meshes within the context of human performance capture. We analyze the pairwise comparison data and establish a ranking of the parameter space and, thus also, a mapping from the parameters to the subjective visual quality. The results of this study may be utilized to tune the parameters of the real-time 3D reconstruction pipeline, optimizing for the best balance between visual quality, bandwidth and overall performance.

1. Introduction

In all contemporary volumetric 3D reconstruction algorithms, like [1], [2], [3] and [4], the volume of the target scene is initially descritized into small cubical volume elements called voxels [5]. The size of those voxels constitutes a parameter of the system and is often referred to as geometry resolution. It is apparent, that the visual quality of the 3D reconstruction’s output is tightly coupled to the value of this parameter. Furthermore, during rendering, the 3D reconstructed meshes are eventually textured using the color images that were captured by the cameras during the acquisition phase of the reconstruction pipeline. The spatial resolution of the color images that are used to texture the reconstructed 3D mesh is a second parameter of the system that also affects the visual fidelity of the output meshes. In most real-time 3D reconstruction systems, higher geometry resolution implies better visual quality at the cost of additional computational burden. Moreover, oftentimes, the image resolution of the captured images also affect the acquisition frame-rate. Commodity RGB-D sensors usually

operate at higher frame-rates at lower resolution. Moreover, in the case of real-time transmission of the 3D reconstructed meshes (e.g in a live tele-immersion scenario) the increased volume of data corresponding to geometry and textures inhibit an additional computational cost to compress and actually send the data to the remote parties. Thus, the benefits of optimizing the geometry and texture resolution parameters of the 3D reconstruction pipeline with respect to a fair balance between visual quality and performance is evident.

In this work, we propose one of the first attempts to subjectively evaluate the visual quality of textured human full-body 3D reconstructions produced by a volumetric method [4]. We jointly evaluate the geometry and texture resolution parameters of the 3D reconstruction algorithm and we compute a model that maps resolution parameters to a visual quality score based on statistical results of subjective pairwise comparisons of images depicting rendered views of the 3D reconstructed human meshes. The results of this study may be utilized to tune the parameters of the real-time 3D reconstruction pipeline, optimizing for the best balance between visual quality, frame-rate and bandwidth.

2. Related Work

The main contribution of the present work is relating the subjective visual quality of reconstructed human 3D meshes to their production parameters: geometry and texture resolution. This is a pioneering direction emerging from the technological advances in digitizing real world content to 3D representations via depth sensing and 3D reconstruction algorithms. This is opposed to traditional 3D mesh creation in 3D modeling tools from specialist artists. To the best of our knowledge there is no other work performing this exact same task. However, in the previous years, significant efforts were made to invent objective metrics that correlate well with human opinion on the visual quality of 3D meshes in general. Those metrics can be separated in two major categories: those that operate on the 3D geometry and those that operate on viewpoint-based rendering of the 3D geometry. Further, there are a few proposed metrics that assess 3D geometry jointly with texture. Besides evaluating the visual quality of meshes directly in 3D space, it is also relevant to consider related work in the area of 2D image quality assessment, since the 3D models are mostly often

presented to the viewers as a rendered 2D image shown on a flat screen. Finally we consider related work of the most common existing surveying methodologies on subjective evaluation of traditional 2D images and videos.

Subjective visual quality evaluation of 3D meshes.

One of the first works on visual quality assessment of 3D meshes taking into account geometry and texture resolution and proposing an objective metric which fits geometry and texture resolution parameters to the collected subjective data is [6]. However, in this work the low geometry resolutions of the 3D meshes are decimated versions of a higher quality reference mesh. In the contrary, in the present work we evaluate geometry resolution as a production parameter, i.e: the output of the 3D reconstruction algorithm operating at a higher / lower voxel size and eventually evaluating the reconstruction algorithm's performance and not the performance of a decimation algorithm.

Later, in [7], a performance comparison between viewpoint-independent and viewpoint dependent metrics and their correlation with human judgments is presented. The paper concludes that visual quality evaluation can be performed either with image based or geometry based metrics. While there are arguments for both, there is not a clear group of metrics that is superior to the other. Moreover, it is found that that visual quality assessment is also depending on the semantic interpretation of the content.

