



HAL
open science

Full-reference saliency-based 3D mesh quality assessment index

Anass Nouri, Christophe Charrier, Olivier Lézoray

► **To cite this version:**

Anass Nouri, Christophe Charrier, Olivier Lézoray. Full-reference saliency-based 3D mesh quality assessment index. IEEE International Conference on Image Processing, IEEE, Sep 2016, Phoenix, United States. pp.1007 - 1011, 10.1109/ICIP.2016.7532509 . hal-01413778

HAL Id: hal-01413778

<https://hal.science/hal-01413778v1>

Submitted on 11 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FULL-REFERENCE SALIENCY-BASED 3D MESH QUALITY ASSESSMENT INDEX

Anass Nouri, Christophe Charrier, Olivier Lézoray

Normandie Université, UNICAEN, ENSICAEN, GREYC UMR CNRS 6072, Caen, France

ABSTRACT

We propose in this paper a novel perceptual viewpoint-independent metric for the quality assessment of 3D meshes. This full-reference objective metric relies on the method proposed by Wang *et al.*[1] that compares the structural informations between an original signal and a distorted one. In order to extract the structural informations of a 3D mesh, we use a multi-scale visual saliency map on which we compute the local statistics. The experimental results attest the strong correlation between the objective scores provided by our metric and the human judgments. Also, comparisons with the state-of-the-art prove that our metric is very competitive.

Index Terms— Objective quality assessment, 3D mesh, Visual Saliency, graph.

1. INTRODUCTION

With the development of 3D acquisition techniques, a large quantities of 3D objects are represented mostly in the form of triangular 3D meshes and are used in several human centered applications. This progress, coupled with the fact that human being are strongly based on their vision, require that the 3D meshes representing the targets to analyse are of high quality. A 3D mesh may undergo various processing steps before being presented to a human observer like compression, watermarking, noise acquisition, etc. Thenceforth, assuming that one or more of the degradations cited above are applied to a 3D mesh, an assessment of the quality is necessary to quantify the visual impact of these distortions on the mesh geometry. A first approach for assessing the quality of 3D meshes is to perform subjective evaluations when seeking human opinions. The problem is that this method is slow, tedious and inadequate for real applications. An alternative approach falls within the objective assessment of quality and aims at predicting the quality in an automatic manner. The goal is to design quality metrics that are correlated with subjective scores provided by humans observers. In the literature, such quality assessment approaches are classified into 3 categories: full reference (the original version of the distorted content is fully available for the comparison), reduced reference (partial informations about the original content are available) and no reference (no information is available about the reference content) metrics. Perceptual metrics can play a significant role in several computers graphics applications

like benchmarking 3D mesh processing algorithms, optimizing and assessing performances of compression and restoration approaches, etc. In this context, we propose a metric for 3D mesh quality assessment based on visual saliency called SMQI (Saliency-based Mesh Quality Index). Indeed, visual saliency captures perceptually important regions on which the human visual attention is focalized. Thereby, if perceptually important regions on the mesh surface are distorted, then the global perceived quality is affected and vice-versa (see figure 1). The paper is organized as follows. Section 2 presents the state-of-the-art approaches. Section 3 describes the link between visual saliency and quality assessment. In the same section, we present the proposed metric with an overview of the pipeline and the associated details: multi-scale saliency detection method, roughness map, and definition of the perceptual distance. In section 4, we present experiments on two subjective databases and compare our metric with state-of-the-art approaches.

