https://doi.org/10.2352/issn.2168-3204.2024.21.1.17

3D Object Quality Metrics and their differences: How can we evaluate quality of digitization?

Markus Sebastian Bakken Storeide; Colourlab, Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Sony George; Colourlab, Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Aditya Suneel Sole; Colourlab, Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Jon Yngve Hardeberg; Colourlab, Department of Computer Science, Norwegian University of Science and Technology; Gjøvik Norway

Abstract
Selecting the optimal resolution and post-processing techniques of 3D objects for cultural heritage documentation is one of the most distinguishable challenges within 3D imaging. Many techniques exist to document a tangible object at very high objective accuracy, but there also exist techniques that can visualize a similar perceptual accuracy without documenting the objective values.
The application difference between storage of complex geometric data and the visualization of it could be fundamentally different, and if the two methods are not disassociated it could lead to either false or inaccurate digital documentation of a cultural heritage object. In this investigation we compare several different metrics for evaluating the quality of a 3D object, both objectively and perceptually, and look at how the different approaches might report greatly different outputs based on the post-processing of a 3D object. We also provide some insight in how to interpret the output of various metrics, and how to compare them.

Introduction
3D objects are constructed mainly by two concepts; geometry and topology. Geometry consists of the number of vertices of a 3D object and their distribution in 3D space, and the topology is the method of how these vertices are connected to create a surface. Manipulation of these features creates the final, optimized model, which based on the application could vary greatly in how the features are processed and prioritized. An investigation of the effect of this was started by the authors in 2022 [1][2] but the various changes cannot be covered by a single paper. Due to the wide application of 3D in the cultural heritage (CH) field, the existing data in various databases greatly differ from each other due to various applied mesh processing steps. Metrics for comparing the quality of these data, in both geometry and topology, are therefore of vital importance to the sustainability of 3D digitization, especially with the exponential growth in data in the last couple of years. The 3D object metrics tested in this paper compute a quality value on the 3D objects themselves, and differ from image quality metrics that could be used on renderings of the objects.

The prominent approach in calculating this orient around the mesh saliency [3]. Saliency is defined as areas that are particularly noticeable or important to our perception of a 3D object, but also regions that are of higher geometric importance. Computations of mesh saliency can be done by considering different features of a mesh, like roughness or curvature, but the output can greatly vary depending on which features you select. Saliency computations are a predictor of what regions of a 3D object an observer would be able to detect changes to the mesh if it is subject to a post-processing step. Similarly, perceptual metrics also primarily considers the features of roughness and curvature within the geometry and topology of a mesh.
Our investigation explores the question of whether 3D objects should be processed and archived based on perceptual or objective metrics, and how different quality measures of 3D objects correlate with each other.

Related Work
Lavoué and Corsini previously provided a comprehensive comparison of perceptually based metrics for objective evaluation of geometry processing [4]. Some of the processing outcomes they investigated, like simplification and noise reduction, are common mesh processing operations in workflows for CH, while others like watermarking or smoothing are not equally as relevant for this field. Additionally, due to their general-purpose approach for computer graphics, they also left out some operations that are more typical for CH like surface reconstruction and making meshes watertight.

Modern 3D objects can be very complex, as the data acquisition technologies have improved. These are often much larger than the ones tested by Lavoué and Corsini, where the highest number of vertices and faces were 100.000. It is common today for models to have several hundred thousand or millions of vertices, and this considerable increase in data might produce different results than previous tests done with lower sample sizes. Testing some developed metrics for evaluating quality on modern 3D objects will therefore provide more accurate data of their usefulness in the field of CH.