Recently, apart from [6] which directly mapped decimation and texture downscale parameters to visual quality, other objective metrics have been introduced in the literature which operate in full, reduced or no reference setting [8]. In [9], a full reference objective metric for evaluating the visual quality of colored 3D point cloud is proposed. For 3D meshes, Dong et al. [10] introduced an objective full-reference metric based on curvature. However, their metric operates only between reference and impaired meshes with the same number of vertices which is not the typical case in varying geometry resolutions of 3D reconstructed meshes. On the other hand, a reduced and no-reference metrics for the same purpose were introduced by Abouelaziz et al. in [11] and [12].

Finally, the most recent work we know in this area is the one presented in [13] which apart from [6] is the only work considering joint distortions in both geometry and texture. In that work the formulated considered dataset includes five types of distortions: geometry quantization (often found in compression algorithms), geometry simplification, geometry smoothing, texture compression and texture downscaling. In the same work, Guo et al. propose two full-reference metrics for textured mesh visual quality prediction. While still relevant, the work of [13] has a different scope than the current paper. The present work computes a model that relates the 3D reconstructed mesh production parameters to subjective visual quality. Thus, the input to our proposed model are geometry and texture resolution. On the other hand, the model of [13] requires a reference and impaired mesh to judge about the subjective quality.

Subjective visual quality evaluation of 2D images. Similar to the previous case for 3D meshes, for 2D images, the

proposed metrics that evaluate the objective quality of a 2D image can be either in a full-reference, reduced-reference or no-reference setting. Typical full-reference objective metrics include PSNR, SSIM, multiscale SSIM (MS-SSIM), VIF, MAD and FSIM [14]. From those metrics, as evaluated in [14], VIF and MAD correlate better to subjective measurements in the root mean squared error (RMSE) sense. However, an improvement of the full-reference SSIM metric has been recently proposed in [15].

In [16], Xue et al. proposed a reduced reference image quality assessment metric based on Weibull statistics, while more recently, in [17] another reduced reference image quality metric was presented based on DCT Sub-band similarity.

In [18] a no reference image visual quality metric is proposed based on fusion of statistical and human visual system based metrics using ϵ -Support Vector Regression while Freitas et al, in [19], proposed another no-reference image assessment metric based on local ternary patterns.

Methodologies for visual quality assessment of images and videos.

The most commonly established methodologies for surveying subjects on image and video quality are the ones standardized by the International Telecommunication Union (ITU). The reference documents include ITU-R BT.500 [20] for images and ITU-T P.910 [21] for videos. In [22] a comparison is conducted between single-stimulus, double-stimulus, forced-choice pairwise comparison, and similarity judgment methodologies, introduced in [21], [23] and [24]. The analysis made in [22] concludes that the results with the lowest variance were obtained by the forced choice pairwise comparison methodology. In this type of subjective assessment, namely pairwise comparisons, two stimuli with the same content are being displayed to the observers forcing them to chose the best stimulus even if both stimuli possess no difference. To minimize the number of pairwise comparisons and analyze the results obtained by the subjects, Perez-Ortiz et al [25] introduced a practical guide and a MATLAB toolbox to compute a model that fits the subjective data. Finally, a more recent work which optimizes the number of pairwise comparisons adaptively in an online setting is introduced in [26].

In this paper, we chose to follow the subjective evaluation methodology described in [25].

3. Experiment

In this work we conduct an experiment to assess the visual quality of human 3D reconstructions produced by the method of [4] with respect to its two production resolution parameters for geometry and texture. In this section we describe all the steps we took in order to realize the experiment, along with all the decisions we had to make throughout the process.

Generating 3D Content. The first step to 3D content generation is the acquisition process. We captured 4 different humans in varying contexts with typical RGB-D sensors in a 360° multi-camera setup. Specifically, using 6 cameras, we captured 5 distinct performances, namely *kicking*, *punching*,



Figure 1. Snapshots from the 5 performances that were captured to produce the 3D sequences.