2. RELATED WORK

Whereas 2D image metrics are extremely developed, the literature is much less important in the field of computer graphics. However, as demonstrated by Rogowitz and Rushmeir [2], 2D metrics aren't effective for 3D meshes since they don't take into account the depth and motion informations. Thus, we quote here only viewpoint-independent metrics that perform directly on the 3D mesh surface. In [3], Corsini *et al.* propose a metric based on the variation of the global roughness measure. The roughness is computed through the variance of the dihedral angles. Lavoué *et al.* [4] propose an extension of the SSIM index proposed for 2D images to the quality assessment of 3D meshes (called Mesh Structural Distorsion Measure: MSDM). Difference of statistics are computed on the curvature maps of the two meshes being compared instead of the pixels intensities that the SSIM index uses. In [5], Lavoué proposes an improvement of MSDM called MSDM2. This time, the metric takes into account the multi-scale aspect and can perform on 3D meshes with different connectivities. Recently, Wang *et al.* [6] proposed a metric based on the variation of local roughness that is derived from the Laplacian of the discrete Gaussian curvature. To the best of our knowledge, there is no metric in the state-of-the-art that considers the visual saliency as basis for 3D mesh quality assessment.

3. THE PROPOSED METRIC

3.1. Visual attention and IQA

Given a 3D object, our visual attention is attracted by particular regions on the surface that are distinct from their surrounding zones. These striking areas, essentially prominent in the field of 3D objects, are content dependent. However, they are not dependent of the behavior or the experience relative to the human observer [1]. This has been proved by a series of subjective experiments where the outcome confirmed that a perceived degradation is highest when the distortion affects a salient region in the content [7]. The same result can be seen in figure 1. Two most outstanding metrics (MSDM2[5] and FMPD[8]) were tested in this example and failed in assessing a content distorted in perceptual salient areas just because they ignore the human visual attention mechanism.

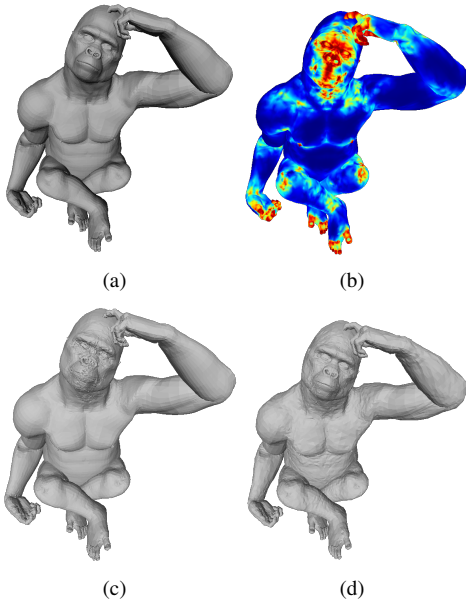


Fig. 1: Comparison of 3D meshes with different perceptual qualities. (a) Original Gorilla 3D mesh. (b) Saliency map of (a) with [9]. (c) Gorilla 3D mesh noised in visual attention areas. FMPD=0.15, MSDM2=0.36, SMQI=0.49. (d) Gorilla 3D mesh noised in less visual attention areas. FMPD=0.54, MSDM2=0.414, SMQI=0.41. (Note that a higher objective quality score denotes a poor quality and vice versa). SMQI is our proposed index.

3.2. Overview of the pipeline

The proposed metric is inspired from the well-known 2D image SSIM [1] index and the MSDM [4] metric for 3D meshes. However, instead of computing local statistics reflecting the structure informations on the image pixels intensities as SSIM-index performs, or on the curvature map as MSDM, we propose to compute a multi-scale saliency map that is used as a basis for computing the statistics over local neighborhoods on 3D meshes. Indeed, we make the assump-

tion that the perceived quality of a mesh is strongly related to the modification of local and global saliency of the mesh surface. Additionally, for the two compared 3D meshes, we use a roughness map on which we compute the differences of mean local roughness of each node. This allows us to capture the visual masking effect that may occur while a rough region is able to hide a geometric distortion. Then we introduce four comparisons functions between corresponding local neighborhoods from two meshes to capture the structure’s differences. Finally we combine them via a weighted Minkowski sum to obtain the final quality score.