A recurring problem in utilizing 3D in CH is to distinguish between computer graphics and cultural heritage documentation. Quality metrics are an important tool to measure the reliability of digital fidelity, but universally agreed-upon implementations are currently lacking in 3D workflows. There are several cases where these metrics might be utilized to evaluate different features of a 3D object, but which are distinctly different in nature. While human perception is important in workflows that target visualization of acquired 3D data, it has a lower priority for applications that solely consider the geometric accuracy of a

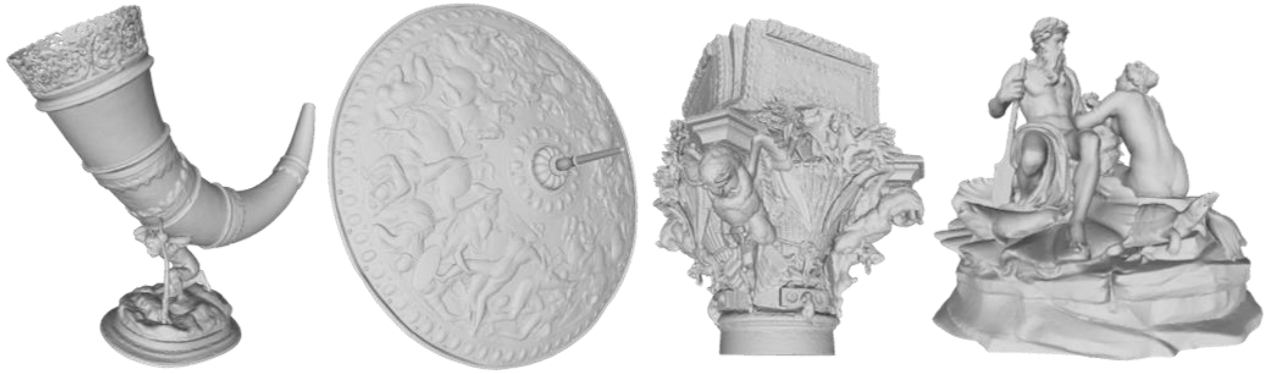

Figure 1: Reference 3D Objects: 'Horn', 'Shield', 'Owl', and 'Statue'.

measured tangible object. However, as visualized by [6], objective measurements are in some cases not able to capture errors in features of 3D objects that have a very high perceptual saliency, but where the objective distance measurement is very low. This also highlights the opposite effect, where a noticeable difference in the mesh topology might not have that great geometric error. Considering both approaches, objective and perceptual, might therefore be the better approach.

In 3D workflows that impel users to post-process their data, through various data optimization algorithms, it is also unknown how these optimization filters affect the objective and perceptual quality of a 3D object, as many of the features they optimize does not relate to this. Lastly, the quality of complex geometric measurements of tangible objects cannot be comprehensively evaluated by a single metric, due to the flexibility and diverse nature of 3D documentation. Metrics works with surface geometry, but not topology, even though both features contribute to the construction of a 3D object.

In Lavoué and Corsini's investigation they found that the metrics have drastically different correlations to subjective evaluation data, and that the geometric metrics had lower correlation than the metrics designed around human perception. This correlation is perhaps as expected, but they did not go into detail about what objective features of the objects might be lost if we only consider the perceptual metrics as a quality measure. Additionally, they utilized synthetic errors on a dataset that is outdated by modern standards, which renders their results less useful when working with novelty 3D objects. In this paper we do the opposite, and only apply filters and post-processing stages that are designed to improve the quality of a mesh in various ways. So instead of testing if a metric can accurately evaluate a reduction in quality from a reference mesh, we test how it interacts with theoretical improvements to a mesh.

Method

We have selected four different objects to test, which vary in tangible size, level of detail, and resolution. This selection was made to have 3D-scans of tangible heritage objects with a high polycount and geometric complexity, and some variation in their shape. Since the tangible objects have different sizes, the digital versions also had a different scale when opened in a 3D viewer. To make it so that the metric outputs could be compared to each other, we set the scale of the objects to be approximately the same. Established and previously used 3D object databases like LIRIS/EPFL general-purpose database [5], LIRIS masking database [6], or IEETA simplification database [7] generally feature lower polycount versions of simpler objects, which we consider an unrealistic representation of the modern 3D objects that are available in the CH field. The selected objects can be seen in Figure 1.

Post-Processing Steps

We tested three post-processing steps: Smoothing, Simplification, and Quadrangulation. All steps were introduced in Meshlab [8]. These are steps that are commonly applied to 3D objects for different reasons, and in general seek to improve the mesh.