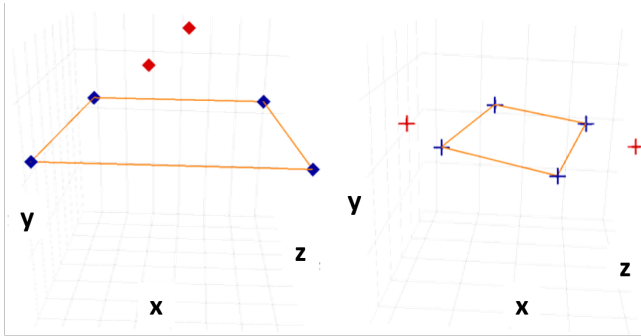


Figure 2. Positions of the *participating* (blue-connected by orange lines) and *non-participating* viewpoints (red). On the left (diamond patterns) the non-participating viewpoints are positioned higher than the participating ones, while on the right (cross patterns) they are on the same height. In all circumstances, they are positioned between two participating viewpoints.

conversing, dancing and physical exercising as presented in Figure 1, with their durations ranging from 11 to 62 seconds. Four out of six cameras are placed symmetrically around the center of the capturing area and their content is used to generate the 3D reconstruction of the captured human. The two remaining cameras are externally calibrated with the original four ones and are later used to acquire objective measurements in order to facilitate the selection process of the 3D content that was finally presented to the surveyed subjects. We refer to the first four viewpoints as the 3D reconstruction *participating* viewpoints and the two extra ones as the *non-participating* viewpoints. The non-participating viewpoints additionally offer un-biased views in the sense that we can examine the shortcomings of the 3D reconstruction process by placing the viewpoints in positions where most distortions would manifest at. The most common distortions would be occlusions that result in "holes", i.e. untextured areas, and the lowest texturing quality as a result of the viewpoint-based texture blending [4]. During the capturing process the non-participating viewpoints are positioned in between of participating ones, either on the same height and looking inwards, or higher and looking downwards towards the center of the capturing space, as illustrated in Figure 2.

From each captured sequence we generate the 3D reconstructed performance using the four participating viewpoints. The 3D mesh production parametric space considered in the present work consists of three discrete levels

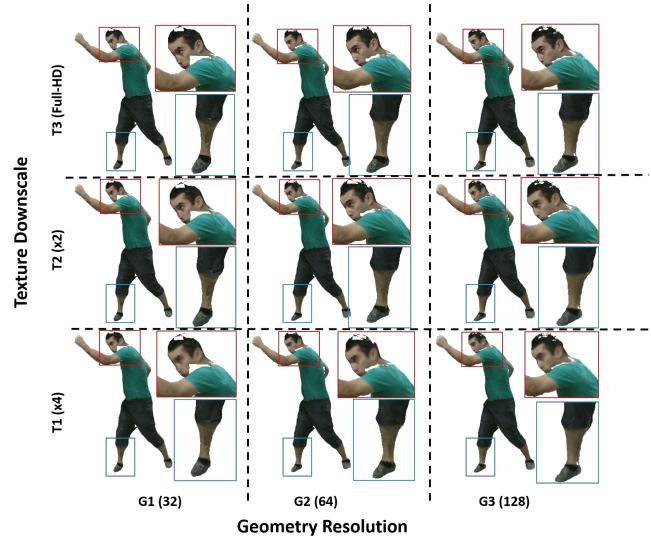


Figure 3. 3D reconstructions of the same frame in varying production parameters. Geometry resolution increases from left to right. Texture resolution increases from bottom to top.

for geometry resolution (32, 64 and 128, with the y axis resolution being the double of the reported number, i.e. the x and z axis resolutions) and three discrete levels for texture resolution (1920×1080 - original Full-HD resolution, 960×540 - downsampled by a factor of 2, and 480×270 - downsampled by a factor of 4). Therefore, we reconstruct each sequence 9 times (using all combinations of geometry and texture resolution) and generate a total of 45 3D reconstructed performances in varying parameterizations with an example presented in Figure 3. The lowest resolution levels for geometry and texture were chosen as such because, for most people, further reduction of those values result into 3D meshes of unacceptable subjective visual quality. On the other hand, the upper value for texture resolution was determined by the capabilities offered by the used RGB-D sensor, while the upper value for geometry resolution was based on the maximum computational load that our hardware can handle in order for the algorithm to be real-time.