3.3. Multi-scale saliency map

In order to detect perceptually important regions on 3D mesh surfaces, we use our multi-scale visual saliency detection approach proposed in [9]. This is a novel multi-scale measure of visual saliency based on an important characteristic of the human visual system (HVS) which is the sensibility to strong fluctuations and discontinuities. We describe here the principal steps of this visual saliency measure. First, we represent a 3D mesh \mathcal{M} by a non-oriented graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{v_1, \dots, v_N\}$ is the set of N vertices and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ the set of edges. The set of edges is deduced from the mesh faces that connect vertices. To each vertex v_i are associated its 3D coordinates $\mathbf{p}_i = (x_i, y_i, z_i)^T \in \mathbb{R}^3$. For a target vertex on the surface mesh, we estimate a 2D tangent plan. Next we consider a spherical neighborhood of radius ε centered on the target vertex and project the neighboring vertices located in the sphere on the 2D plan defined above. Then we define the size of the patch related to the target vertex by computing the maximum distances between 2D projections along the x -axis and the y -axis. Afterwards, we divide the patch into cells and increment them with the neighborhood projection height field. The next step consist of computing the similarities between the patch of a target vertex and the patches of its neighboring vertices that are affected to the weights of edges vertices they belong to:

$$w_{\mathcal{P}}(v_i, v_j) = \exp \left[- \frac{\kappa(v_j) \cdot \|\mathcal{P}(v_i) - \mathcal{P}(v_j)\|_2^2}{\sigma_{\mathcal{P}}(v_i) \cdot \sigma_{\mathcal{P}}(v_j) \cdot \|\mathbf{p}_i - \mathbf{p}_j\|_2^2} \right] \quad (1)$$

where $\mathcal{P}(v_i) \in \mathbb{R}^{l \times l}$ is the vector of accumulated heights into the cells of the patch, $\kappa(v_j)$ is the curvature at the vertex v_j , and $\|\mathbf{p}_i - \mathbf{p}_j\|_2^2$ represents the Euclidean distance between the points of vertices v_i and v_j . We propose to locally compute the scale parameters $\sigma_{\mathcal{P}}(v_i)$. The scale parameter is defined as $\sigma_{\mathcal{P}}(v_i) = \max_{v_k \sim v_i} (\|\mathbf{p}_i - \mathbf{p}_k\|_2)$. As a result, we obtain a single-scale saliency value of the target vertex defined by the mean of its degree:

$$\text{ss-saliency}_{\mathcal{P}}(v_i) = \frac{1}{|v_j \sim v_i|} \sum_{v_i \sim v_j} w_{\mathcal{P}}(v_i, v_j) \quad (2)$$

To enhance the quality of the measured saliency, we compute it at 3 scales by varying the radius ε of the spherical neighborhood while constructing the local patches. The

multi-scale saliency value of a vertex denoted $MS(v_i)$ is defined as the average of single-scale saliencies weighted by their respective entropies. Figure 1(b) presents an example of saliency computation.

3.4. Roughness map and perceptual distance

To compute the local statistics that reflect the structural information of a 3D mesh, we use our multi-scale saliency map as a basis. For a local neighborhood $N(v_i)$ representing the set of adjacent vertices of v_i on the mesh surface, we define the local mean saliency and standard deviation saliency respectively denoted $\mu_{N(v_i)}$ and $\sigma_{N(v_i)}$ as :

$$\mu_{N(v_i)} = \frac{1}{|N(v_i)|} \sum_{v_j \in N(v_i)} MS(v_j) \quad (3)$$

$$\sigma_{N(v_i)} = \sqrt{\frac{1}{|N(v_i)|} \sum_{v_j \in N(v_i)} (MS(v_j) - \mu_{N(v_i)})^2} \quad (4)$$

where $|N(v_i)|$ is the cardinality of $N(v_i)$. For two corresponding local neighborhoods $x = N_{\mathcal{M}_1}(v_i)$ and $y = N_{\mathcal{M}_2}(v_i)$ from two 3D meshes \mathcal{M}_1 and \mathcal{M}_2 , we define the covariance σ_{xy} as follows:

$$\sigma_{xy} = \frac{1}{|x|} \sum_{v_j \in x, y} (MS_{\mathcal{M}_1}(v_j) - \mu_x)(MS_{\mathcal{M}_2}(v_j) - \mu_y) \quad (5)$$

where $MS_{\mathcal{M}_1}$ and $MS_{\mathcal{M}_2}$ respectively represent the multi-scale saliency maps of the two compared meshes \mathcal{M}_1 and \mathcal{M}_2 .