Smoothing [9] consists of averaging a vertex' location based on weights from surrounding vertices, thereby ironing out very rough surfaces. When capturing 3D data, you also capture a lot of noise, rendering otherwise smooth surfaces with a rougher finish even after noise filtering. Averaging the surface in a smoothing step flattens vertices that are especially different from their surroundings, creating a more uniform surface. This improvement is primarily perceptual, but might remove correctly digitized rough surfaces because they are interpreted as noise.

Most 3D-meshes have very dense data that include a lot of redundancies, which is why a **simplification** step is applied to make the files easier to manage. This step reduces the amount of data used to create the geometry while attempting to keep it as close to the original as possible. For our simplification we utilized the Quadric Error Metrics algorithm [10] and reduced the data to 25% of the original, based on the previously mentioned simplification research.

Meshes are always rendered and stored as triangles, therefore all outputs from a 3D-scanning process are triangulated meshes. A drawback with this is that triangulated meshes are very impractical to work with if you want to do some manual processing of the mesh, due to the lack of edge-loops and clean topology. **Quadrangulating** the mesh transforms these triangles into quads, or rectangular polygons, which are much easier to work with for a human, but also changes the topology of the object quite drastically. These can be rendered and stored as normal without retriangulating them, as rendering software automatically converts quads to triangles at render time. We utilized the smart triangle pairing filter for this process.

Each of these processes have been applied to the high-resolution version of each of the four 3D objects in the same manner. We do not imply that these processes necessarily reduce or improve the quality of the mesh; however, they certainly introduce changes to it.

Objective and Perceptual Metrics

We test two categories of metrics, objective and perceptual. Tested metrics based on objective distance calculation with no relation to perceptual tests are the **Hausdorff Distance**, **Chamfer Distance**, and **Earth Mover's Distance**. Metrics based on subjective experiments with perceptual quality are **FMPD**, **GL2**, **3DWPM2**, and **MSDM2.** Their functionality is explained below.
We note that all metrics tested in this paper are full-reference types, directly comparing the information in the reference object to the processed object. They are therefore only applicable to meshes that have undergone some post-processing step, and report the relative quality of the object subsequent to the changes introduced by this step.

Objective Metrics

Hausdorff distance [11] computes the lowest distance between a sample point in the reference mesh to the sampled mesh. Output values are given relative to the bounding box diagonal of the 3D object, meaning that the Hausdorff values between two different objects can be compared against each other. We applied the computation to sample each polygonal face of our reference object, and the output value is the mean of all of these samples.

Chamfer Distance [12] computes the nearest neighbour correspondence of two point clouds and sums the square distances between them. Values are sampled from all vertices in the reference mesh and again reported relative to the objects bounding box diagonal, meaning it can be directly compared to the output of the Hausdorff distance. The single value is the mean of all the sample values.

Earth Mover's Distance [13], also known as the Wasserstein Metric, computes the differences in the distributions of vertices instead of a direct distance between them. We can interpret this output value as how much it would cost to transform the reference mesh to the sampled mesh. The cost is defined as the amount of points moved multiplied by the Euclidian distance it is moved. During a transformation, denser 3D objects would have more points to move, but each point would have to move less. Sparser 3D objects would have fewer points to move, but over a larger distance. A difference in resolution should therefore not greatly affect the output of the metric. Number wise, we cannot directly compare the Earth Mover's distance to the two other distances. It rather tells us how much the location of the vertices used to create the geometry changed from the reference to the sampled mesh, even though the distance of the surface could be relatively low. To compare the Earth Mover's Distance output between the different 3D objects, they need to be approximately the same size in a 3D geometric space.