Experimental Setup. The subjective evaluation methodology adopted in this paper was forced choice pairwise comparisons (PC) [25], during which the participants are asked to choose the preferred content between two choices. Compared to other methods, PC is easier for the participants, especially for non experts, and simpler to implement. Additionally, in the literature, this method was found to also produce measurements of smaller variance compared to other alternatives [22]. However the number of possible comparisons can get quite large especially if more than one parameters are modified. This problem can be mitigated by always presenting content of similar quality. This can significantly reduce the number of required comparisons. In our case, and by following the guidelines proposed in [25], we choose to compare 3D generated content for which the

one production parameter is constant and the second differs only by one level. This reduces the number of comparisons for every 3D content from 36 to 12. Moreover, this also satisfies the requirement described in the above work to avoid pairwise comparisons where one content is overwhelmingly preferred over the other, as this case tends to give biased or even wrong results.

Since the content is 3D, our typical presentation options include unrestricted free view-point viewing, pre-defined continuous changing viewpoint animations and static pre-render views. Each presentation option has its own benefits and shortcomings [7], [13] with no method dominating the others in all relevant aspects. We chose to adopt static pre-rendered views for simplicity, fairness for comparison between different subjects (as they all see the same view) and minimization of the survey’s duration while maximizing variation in human performance types.

After producing the human 3D-meshes from pre-recorded captures of human performances in all discrete levels of the production parameters, we were required to follow a strategy in order to select the actual content (static frames) that would be showcased to the subjective survey. Initially, the selection process was mainly driven by the total expected duration of the survey for each subject. With a target of 30 mins per subject and asserting that a pairwise choice can be made in 5 seconds we can afford a total of 360 comparisons. Since we need 12 comparisons of different quality levels for the same content (pre-rendered frame), the total number of distinct frames that we can present to the subjects are 30. Those 30 pre-rendered frames are equally distributed among all the 5 distinct captured performances resulting into 6 total frames per captured sequence.

For the same value of the geometry resolution parameter it is expected that the 3D reconstructed mesh will be of better visual quality when the captured user’s bounding box is of smaller size. This is due to the nature of the volumetric 3D reconstruction algorithm which operates on the discretization of the captured space in constant number of voxels. The higher the captured space (bounding box) the higher the voxel size and thus the lower the fidelity of the output 3D mesh. Depending on the type of the performance executed by the captured users, it is common that the bounding box of the captured user varies significantly across time. In our survey we would like to include frames corresponding to small, medium and large voxel sizes from each captured sequence. In order to accomplish that, we esteem that along a sequence of 3D meshes produced with the same production parameters, the frames that correspond to lower voxel size are the ones producing better scores with respect to a reference ground-truth image when compared using an objective image similarity metric. In our case we were able to automatically extract frames whose bounding box is low, medium or large by computing SSIM [27] scores between the reference ground-truth non-participating captured images of the reconstructed users and a rendered image of the 3D reconstructed human from the same view-point. We distributed the 6 total frames of each captured sequence we computed in the previous paragraph among the

Age			Gender		Experienced	
Min	Max	Average	Male	Female	Yes	No
25	43	30	26	4	10	20

TABLE 1. DEMOGRAPHICS OF SURVEYED SUBJECTS.

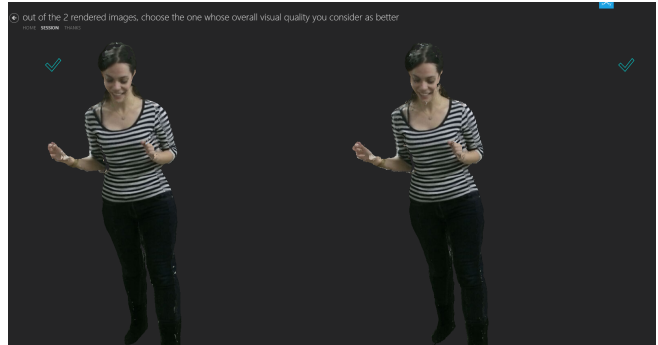


Figure 4. A screenshot of the forced pairwise comparison application used for the survey. Users watch the content side by side and select their preference by clicking on the respective (left or right) green tick mark.

