3D Blind Mesh Quality Assessment Index

Anass Nouri, Christophe Charrier, Olivier Lézoray - Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, FRANCE.

Abstract

After the sound, 2D images and videos, 3D models represented by polygonal meshes are the actual emergent content due to the technological advance in terms of 3D acquisition [1]. 3D meshes can be subject to several degradations due to acquisition, compression, pre-treatment or transmission that distort the 3D mesh and therefore affect its visual rendering. Because the human observer is generally located at the end of this line, quality assessment of the content is required. We propose in this paper a view-independent 3D Blind Mesh Quality Assessment Index (BMQI) based on the estimation of visual saliency and roughness. Given a 3D distorted mesh, the metric can assess the percived visual quality without the need of the reference content as humans do. No assumption on the degradation to evaluate is required for this metric, which makes it powerful and usable in any context requiring quality assessment of 3D meshes. Obtained results in terms of correlation with subjective human scores of quality are important and highly competitive with existing full-reference quality assessment metrics.

Introduction

Nowadays, with the development of 3D scanners, 3D meshes represent the most emergent content. These latters are used in many fields and applications. Medical industry uses 3D meshes for organ analysis and detailed representation of chemistry components. The architectural domain uses them for the modelization and the visualisation of buildings, bridges, etc. The automative industry adopts 3D meshes for representing new concepts and design cars. 3D meshes find their place also in cinema, video games, fashion and 3D printing which is a prominent application of these last decades. Other futuristic applications like the Holoportation [2] and the 3D mainstream photography [3] benefit from the flexebity of 3D meshes. In this context, it is evident that the quantity and the frequency of exchanges of 3D meshes will increase exponentially. This leads to new challenges concerning the implementation of methods that assess their objective visual quality while taking into account some properties of the human visual system (HVS).

A first approach for assessing the quality of 3D meshes is to perform subjective evaluations when seeking human opinions. However, this method is slow, tedious and inadequate for real time 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: 1) full reference (the original version of the distorted content is fully available for the comparison), 2) reduced reference (partial informations about the original content and the distorted one are available) and 3) no reference (no information is available about the reference content) metrics. In the majority of real time applications manipulating 3D meshes, the reference version (considered as free distorted) of the 3D mesh is not available which makes the objective quality assessment of the 3D content more difficult. This capacity of assessing objects without their reference version is an easy task for humans but this is far from being the case for machines and algorithms. Many quality assessment metrics for 3D meshes were proposed in the state-of-the-art, however they are still limited due to their dependence to the reference version of the 3D mesh. Perceptually-based no-reference/blind quality assessment algorithms can play a significant role in several computers graphics applications such as optimizing and assessing performances of compression and restauration approaches, dynamically adjusting the quality of a monitor, 3D TV or parameters of mesh processing methods in a transmission application, etc. In this context, we propose a no-reference perceptual metric for the quality assessment of a 3D mesh based on saliency and roughness called BMQI (Blind Mesh Quality Index). The paper is organized as follows. Section 2 describes the link between visual saliency, visual roughness and quality assessment. We present in the same section an overview of the pipeline of our approach. In section 3, we present the proposed metric with its associated details: multi-scale saliency detection method, roughness estimation, patch segmentation and regression stage. In section 4, we present the considered subject-rated mesh datasets and analyze the correlation results of the proposed metric with the human subjective scores. Finally, we conclude and point some perspectives of this work in section 5.

The proposed metric

Visual saliency, roughness and quality assessment

The principal challenge met while designing this no-reference quality assessment metric was to select visual features that have the capability to quantify the structural deformation that the 3D mesh undergo and that are correlated with human perception. To do this, we use multi-scale visual saliency and roughness maps. Visual saliency is an important characteristic for human visual attention. Its use in computer graphics applications like mesh quality assessment [4], optimal view point selection [5] and simplification [6] has proven beyond any doubt its correlation with human visual perception. We suppose as in [4] that visual quality of a 3D mesh is more affected when salient regions are affected rather than less or not salient regions. This characteristic was studied in [7][8]. Likewise, variations of 3D mesh roughness appear to be correlated with human perception [9]. Indeed, a roughness map points regions that expose a strong visual masking effect. Regions with high roughness magnitude expose an important degree of visual masking effect since distorsions are less visible on these ones. We show that local variations of saliency

and roughness combined together succeed in assessing the visual quality of a distorted 3D mesh without the need of its reference version.