For the Hausdorff and Chamfer distance values, a lower value signifies a smaller difference between the reference and the sampled object. As the post-processing stages should preferably introduce as little change to the object as possible, we say that the lower number the better. The Earth Mover's Distance value can be used to contextualize the output of all the metrics, as it says nothing about the geometric accuracy of the mesh but its change in vertex distribution. A high value in this metric signifies that the cost of changing the mesh is high, or how much the location of the vertices changed. For example, if the geometric change introduced by a post-processing stage reports medium error value with the Hausdorff distance, and a very high cost value with the Earth Mover's Distance, we could for example argue that the post-processing step is not worth the computing cost it requires.

Distances all output a value representing the differences between the reference and sampled object in a global, objective way, without consideration to the local geometry of the 3D object. The Earth Mover's distance is especially abstract as it samples the distribution of vertices in 3D space rather than the actual distance between the surfaces of the two objects. It therefore tells by how much the geometry changes, but not where or in what way.

Perceptual Metrics

FMPD (*Fast Mesh Perceptual Distance*) [14] considers the global roughness of an objects' surface relative to its local roughness, and compares this between the referenced and sampled mesh. Size of local roughness regions are set so that the whole surface is covered, providing data to compare roughness measures across the salient areas of both objects. This metric is symmetric, meaning that it still works even if the sampled mesh is smoother than the reference, although we expect the output to be of limited use if this is the case. The output value is scaled to be between 0 and 1, where extreme values of difference are thresholded to be 1.

GL2 (*Geometric Laplacian Measures*) [15] measures local geometric differences after a vertex smoothing step, where the metric output is the mean of the root-mean-square of these differences. A lower value represents a smaller noticeable difference between the two meshes. This metric only works for *consistent* meshes, meaning that they have the same number of vertices. For the simplifications step which reduces the number of vertices in the mesh, this metric is therefore not tested.

3DWPM2 (*Watermarking Roughness Measure*) [16] is like the GL2 metric, but considers the whole model at once instead of local smoothing steps. The difference in roughness is then evaluated, which is the output of the metric. The lower the score, the smaller the difference between the two meshes.

MSDM2 (*Mesh Structural Distortion Measure*) [17] evaluates structural similarity, which by their calculations consist of curvature, contrast, and structure computed within a neighbourhood around a vertex. The final global metric value is the Minowski sum of the distances between these neighbourhoods. For the MSDM2 metric the values are scaled between 0 and 1, where 0 are identical meshes. This metric was calculated with MEPP2 software development kit. [18].

For the perceptual metrics that are scaled between 0 and 1, FMPD and MSDM2, we can directly compare the two outputs. Other metrics must be regarded as a gradient, where a proportionately higher number represents a proportionally higher error.

All the metrics were computed for every object/post-processing step combination, the outputs are shown in Tables 1-4.

Metric Outputs

As the tested metrics compute their output very differently, their outputs are also reported relative to different aspects of the mesh. Some have very high values due to a greater change in whatever characteristic they are measuring, while others can have very low change in another measured characteristic. Numeric output values can therefore not be compared to each other directly, but with understanding of the metrics' functionality we can make some assumptions by visually inspecting the data.

Table 1: Objective and perceptual metric outputs for the Horn object.

Horn	H_d	C_d	EM_d	$FMPD$	GL_2	$3DWPM$	$MSDM2$
Smoothing	0.297525	0.127241	0.4168304	0.0000022	0.000163	0.1515	0.216893
Simplification	0.032994	0.880154	2.371780	0.0000736	-	6.723	0.392662
Quadrangulation	0.028034	0.000000	0.0000001	0.2500375	0.001559	1.533	0.125561

Table 2: Objective and perceptual metric outputs for the Shield object.

Shield	H_d	C_d	EM_d	$FMPD$	GL_2	$3DWPM$	$MSDM2$
Smoothing	0.004265	0.004940	0.000335	0.0000018	0.000095	0.8363	0.419405
Simplification	0.001257	0.043216	0.494611	0.0000176	-	1.2352	0.180587
Quadrangulation	0.000046	0.000002	0.000001	0.0000011	0.000839	0.9382	0.099297

Table 3: Objective and perceptual metric outputs for the Owl object.