2 non-participating view points and among 3 levels of SSIM (low, mid, high implicitly corresponding to high, mid and low voxel sizes). The additional advantage of employing an image objective metric (SSIM) in our frame selection strategy, is the increased trust we can have in our selection process to not be biased towards only good or bad 3D reconstructions. The objective metric gives an indication of the quality of the 3D reconstruction with respect to the non-participating viewpoint which is used as a reference.

Generating pre-rendered views. The final pre-rendered views that were presented to the surveyed subjects were produced by the following process. Using the extra sensor viewpoints which are externally calibrated onto the same coordinate system as the participating viewpoints, we render each sequence’s frame after positioning the virtual camera at the external viewpoint’s position and set the projection matrix according to the sensor’s intrinsics. We set the rendering output to be Full-HD. In this way, we acquire novel views of the 3D reconstructed sequence (i.e. views that are not aligned with the capturing setup which participated in the 3D reconstruction) and we can additionally calculate objective metrics for these rendered views against the sensor acquired color data as described in the previous subsection. It is maybe worth to note that the objective metric (SSIM) is calculated by taking into account only the area of the image that corresponds to the rendered portion of the 3D reconstructed mesh. Finally, for the rendering process of the 3D reconstructions we used no shading or lighting calculations, faithfully reproducing the production’s reconstructed mesh.

Survey. We developed an application implementing the aforementioned pairwise comparison methodology with unrestricted voting times that was then presented to the subjects participating in the experiment. The application would rotate along the 360 total comparisons enforcing a minimum viewing time of 3 seconds before allowing voting. Moreover, to counter potential biases of the users consistently voting

to the same side (right or left) when they see no difference, the application would randomize the order (left/right) of the presented rendered images. Finally, the application would also collect the demographics of the users namely their age, gender and whether they have previous experience with 3D reconstructions.

We scheduled a 45 min survey per subject, giving room for average voting times of slightly higher than 5 seconds and a small break in the middle to avoid fatigue. We were not strict in enforcing the aforementioned time limit in the total duration of the survey for each subject. Each one would participate in their own pace, with some finishing in 20 mins and others in 1 hour. The average duration of the survey per subject was 45.2 mins. In total, 30 subjects aged from 25 to 43 were surveyed with their detailed demographics presented in Table 1. We chose the number of subjects in accordance with [25] where in their analysis it is shown that increasing this number beyond 30 would not significantly decrease statistical metrics like confidence intervals or RMSE.

4. Results

We chose to process the survey data by using the toolbox provided by Perez-Ortiz et. al. [25]. The computed mathematical model gives a rating to each pair of production parameters based on the subjective data collected by the survey. By convention, the first examined quality is given a zero rating and the rest are assigned a relative score with respect to the initial, based on the subjective data. As a result of this, we can directly compare the various qualities and gain insight on which parameter has the biggest impact on the user's quality of experience.

Furthermore, following the guidelines of the respective paper [25], we performed outlier detection on the collection of our subjective data. However, quite surprisingly, the outlier detection method proposed in [25] did not find any outliers in our data collection. This probably means that the opinions of all subjects more or less converge to the same values. Moreover, this also strengthens our position in claiming that the generated dataset had meaningful pairwise comparisons.

The final visual quality score of the production parameters given by [25] is shown in Figure 5 in the form of a heat-map, while the actual score values are presented inside the cells. The three geometry resolution levels ($G1$, $G2$, $G3$) correspond to geometry resolutions 32, 64 and 128, while the three texture levels ($T1$, $T2$, $T3$) correspond to texture resolutions of downscale $\times 4$, downscale $\times 2$ and finally Full-HD. From the heat-map it is evident that the subjective visual quality is a monotonically increasing function of geometry and texture resolution with respect to both parameters. For constant geometry resolution, the visual quality is a monotonically increasing function of texture resolution and vice versa. However, the visual quality of all meshes produced by texture resolution level $T1$, regardless of geometry resolution, are almost equally bad. Moreover, for geometry resolution level $G1$ there are more substantial

differences in subjective visual quality when texture resolution increases. This is a well known fact in the 3D Graphics research community as it describes the case where higher texture resolution masks geometric artifacts [7].