Method

Given a 3D distorted mesh, we start by computing a multi-scale saliency map MS with our method proposed in [5] and a roughness map R with the method proposed in [10]. Then, we adapt the approach of [11] to our context and segment the 3D mesh into a number of superfacets N_{SF}. In our context, these Superfacets will play the role of local patches since the human visual system (HVS) locally processes information. Once the segmentation is performed, we affect to each vertex v_i of a superfacet SF_j its respective values of saliency $MS(v_i)$ and roughness $R(v_i)$. Afterwards, we construct a feature vector of 4 attributes for each superfacet SF_j:

$$\phi_j = \left[\mu_{SF_j}, \sigma_{SF_j}, \delta_{SF_j}, \gamma_{SF_j}\right] \quad \text{with} \quad j \in [1, N_{SF}]. \tag{1}$$

where μ_{SF_j} and σ_{SF_j} represent respectively the local mean saliency and local standard deviation saliency of the superfacet SF_j and are defined as:

$$\mu_{SF_j} = \frac{1}{|SF_j|} \sum_{v_i \in SF_j} MS(v_i) \tag{2}$$

$$\sigma_j = \sqrt{\frac{1}{|SF_j|} \sum_{v_i \in SF_j} (MS(v_i) - \mu_j)^2} \tag{3}$$

where $|SF_{SF_j}|$ represents the cardinality (*i.e*, the number of vertices) of the superfacet SF_j.

δ_{SF_j} and γ_{SF_j} denote respectively the local mean roughness and the local standard deviation roughness and are defined as:

$$\delta_{SF_j} = \frac{1}{|SF_j|} \sum_{v_i \in SF_j} LRF(v_i) \tag{4}$$

$$\gamma_{SF_j} = \sqrt{\frac{1}{|SF_j|} \sum_{v_i \in SF_j} (LRF(v_i) - \delta_{SF_j})^2} \tag{5}$$

Finally we perform a learning step using the constructed feature vector. This is done using the Support Vector Regression (SVR) [12] that is also used for scoring the visual quality of the 3D mesh. Figure 1 presents the block-diagram of our approach.

Segmentation, learning and regresssion

Superfacets segmentation

One of the novelties of the proposed approach falls within the use of the superfacets - the result of an over-segmentation of the mesh surface into regions whose borders fit well the semantic entities of the mesh - into the pipeline of a mesh quality assessment metric. To segment the mesh, we modified the approach of [11] which, for a 3D mesh $\mathcal{M}$ and a number of desired superfacets, execute the following steps based on the *farthest point* principle:
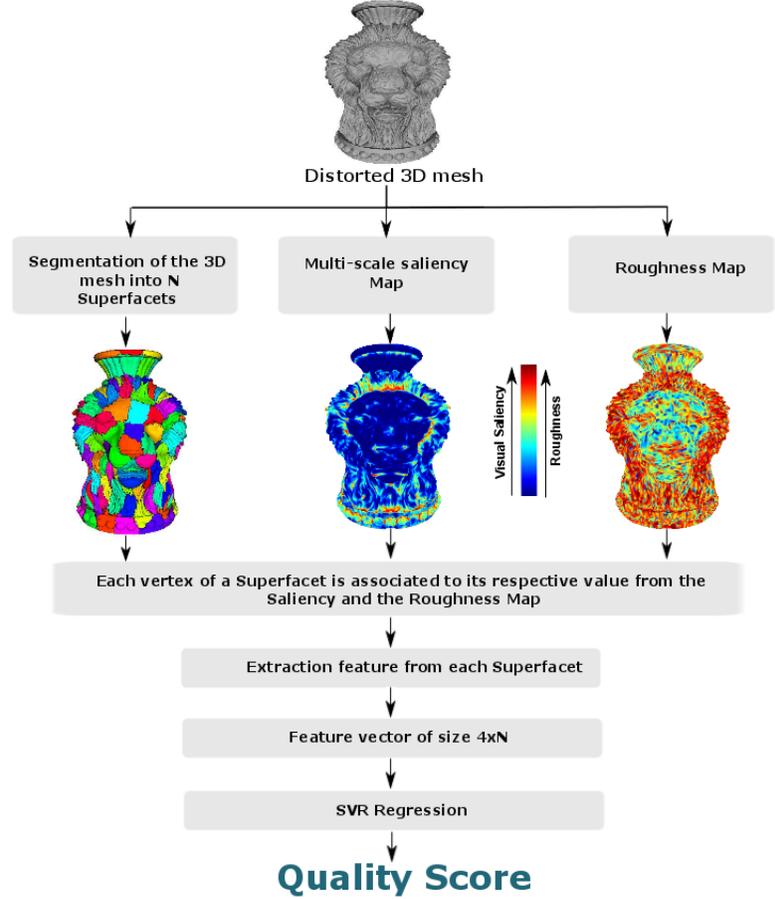
Figure 1: Block diagram of our metric