Owl	H_d	C_d	EM_d	$FMPD$	GL_2	$3DWPM$	$MSDM2$
Smoothing	0.006607	0.006717	0.000289	0.0000003	0.019536	3.3013	0.360777
Simplification	0.001914	0.031733	1.616149	0.0000084	-	7.3026	0.216110
Quadrangulation	0.001377	0.000000	0.000000	0.0915861	0.046382	0.9503	0.400351

Table 4: Objective and perceptual metric outputs for the Statue object.

Statue	H_d	C_d	EM_d	$FMPD$	GL_2	$3DWPM$	$MSDM2$
Smoothing	0.002783	0.003466	0.000244	0.0000014	0.000268	0.2195	0.265384
Simplification	0.000711	0.044154	1.037726	0.0000385	-	1.6382	0.223125
Quadrangulation	0.000271	0.000000	0.000000	0.0824165	0.000182	0.1684	0.316858

Results & Discussion

Inspecting the data, there are a few assumptions we could make. For example, the highest values for the Earth Mover's Distance (EMD) are always for the simplification step, which makes sense since the number of vertices was reduced by 25%. The distribution between the reference and simplified object is therefore quite drastic. Similarly, we can see that the Chamfer Distance (CD) also has high values during simplification, as this introduces higher changes to the geometry of the objects than the other post-processing steps. It is therefore expected that these two metrics have a high correlation in their outputs. Simplification introduces a lot of changes to the geometry, but as seen with the values of the Hausdorff Distance (HD) and the perceptual metrics, it might not be too noticeable. Contextualizing the error values of HD and CD with EMD and the perceptual metrics might therefore allow a user to make informed choices of how accurate a 3D object that is designed purely for visualization is to the high-resolution original. This could be important when visualizing the object to an audience, where the user would preferably show an object that is as close to the tangible original as possible.

The roughness-based metrics of FMPD and MSDM2 seem to have higher values for the quadrangulation step compared to the curvature-based metrics of GL2 and 3DWPM. Considering that the quadrangulation step creates new topology within the model, we could assume that the roughness is more affected by topology than curvature. As the HD and CD values are relatively low for this step, these metrics might then be able to detect changes to the perceptual quality of the object even though the geometric error is very low.
Our assumptions are made with a limited amount of data, so more testing should be done with a variety of 3D objects to make more comprehensive conclusions.

Correlation matrixes between the tested metrics are good ways of visualizing how the output values can be interpreted against each other, instead of visually inspecting the output tables. Below are the correlation matrixes for the 'Owl' and 'Statue' objects. Not that these matrixes are between all the values of the different post-processing steps.