Then, similarly to [4], we define 3 comparisons functions between two corresponding neighborhoods x and y for quantifying the deformation that affect the structural informations of the distorted 3D mesh :

$$L(x, y) = \frac{\|\mu_x - \mu_y\|_2^2}{\max(\mu_x, \mu_y)} \quad (6)$$

$$C(x, y) = \frac{\|\sigma_x - \sigma_y\|_2^2}{\max(\sigma_x, \sigma_y)} \quad (7)$$

$$S(x, y) = \frac{\|\sigma_x \sigma_y - \sigma_{xy}\|_2^2}{\sigma_x \sigma_y} \quad (8)$$

where L , C and S respectively refer to the saliency comparison, the contrast comparison and the structure comparison. Having these 3 comparison functions, we noticed that the visual masking effect on the 3D meshes isn't captured well by these functions when a rough region is present on the reference surface mesh. Indeed, given a rough and a smooth region, a distortion will be much more visible on the smooth region than on the rough one. To deal with this problem, we implemented the work of [8] that provides a roughness map of a 3D mesh based on the Laplacian of the Gaussian curvature. Consequently, we introduced a fourth function based on the comparison of the mean local roughness. The aim of this function is to induce a large difference when a smooth region becomes a rough region and is defined as follows:

$$R(x, y) = \frac{\|\delta_x - \delta_y\|_2^2}{\max(\delta_x, \delta_y)} \quad (9)$$

with $\delta_x = \frac{1}{|x|} \sum_{v_j \in x} \text{RoughnessMap}(v_j)$. It is important to note that a saliency map is different from a roughness map, since only novel and non-redundant informations are pointed on a saliency map. Finally our Saliency-based Mesh Quality Index (SMQI) between two 3D meshes \mathcal{M}_1 and \mathcal{M}_2 is defined by a weighted Minkowsky sum of their local distances:

$$SMQI(\mathcal{M}_1, \mathcal{M}_2) = \left(\frac{1}{|V|} \sum L(x, y) \right)^\alpha + \left(\frac{1}{|V|} \sum C(x, y) \right)^\beta + \left(\frac{1}{|V|} \sum S(x, y) \right)^\gamma + \left(\frac{1}{|V|} \sum R(x, y) \right)^\delta \quad (10)$$

where α, β, γ and δ are obtained from an optimization based on genetic algorithms. Details on the genetic optimization will be given in the experiments section.

4. EXPERIMENTAL RESULTS

4.1. Datasets

To compare our proposed full reference metric with the state-of-the-art methods, two publicly available databases are used: 1) The Liris/Epfl General-Purpose Database [4] and 2) the Liris-Masking database [11]. The first database contains 4 reference 3D meshes. They are affected by two types of distortions: Noise addition and Smoothing. These distortions are applied in 3 different strengths either uniformly over the 3D mesh surface, specifically to rough or smooth regions (for simulating the masking effect) and to transitional areas between rough and smooths regions. In total, 22 distorted 3D meshes of each reference mesh are generated and evaluated by 12 human observers. The Liris Masking Database consists of 4 reference 3D meshes which are distorted by adding noise of three different strengths in either rough and smooth areas to generate 6 distorted versions of each 3D mesh. 12 observers have evaluated the database. The performance of our method is measured by the Spearman Rank Ordered cOrrelation Coefficient (SROOC).

4.2. Results

Before measuring the performance of our proposed metric and since our perceptual distance includes 4 independent parameters (α, β, γ and δ) of which the manual optimization may be ineffective, we choose to use a genetic optimization for tuning these parameters. It's important to note that the number of 3D meshes contained in the two datasets described above is small in comparison with available datasets of 2D images. Consequently, to deal with this weakness, we opted for a Leave-One-Out training on both databases. The goal of this approach is to perform the learning of the model on $k - 1$ observations and to validate it on the k^{th} one. This process is repeated $k \times 999$ times.