Initialization : The method begins by associating the center of the first superfacet to the triangle of which the centroid is the nearest to the global centroid of the mesh. Then, each center of a new superfacet is affected to the triangle with the euclidian distance to the latest considered triangle is the highest.

Update of the centers: Once the triangles have been affected to different superfacets, it's necessary to compute the new center of each superfacet. For this, the method computes the mean area of all triangles belonging to a superfacet and associates the new center to the triangle of which the area is the nearest to the computed mean area. If the new center is different from the prior one, the algorithm stops. Otherwise, the classification step is computed.

Classification: For each triangle, the method computes, using the Dijkstra Algorithm, the shortests paths between the centers of the defined superfacets and the triangles of the mesh. When a triangle is considered while computing the shortest path from a superfacet center and if the current computed distance is less than the prior stored one (obtained from the initialization step or from an expansion that started from a different center) then both distance and label associated to the considered triangle are updated (the superfacet that

contains the triangle is fixed).

Geodesic weight: Given two adjacent faces f_i and f_j sharing an edge $e_{i,j}$ with a median point $m_{i,j}$ and two respective centroids c_i and c_j, the geodesic weight is defined as $geo(f_i, f_j) = ||c_i - m_{i,j}|| + ||m_{i,j} - c_j||$. This latter is affected to the weight $w(f_i, f_j)$ of the edge $e_{i,j}$ as follows:

$$w(f_i, f_j) = \frac{geo(f_i, f_j)}{d} \qquad (6)$$

where d is the length of the diagonal of the bounding box including the 3D mesh.

Learning and regression

Even if it's unusual that a human observer associates a quality score in the form of a scalar to a 3D mesh but rather proceeds to a classification of the quality according to the perceived sensation (for exemple: "good" or "bad" quality), nevertheless, the applications context of the quality metrics forces us to provide a single scalar reflecting the perceived quality. For this, we use the extension of Support Vector Machines (SVM) to regression: SVR. The aim is to estimate a function f presenting at most a maximal deviation ε reflecting the dependance between a features vector x_i and an affiliation class y_i. Thus, for a features vector x_i of a distorted 3D mesh $\mathcal{M}_i$ with a subjective quality score y_i, the regression function of an observation x to classify is defined as :

$$f_{SVR} = \sum_{x_i \in V_S} \alpha_i y_i K(x_i, x) + b \qquad (7)$$

where V_S are the support vectors, (x_i, y_i) is the learning set, α is the Lagrange coefficient obtained from a minimization process and $K(x_i, x)$ represents the RBF(Radial Basis Function) kernel defined by:

$$K(x_i, x) = exp(\gamma ||x_i - x_j||^2) \qquad (8)$$

Indeed, the RBF function is often used as a kernel function due to its resemblance to a similarity measure between 2 examples to classify. Also, the motivations related to the use of the SVR are as follows:

1. The regression solution includes a small number of examples x_i (rapidity and efficiency).
2. Results of the regression depend of the used kernel. The test of different kernels is beneficial in so far as the correlation rate between the objective scores and subjective ones may depends on the chosen kernel.