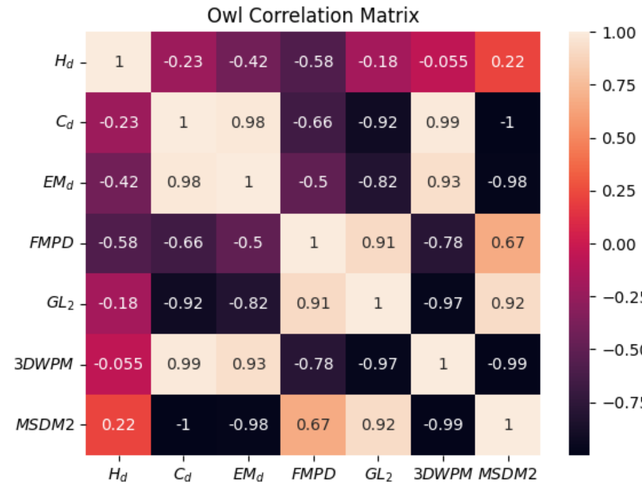

Figure 2: *Correlation matrix between metrics for 'Owl' object.*

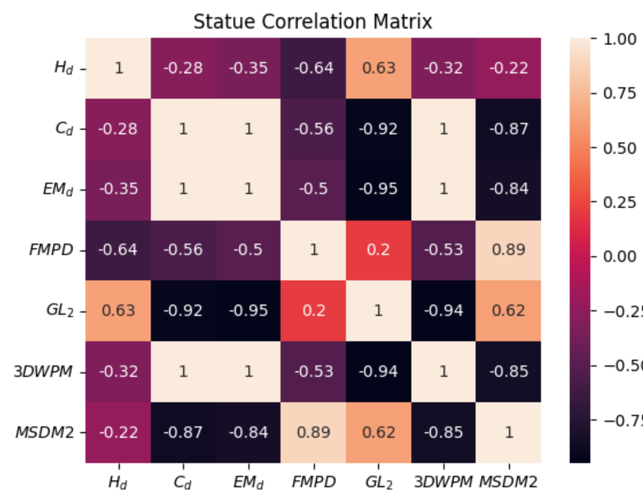

Figure 3: *Correlation matrix between metrics for 'Statue' object.*

Utilizing and understanding the metrics tested in this paper can lead to good discussions on what is the preferred and sustainable documentation practice; 3D objects that look good, or 3D objects that are geometrically accurate. The data shows that the post-processing steps affect the quality and accuracy of the different 3D objects very differently, meaning that we should consider the object geometry when applying introducing these steps. Additional testing on more 3D objects using the same metrics can increase our understanding of their behavior, leading to more informed interpretation.

Values reported by the metrics are a global average across the whole model, and can therefore provide a limited understanding of the object's local quality. Some regions might have a very high score, and others a very low score. But as all the metrics apply local values to local neighborhoods across the objects it is possible to visually and statistically compare the distributions of the errors, providing a more comprehensive understanding of the various metrics outputs compared to a single statistic. Some of the metrics visualize these values in a scalar field or heat-map across the model, allowing for visual inspection. But further computational analysis of local geometric would require a segmentation or clustering step, creating a more complicated process compared to an average global metric output. This could be a prospect for future work.

Conclusion

The outputs of state-of-the-art metrics for evaluating the quality of 3D objects have no direct means of comparison between each other, meaning that their output provides little information without context. The difference in how they are computed also means that each metric only tells us something about the quality of some features of the 3D object and is therefore not comprehensive.
We speculate that no 3D metric will ever comprehensively evaluate all of an object's features, and that application-dependent metrics would be more beneficial to the archiving and heritage preservation field. In this paper, we have interpreted the output of some objective and perceptual metrics relative to some customary post-processing steps. Clearly, some of the metrics can detect the changes introduced by some post-processing step better than others, and the objective and geometric change introduced to the mesh and the perceptually detectable mesh differences can be contextualized by a third cost-metric. Knowledge about this interaction and the metric computation can help in making decisions about a 3D digital object’s quality, especially if it is known if the object will be used for detailed documentation or purely visualization.

References

[1] Storeide, M. S. B., George, S., Sole, A. S., & Hardeberg, J. Y. (2023, June). Statistical Evaluation of 3D Manifolds Shape Retention during Simplification Stages. In Archiving Conference (Vol. 20, pp. 18-22). Society for Imaging Science and Technology.

[2] Storeide, M. S. B., George, S., Sole, A., & Hardeberg, J. Y. (2023). Standardization of digitized heritage: a review of implementations of 3D in cultural heritage. Heritage Science, 11(1), 249.

[3] Lee, Chang Ha, Amitabh Varshney, and David W. Jacobs. "Mesh saliency." ACM SIGGRAPH 2005 Papers. 2005. 659-666.

[4] Lavoué, Guillaume, and Massimiliano Corsini. "A comparison of perceptually-based metrics for objective evaluation of geometry processing." IEEE Transactions on Multimedia 12.7 (2010): 636-649.

[5] Lavoué, G., Gelasca, E. D., Dupont, F., Baskurt, A., & Ebrahimi, T. (2006, August). Perceptually driven 3D distance metrics with application to watermarking. In Applications of Digital Image Processing XXIX (Vol. 6312, pp. 150-161). Spie.

[6] Lavoué, G. (2009). A local roughness measure for 3D meshes and its application to visual masking. ACM Transactions on Applied perception (TAP), 5(4), 1-23.