However, the geometry resolution level $G1$ with the best possible texture does not outperform geometry resolution level $G2$ with the medium texture level. This is in contrast to geometry resolution level $G2$ with the best texture scoring slightly higher than geometry resolution level $G3$ with medium texture. This fact can be explained by the following argument. The geometry resolution, besides the overall fidelity of the produced triangle mesh, is also directly linked to the reconstruction of finer details like the hands and the face. In addition, given the way the textures are applied to the mesh, it also affects the overall visual quality as lower geometry resolutions manifest into stronger color discontinuities and deteriorate the accuracy of the texture mapping process [4]. From the experimental results it is deduced that transitioning from geometry resolution level $G1$ to $G2$ is more substantial than transitioning from level $G2$ to level $G3$.

Overall, as a conclusive point we can say that 3D reconstructions with the worst texture or geometry resolution value are highly undesirable and those pairs of parameters should be avoided in respective applications. On the other hand, the heat-map's mid point (geometry resolution level $G2$ and texture resolution level $T2$) seems to be a good compromise between visual quality and performance of the real-time 3D reconstruction pipeline as this combination of parameters can boost performance in both frame-rate and resulting network bandwidth when used in network applications. However, starting from this optimal quality and looking to improve it, we are faced with a dilemma: which resolution do we increase? While the scores acquired in this study point towards improving the texture resolution, this also comes at a great bandwidth cost as typically 4 images are transmitted per frame. On the other hand, improving the geometry resolution would result in less - but comparable - gain, but instead increase the processing time. This issue can possibly be addressed either adaptively or be tuned differently for each specific application.

5. Conclusion

In this work we have conducted a survey to map the production parameters of live 3D reconstructed meshes to the resulting subjective visual quality. It involved pairwise comparisons between pre-rendered views of 3D content produced in different geometry and texture resolutions. Our findings highlight the challenge associated with the selection of those parameters in live 3D content productions. There are two directions to improve the visual quality, one is associated with higher bandwidth costs while the other involves higher computational load. Whilst there are also other aspects to be considered (e.g. compression), these findings can be used to fine-tune the quality of experience for live 3D content streaming on the production side.

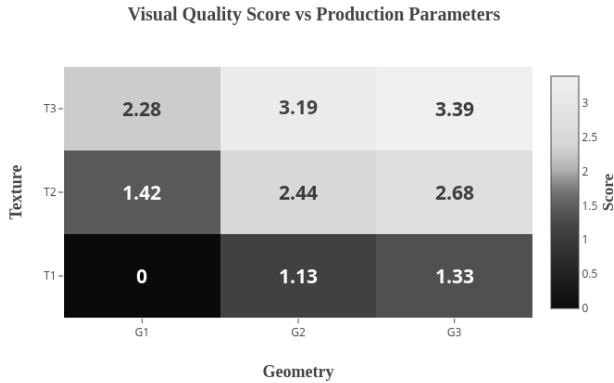


Figure 5. Visual Quality Score vs Production Parameters. From Left to Right: Geometry resolution: ($G1$)32, ($G2$)64 and ($G3$)128. From bottom to top: Texture resolution: ($T1$) downsampled $\times 4$, ($T2$) downsampled $\times 2$, ($T3$) Full HD.

Acknowledgments

This work has been supported by the EC Project 5G-MEDIA (www.5gmedia.eu). This project has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 761699.