Experimental results

Datasets

In order to compare our no-reference metric with the full-reference methods proposed in the state-of-the-art, we use 2 publicly available subject-rated mesh databases: 1) Liris/Epfl General-Purpose database [13] and 2) Liris-Masking database [14]. The first database contains 4 reference meshes. These are affected with two distorsions: additive noise and smoothing. The distorsions are performed according to 3 strengths on 3 different regions of the surface mesh: 1) uniformly on the surface mesh,

2) specifically on rough or smooth regions and 3) specifically on transitional regions (between rough and smooth regions). In total, 22 distorted meshes of each 3D reference mesh are generated and evaluated by 12 human observers. Figure 2 shows some 3D meshes of the Liris/Epfl General-Purpose database with their respective normalized MOS(Mean Opinion Score).

The second database Liris-Masking consists in 4 reference meshes. 6 distorted versions with the additive noise are generated for each reference 3D mesh. Only rough and smooth regions are considered while distorting the 3D meshes according to 3 strengths to simulate the masking effect. The main goal of this database is to evaluate the masking effect detection capacity of proposed metrics. 12 human observers have assessed the visual quality of the corpus. We present some 3D meshes of the Liris-Masking database with their respective normalized MOS in figure 3.

Performance and analysis

We begin by carrying out a learning step onto the Liris-Masking database in order to determine the RBF kernel's parameters (γ and C which represents the error penalty coefficient) with a 4 parts cross-validation. Each of these parts represents the distorted versions of one of the four reference mesh associated to their subjective quality score. The SVR regression was performed using the LIBSVM [15] library. The selected parameters of the RBF kernel for the Liris-Masking database are: gamma = 0.002 and $C = 32$. For the Liris-Epfl General Purpose database, the selected parameters are : *gamma = 0.005* and $C = 2$ (where C is penalty error).

To evaluate the performance of the proposed metric, we compute the Spearman Rank Ordered Correlation Coefficient (SROCC) between the predicted scores and the subjective human scores of quality provided by the subject rated databases. We present in table 1 the SROOC correlation values of our no-reference metric and the correlation values of 7 full-reference metrics from the state-of-the-art associated to the Liris-Masking database. We can notice that our approach BMQI produces important correlation values for all the 3D meshes without the need of their reference version on the contrary of the full-reference metrics. These results confirm that our metric succeeds very well in taking into account the visual masking effect.

We do not show the correlation over the whole database since the subjective evaluation protocol used while designing the Liris-Masking database has 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 [16].

Table 2 shows the correlation values of our metric associated to the Liris/Epfl General Purpose database. We notice that the performances of BMQI on this database are not as good as those on Liris-Masking database. Indeed, distorsions on the Liris-Epfl General Purpose database (noise addition and smoothing) are applied on 4 distinct regions of the surface mesh (uniform regions, rough regions, smooth regions, and transitional regions). This aims at reflecting the distorsions associated to common mesh processing methods like simplification, compression and watermarking [17] which makes the quality assessment more difficult. From the results presented in table 2, it seems that BMQI assesses
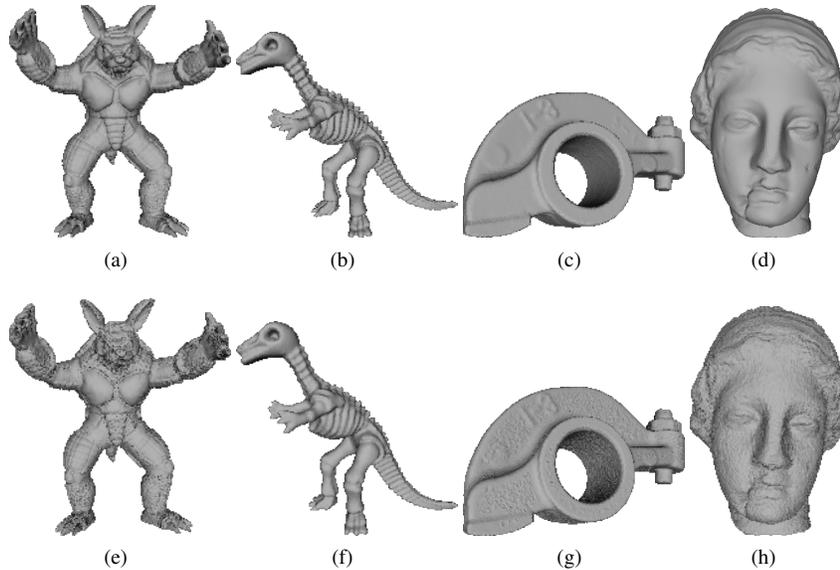
Figure 2: Examples of 3D meshes belonging to the Liris/Epfl General-Purpose database. The top row shows the 4 reference meshes. The second row presents 4 distorted 3D meshes: e) 3D mesh Armadillo affected with noise on rough regions (MOS=0.84), f) 3D mesh Dinosaur uniformly smoothed (MOS=0.43), g) 3D mesh RockerArm affected with noise on smooth regions (MOS=0.75) and h) 3D mesh Venus affected uniformly with noise (MOS=1).