[7] Silva, S., Santos, B. S., Ferreira, C., & Madeira, J. (2009, July). A perceptual data repository for polygonal meshes. In 2009 Second International Conference in Visualisation (pp. 207-212). IEEE.

[8] Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., & Ranzuglia, G. (2008, July). Meshlab: an open-source mesh processing tool. In Eurographics Italian chapter conference (Vol. 2008, pp. 129-136).

[9] Sorkine, O. (2005). Laplacian mesh processing. Eurographics (State of the Art Reports), 4(4).

[10] Garland, M., & Heckbert, P. S. (1997, August). Surface simplification using quadric error metrics. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques (pp. 209-216).

[11] Cignoni, P., Rocchini, C., & Scopigno, R. (1998, June). Metro: measuring error on simplified surfaces. In Computer graphics forum (Vol. 17, No. 2, pp. 167-174). Oxford, UK and Boston, USA: Blackwell Publishers.

[12] Wu, T., Pan, L., Zhang, J., Wang, T., Liu, Z., & Lin, D. (2021). Balanced chamfer distance as a comprehensive metric for point cloud completion. Advances in Neural Information Processing Systems, 34, 2088-29100.

[13] Batra, R., & Hesselink, L. (1999, October). Feature comparisons of 3-D vector fields using earth mover's distance. In Proceedings Visualization'99 (Cat. No. 99CB37067) (pp. 105-114). IEEE.

[14] Wang, K., Torkhani, F., & Montanvert, A. (2012). A fast roughness-based approach to the assessment of 3D mesh visual quality. Computers & Graphics, 36(7), 808-818.

[15] Sorkine, O., Cohen-Or, D., & Toledo, S. (2003, June). High-pass quantization for mesh encoding. In Symposium on geometry processing (Vol. 42, p. 3).

[16] Gelasca, E. D., Ebrahimi, T., Corsini, M., & Barni, M. (2005, September). Objective evaluation of the perceptual quality of 3D watermarking. In IEEE International Conference on Image Processing 2005 (Vol. 1, pp. I-241). IEEE.

[17] Lavoué, G. (2011, August). A multiscale metric for 3D mesh visual quality assessment. In Computer graphics forum (Vol. 30, No. 5, pp. 1427-1437). Oxford, UK: Blackwell Publishing Ltd.

[18] Vidal, V., Lombardi, E., Tola, M., Dupont, F., & Lavoué, G. (2020, May). MEPP2: a generic platform for processing 3D meshes and point clouds. In EUROGRAPHICS 2020 (Short Paper).

Author Biographies

Markus Sebastian Bakken Storeide is a PhD candidate in Computer Science at the Colourlab. Norwegian University of Science and Technology (NTNU). His PhD project regards the fusion of multiresolution 3D data for digitization of large cultural heritage objects, and the fusion of heterogeneous texture data for additional digitization.

Sony George, PhD is an Associate Professor at the Colourlab, Norwegian University of Science and Technology (NTNU), Norway. His research interests include color imaging, multi/hyper spectral imaging, imaging applications in cultural heritage. He has been involved in several national and EU projects related to cultural heritage imaging, including H2020 MSCA CHANGE-ITN and PERCEIVE.

Aditya Sole completed received his doctoral degree from the Department of Computer Science at the Norwegian University of Science and Technology, Gjøvik, Norway in 2019 and currently is working as an Associate Professor at the computer science department at NTNU Gjøvik campus mainly focusing his research in the field of measuring, understanding, and reproduction of visual appearance and 3D printing. He is also involved in MSCA-ITN-ETN funded ApPEARS project as a deputy scientific co-ordinator and RCN-INTPART funded MANER project as a project manager.

Jon Y. Hardeberg received his PhD degree from Ecole Nationale Supérieure des Télécommunications in Paris, France in 1999. He is now Professor of Colour Imaging at the Colourlab at NTNU - Norwegian University of Science and Technology, Gjøvik, Norway. His current research interests include spectral imaging, image quality, colour management, material appearance, and cultural heritage imaging, and he has co-authored more than 300 publications within the field.