References

- [1] A. Dai, M. Niessner, M. Zollhöfer, S. Izadi, and C. Theobalt, "Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration," *ACM Trans. Graph.*, vol. 36, no. 3, pp. 24:1–24:18, May 2017.
- [2] M. Dou, P. Davidson, S. R. Fanello, S. Khamis, A. Kowdle, C. Rhemann, V. Tankovich, and S. Izadi, "Motion2fusion: Real-time volumetric performance capture," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 246:1–246:16, Nov. 2017.
- [3] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson, S. R. Fanello, A. Kowdle, S. O. Escolano, C. Rhemann, D. Kim, J. Taylor, P. Kohli, V. Tankovich, and S. Izadi, "Fusion4d: Real-time performance capture of challenging scenes," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 114:1–114:13, Jul. 2016.
- [4] D. Alexiadis, D. Zarpalas, and P. Daras, "Fast and smooth 3d reconstruction using multiple rgb-depth sensors," in *Visual Communications and Image Processing Conference, 2014 IEEE*. IEEE, 2014, pp. 173–176.
- [5] P. Eisert, "Reconstruction of volumetric 3d models," *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*, p. 133, 2005.
- [6] Y. Pan, I. Cheng, and A. Basu, "Quality metric for approximating subjective evaluation of 3-d objects," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 269–279, 2005.
- [7] A. Bulbul, T. Capin, G. Lavoué, and M. Preda, "Assessing visual quality of 3-d polygonal models," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 80–90, 2011.
- [8] International Communication Union, "Recommendation itu-t g.1011: Reference guide to quality of experience assessment methodologies,"
- [9] J. Zhang, W. Huang, X. Zhu, and J.-N. Hwang, "A subjective quality evaluation for 3d point cloud models," in *Audio, Language and Image Processing (ICALIP), 2014 International Conference on*. IEEE, 2014, pp. 827–831.
- [10] L. Dong, Y. Fang, W. Lin, and H. S. Seah, "Perceptual quality assessment for 3d triangle mesh based on curvature," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2174–2184, 2015.
- [11] I. Abouelaziz, M. Omari, M. El Hassouni, and H. Cherifi, "Reduced reference 3d mesh quality assessment based on statistical models," in *Signal-Image Technology & Internet-Based Systems (SITIS), 2015 11th International Conference on*. IEEE, 2015, pp. 170–177.
- [12] I. Abouelaziz, M. El Hassouni, and H. Cherifi, "No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression," in *International Conference on Image and Signal Processing*. Springer, 2016, pp. 369–377.
- [13] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoue, "Subjective and objective visual quality assessment of textured 3d meshes," *ACM Transactions on Applied Perception (TAP)*, vol. 14, no. 2, p. 11, 2017.
- [14] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *arXiv preprint arXiv:1406.7799*, 2014.
- [15] M. Fidalgo-Fernandes, M. V. Bernardo, and A. M. Pinheiro, "A bag of words description scheme based on ssim for image quality assessment," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*. IEEE, 2016, pp. 1–6.
- [16] W. Xue and X. Mou, "Reduced reference image quality assessment based on weibull statistics," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*. IEEE, 2010, pp. 1–6.
- [17] A. Balanov, A. Schwartz, and Y. Moshe, "Reduced-reference image quality assessment based on dct subband similarity," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*. IEEE, 2016, pp. 1–6.
- [18] M. D. Dimitrievski, Z. A. Ivanovski, and T. P. Kartalov, "No-reference image visual quality assessment using nonlinear regression," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*. IEEE, 2011, pp. 78–83.
- [19] P. G. Freitas, W. Y. Akamine, and M. C. Farias, "No-reference image quality assessment based on statistics of local ternary pattern," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*. IEEE, 2016, pp. 1–6.
- [20] International Telecommunication Union, "Recommendation ITU-R BT.500: Methodology for the subjective assessment of the quality of television pictures," 2012.
- [21] —, "Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications," 2008.
- [22] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," in *Computer Graphics Forum*, vol. 31, no. 8. Wiley Online Library, 2012, pp. 2478–2491.
- [23] H. de Ridder and G. M. Majoor, "Numerical category scaling: an efficient method for assessing digital image coding impairments," in *Proc. SPIE*, vol. 1249, 1990, pp. 65–77.
- [24] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 640–648.
- [25] M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," *ArXiv e-prints*, Dec. 2017.
- [26] R. Heckel, M. Simchowicz, K. Ramchandran, and M. J. Wainwright, "Approximate Ranking from Pairwise Comparisons," *ArXiv e-prints*, Jan. 2018.
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.