Mesh database	Full-reference								No-reference
Liris Masking	HD	RMS	3DWPM1	3DWPM2	MSDM2	FMPD	TPDM	SMQI	BMQI
Armadillo	48.6	65.7	58.0	48.6	88.6	88.6	88.6	88.6	**94.3**
Lion-vase	71.4	71.4	20.0	38.3	**94.3**	**94.3**	82.9	83.0	**94.3**
Bimba	25.7	71.4	20.0	37.1	**100.0**	**100.0**	**100.0**	100.0	**100.0**
Dinosaur	48.6	71.4	66.7	71.4	**100.0**	94.3	**100.0**	100	83.0

Table 1: SROOC values (%) of different viewpoint-independent metrics on the Liris-Masking database.

the visual quality in a multi-distorsion context with less precision than in a mono-distorsion context even if the correlation values of the three groups of meshes (Dinosaur, Venus and RockerArm) are important. This is mainly due to the correlation value of the sub-corpus Armadillo, this one is lower in comparison to other correlation values. This can be explained by the generated multi-scale saliency map that may not reflect well the distorted salient regions. Thus, the objective quality scores aren't consistant with the quality scores of human observers. The number of superfacets and their size are 2 parameters that could influence the performance of the proposed metric. A precise definition of theses parameters may improve the results. Finally, when we consider the whole corpus of the Liris-Epfl General Purpose database, our approach provides a correlation value relatively low in comparison to the full-reference metrics of the state-of-the-art. This is related on one hand to the low correlation value of the Armadillo subcorpus and on the other hand to the number of 3D meshes considered in the learning step that is very small. Indeed, a corpus consisting in 88 meshes with their associated MOS wouldn't allow the design of an effective quality metric in a multi-distortion context.

In the light of these results, and given the capacity of our approach to assess the perceived quality of a distorted mesh without the need of its reference version, our no-reference metric seems nevertheless competitive with the full-reference methods. For ex-ample, our no-reference approach obtains better correlation rates associated to the Liris-Masking database in comparison to our previous full-reference approach [4]. This could be explained by the segmentation of the mesh into superfacets.

Performance on independant 3D meshes

We have also tested our no-reference metric for the quality assessment of 3D meshes not belonging to any database. This permits to analyse the behavior of our metric when assessing the visual quality of any 3D mesh. Figure 4 presents two reference 3D meshes with their distorted versions. The distorsions considered are: additive noise and simplification. It's important to note that the simplification distorsion wasn't considered in the learning process for selecting the parameters of the RBF kernel since both subject rated databases doesn't include this type of distorsion. In experimentations, we use the selected parameters associated to the Liris/Epfl General Purpose (see subsection Performance Analysis). From the top row of figure 4, we can notice that BMQI provides coherent scores of quality in accordance with human perception. The reference mesh (figure 4(a)) obtains a perceived quality score equal to 6.13. Its noised version obtains a quality score equal to 6.25 and its simplified version (more visually distorted) obtains a quality score equal to 6.74 (note that a low score signifies a good quality score and vice versa). The same remarks could be made to the second row of figure 4.

Mesh database	Full-reference								No-reference
Liris/Epfl G.-Purpose	HD	RMS	3DWPM1	3DWPM2	MSDM2	FMPD	TPDM	SMQI	BMQI
Armadillo	69.5	62.7	65.8	74.1	81.6	75.4	**84.9**	77.5	20.1
Venus	1.6	90.1	71.6	34.8	89.3	87.5	**90.6**	91.6	88.9
Dinosaur	30.9	0.3	62.7	52.4	85.9	89.6	**92.2**	84.4	83.5
RockerArm	18.1	7.3	87.5	37.8	89.6	88.8	92.2	91.8	**92.7**
Base entire	13.8	26.8	69.3	49.0	80.4	81.9	**89.6**	84.6	78.1

Table 2: SROOC values (%) of different viewpoint-independent metrics on the LIRIS/EPFL Genaral Purpose database.