Liris/Epfl General Purpose	HD [12][13]	RMS [12][13]	3DWPM1 [3]	3DWPM2 [3]	MSDM2 [5]	FMPD [8]	SMQI
Armadillo	69.5	62.7	65.8	74.1	81.6	75.4	77.5
Venus	1.6	90.1	71.6	34.8	89.3	87.5	91.6
Dinosaur	30.9	0.3	62.7	52.4	85.9	89.6	84.8
RockerArm	18.1	7.3	87.5	37.8	89.6	88.8	91.8
Entire database	13.8	26.8	69.3	49.0	80.4	81.9	84.6

Table 1: SROOC values (%) of different viewpoint-independent metrics on the LIRIS/EPFL General Purpose database

Liris Masking	HD [12][13]	RMS [12][13]	3DWPM1 [3]	3DWPM2 [3]	MSDM2 [5]	FMPD [8]	SMQI
Armadillo	48.6	65.7	58.0	48.6	88.6	88.6	88.6
Lion-vase	71.4	71.4	20.0	38.3	94.3	94.3	83.0
Bimba	25.7	71.4	20.0	37.1	100.0	100.0	100.0
Dinosaur	48.6	71.4	66.7	71.4	100.0	94.3	100.0
Entire database	26.6	48.8	29.4	37.4	89.6	80.2	81.0

Table 2: SROOC values (%) of different viewpoint-independent metrics on the Liris-Masking database

In our case, an observation refers to the MOS values of a reference 3D mesh and its distorted versions. The fitness function used to perform the genetic optimization is defined as: $f(\alpha, \beta, \gamma, \delta) = \sqrt{\sum_{i=0}^{k-1} (MOS_i - SMQI_i(\mathcal{M}_1, \mathcal{M}_2))^2}$ where MOS_i is the vector of MOS values of the observation i and $SMQI_i(\mathcal{M}_1, \mathcal{M}_2)$ is the perceptual distance computed with the equation 10. After genetic optimization, we obtain: $\alpha = 23.63$, $\beta = 3.26$, $\gamma = 5.04$ and $\delta = 0.77$. Note that in figure 1, the Gorilla 3D mesh was assessed with these parameters. Table 1 presents the performance of our metric and the state-of-the-art metrics in term of the Spearman correlation with the subjective scores provided by the Liris/Epfl General Purpose. We can notice that SMQI produces important correlation values for all the 3D meshes and particularly for the RockerArm and Venus 3D meshes where the SROOC values are the highest among the reminder values. Moreover, our perceptual metric SMQI outperforms the 2 best metrics so far MSDM2 and FMPD over the entire database (the SROOC values are 84.6% for SMQI, 81.9% for FMPD and 80.4% for MSDM2). We have also tested and compared our proposed metric with the state-of-the-art metrics on on the Liris-Masking database. Table 2 provides the Spearman correlation values of the different metrics on this database. The choice of the Spearman correlation is motivated by we do not need to verify the normality of the objective scores to compute the correlation with the subjective ones. From these results, three main observations can be made. The first one is that SMQI is very competitive with MSDM2 and FMPD and succeed in capturing the masking effect. The second observation deals with the correlation value associated to the Lion-vase 3D mesh that is slightly lower in comparison with the values of FMPD and MSDM2. This can be explained by the fact that the multi-scale saliency map of the distorted 3D mesh on which are computed the statistics doesn't reflect well the distorted salient areas. Indeed, the reference and the distorted 3D meshes use the same radius (defined

manually) associated to the spherical neighborhoods for the computation of the local adaptive patches for the estimation of the multi-scale saliency. We expect that an automatic method defining the radius will lead to a better estimation of saliency for the distorted 3D mesh. This constitutes a perspective of our future work. The third observation concerns the correlation values on the entire database. We can notice that the correlation values of SMQI and FMPD are lower than the MSDM2 correlation. However in this context and in the contrary of Liris/Epfl General Purpose database, the subjective evaluation protocol used while designing the Liris-Masking database have established the referential range for the rating separately for each 3D mesh and therefore the correlation values over the whole set of 3D meshes are not really meaningful [14]. From the above results and comparisons, it appears that the SMQI metric is strongly correlated to the human perception because of the integration of visual saliency. Additionally, SMQI outperforms the state-of-the-art metrics on the Liris/Epfl General Purpose Database and is very competitive on the Liris-Masking database.

5. CONCLUSION

We have proposed in this paper a new objective metric, called SMQI, for the assessment of 3D mesh visual quality. This perceptual metric compares structural informations of a reference 3D mesh and a distorted one. For this, we use a multi-scale visual saliency map as a basis for computing the local statistics. In order to take into account the masking effect, a roughness map is used to measure the difference of local mean roughness between a pair of meshes. Consequently, four comparison functions are combined via a weighted Minkowski sum to provide an objective score that quantify the visual similarity between two meshes. Experimental results as well as comparisons with the state-of-the-art demonstrate the strong correlation of our approach with the subjective results and its high competitiveness.

6. REFERENCES

- [1] Z. Wang, A.C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE T IMAGE PROCESS*, vol. 13, no. 4, pp. 600–612, 2004.
- [2] Bernice E. Rogowitz and Holly E. Rushmeier, "Are image quality metrics adequate to evaluate the quality of geometric objects?," in *Proc. SPIE*, 2001, vol. 4299, pp. 340–348.
- [3] M. Corsini, E.D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3D mesh quality assessment," *IEEE T MULTIMEDIA*, vol. 9, no. 2, pp. 247–256, Feb 2007.
- [4] G. Lavoué, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," in *Proc. SPIE*, 2006, vol. 6312, pp. 63120L–63120L–12.
- [5] G. Lavoué, "A multiscale metric for 3D mesh visual quality assessment," *Computer Graphics Forum*, vol. 30, no. 5, pp. 1427–1437, 2011.
- [6] K. Wang, F. Torkhani, and A. Montanvert, "Technical section: A fast roughness-based approach to the assessment of 3D mesh visual quality," *Comput. Graph.*, vol. 36, no. 7, pp. 808–818, Nov. 2012.
- [7] F. Boulos, B. Parrein, P. Le Callet, and D. Hands, "Perceptual effects of packet loss on H.264/AVC encoded videos," in *VPQM workshop*, 2009.
- [8] K. Wang, F. Torkhani, and A. Montanvert, "Technical section: A fast roughness-based approach to the assessment of 3D mesh visual quality," *Comput. Graph.*, vol. 36, no. 7, pp. 808–818, Nov. 2012.
- [9] A. Nouri, C. Charrier, and O. Lézoray, "Multi-scale mesh saliency with local adaptive patches for viewpoint selection," *Signal Processing: Image Communication*, vol. 38, pp. 151–166, 2015.
- [10] S. Coren, L. M. Ward, and J. T. Enns, *Sensation and perception*, Wiley, 2003.
- [11] G. Lavoué, "A local roughness measure for 3D meshes and its application to visual masking," *ACM Trans. Appl. Percept.*, vol. 5, no. 4, pp. 21:1–21:23, Feb. 2009.
- [12] N. Aspert, Diego S. Cruz, and T. Ebrahimi, "Mesh: Measuring error between surfaces using the hausdorff distance," in *ICME*, 2002, vol. 1, pp. 705–708.
- [13] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring Error on Simplified Surfaces," *Computer Graphics Forum*, vol. 17, no. 8, pp. 167–174, 1998.
- [14] G. Lavoué and M. Corsini, "A comparison of perceptually-based metrics for objective evaluation of geometry processing," *IEEE T MULTIMEDIA*, vol. 12, no. 7, pp. 636–649, Nov 2010.