(a)   (b)

(c)   (d)
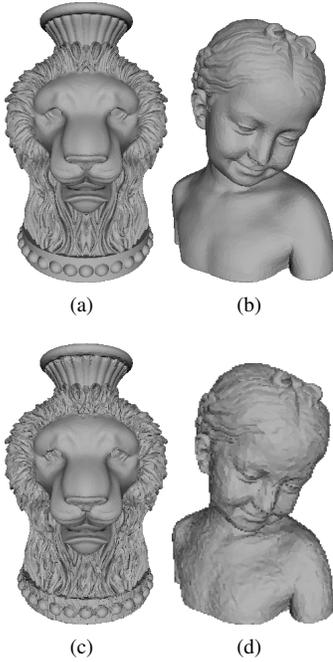
Figure 3: Examples of 3D meshes belonging to the Liris-Masking. The top row shows 2 reference 3D meshes. The second row presents 2 distorted 3D meshes: c) 3D mesh Vase-Lion affected with noise on rough regions (MOS=0.20) and d) 3D mesh Bimba affected with noise on smooth regions (MOS=1.0).

(a) BMQI=6.13   (b) BMQI=6.25   (c) BMQI=6.74

(d) BMQI=5.97   (e) BMQI=6.56   (f) BMQI=7.20
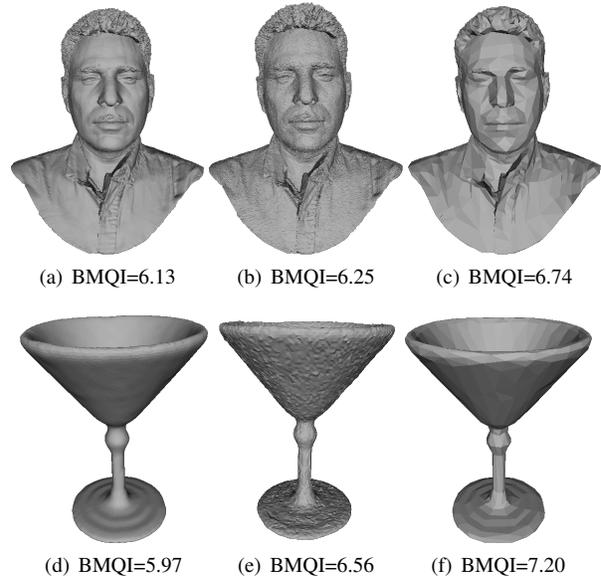
Figure 4: Results of our metric (BMQI) on 3D meshes not belonging to any database used in the learning or the regression step : a) Reference 3D mesh representing a scan of a head (276402 faces) b) distorted version of (a) with an additive noise, c) simplified version of (a) with QSLIM [18] (5000 faces), d) reference 3D mesh representing a glass(15006 faces), e) distorted version of the (d) with an additive noise and f) simplified version of (d) with QSLIM [18] (1000 faces).

Conclusion

In this paper we have proposed an approach to address the difficult problem of the blind quality assessment of 3D meshes. This new index uses simple characteristics computed on a multi-scale saliency and roughness maps in order to asses the perceived quality of a 3D mesh without the need of its reference version. The good performance in terms of correlation with humans judgments proofs that our measure is competitive with full reference metrics. Our future work consists on enhancing both the multi-scale saliency maps and the learning/regression step. Indeed, we believe that the subject rated databases have to be more important in terms of dimension to provide an optimal learning and thus leading to a better prediction of quality in a context of different types of distorsions. Another improvement would consist in considering the multi-scale aspect from the superfacets sizes.

References

[1] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz. Coding algorithms for 3dtv x2014-a survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11):1606–1621, Nov 2007.

[2] Holoportation Microsoft. `https://www.microsoft.com/en-us/research/project/holoportation-3/`,, 2016.

[3] K. Kolev, P. Tanskanen, P. Speciale, and M. Pollefeys. Turning mobile phones into 3d scanners. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3946–3953, June 2014.

[4] A. Nouri, C. Charrier, and O. Lézoray. Full-reference saliency-based 3d mesh quality assessment index. In *International Conference on Image Processing (IEEE)*, 2016.

[5] A. Nouri, C. Charrier, and O. Lézoray. Multi-scale mesh saliency with local adaptive patches for viewpoint selection. *Signal Processing: Image Communication*, 38:151–166, 2015.

[6] Chang Ha Lee, Amitabh Varshney, and David W. Jacobs. Mesh saliency. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 659–666, New York, NY, USA, 2005.

ACM.

[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] U. Engelke, R. Pepion, P. Le Callet, and H Zepernick. Linking distortion perception and visual saliency in h.264/avc coded video containing packet loss. In *Proc. SPIE*, 2010.

[9] K. Wang, F. Torkhani, and A. Montanvert. Technical section: A fast roughness-based approach to the assessment of 3D mesh visual quality. *Comput. Graph.*, 36(7):808–818, November 2012.

[10] K. Wang, F. Torkhani, and A. Montanvert. Technical section: A fast roughness-based approach to the assessment of 3D mesh visual quality. *Comput. Graph.*, 36(7):808–818, November 2012.

[11] Patricio Simari, Giulia Picciau, and Leila De Floriani. Fast and scalable mesh superfacets. *Comput. Graph. Forum*, 33(7):181–190, October 2014.

[12] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[13] G. Lavoué, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi. Perceptually driven 3D distance metrics with application to watermarking. In *Proc. SPIE*, volume 6312, pages 63120L–63120L–12, 2006.

[14] G. Lavoué. A local roughness measure for 3D meshes and its application to visual masking. *ACM Trans. Appl. Percept.*, 5(4):21:1–21:23, February 2009.

[15] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[16] G. Lavoué and M. Corsini. A comparison of perceptually-based metrics for objective evaluation of geometry processing. *IEEE T MULTIMEDIA*, 12(7):636–649, Nov 2010.

[17] G. Lavoué, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi. Perceptually driven 3D distance metrics with application to watermarking. In *Proc. SPIE*, volume 6312, pages 63120L–63120L–12, 2006.

[18] Michael Garland and Paul S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 209–216, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

Author Biography

Anass Nouri received both the M.Sc. degree in intelligent systems and imaging from the IBN TOFAIL-University, Kenitra, Morocco, and the M.Sc. degree in telecommunications and image processing from the University of Poitiers, France in 2013. He obtained his Ph.D. in computer science from the University of Normandie in 2016. From 2016 to 2017, he was an assistant professor in the Computer Science department of the national school of engineering of Caen (ENSICAEN). His current research interests include 3D mesh processing, quality assessment of 3D meshes/2D digital images, and computational vision.

Christophe Charrier, received the M.Sc. degree from the Nantes University of Science and Technology, Nantes, France, in 1993, and the Ph.D. degree from the University Jean Monnet, Sainttienne, France, in 1998. He has been an Associate Professor with the Communications, Networks and Services Department, Cherbourg Institute of Technology, Cherbourg, France, since 2001. From 1998 to 2001, he was a Research Assistant with the Laboratory of Radio Communications and Signal Processing, Laval University, Quebec, QC, Canada. In 2008, he was a Visiting Scholar with the Laboratory for Image and Video Engineering, University of Texas, Austin. From 2009 to 2011, he was an Invited Professor with the Computer Department, University of Sherbrooke, Sherbrooke, QC, Canada. His current research interests include digital image and video coding, processing, quality assessment, and computational vision.

Olivier Lézoray received the M.Sc. and doctoral degrees in computer science from University of Normandie, France, in 1996 and 2000, respectively. From 1999 to 2000, he was an assistant professor in the Computer Science Department at the same university. From 2000 to 2009, he was an associate professor at the Cherbourg Institute of Technology in the Communication Networks and Services Department. Since 2010, he has been a full professor at the Cherbourg Institute of Technology. His research is focused on color image segmentation and filtering (graph-based variational and morphological methods) and machine-learning techniques for image mining (neural networks and support vector machines).