Evaluating the local visibility of geometric artifacts

Jinjiang Guo, Vincent Vidal, Atilla Baskurt, Guillaume Lavoué*
LIRIS, Université de Lyon, CNRS, France

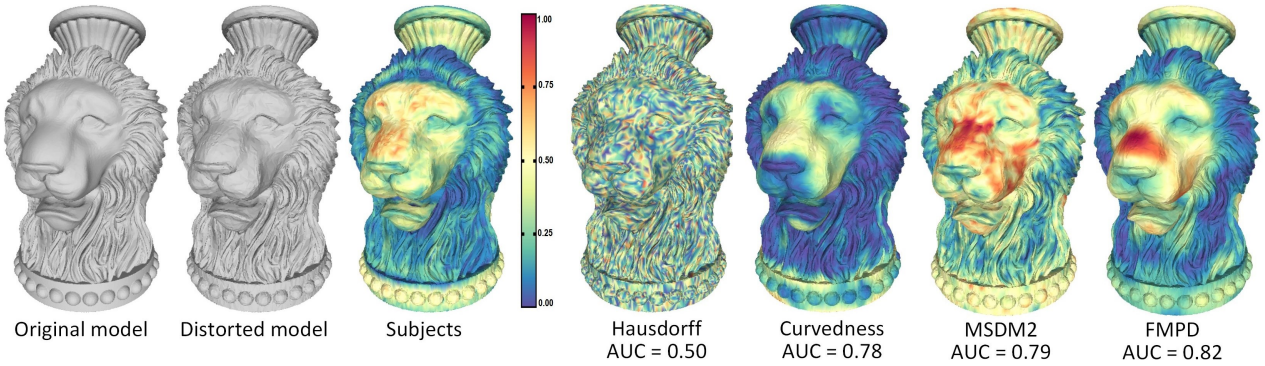

Figure 1: *Reference and distorted models from our dataset (stimulus #9) along with human-perceived distortions (color-coded probabilities of artifact detection) and distortion maps from different metrics (colors mapped in the min-max range). The distortion prediction performance of each metric is given in term of Area Under ROC Curve (AUC).*

Abstract

Several perceptually-based quality metrics have been introduced to predict the global impact of geometric artifacts on the visual appearance of a 3D model. They usually produce a single score that reflects the global level of annoyance caused by the distortions. However, beside this global information, it is also important in many applications to obtain information about the *local visibility* of the artifacts (i.e. estimating a localized distortion measure). In this work we present a psychophysical experiment where observers are asked to mark areas of 3D meshes that contain noticeable distortions. The collected per-vertex distortion maps are first used to illustrate several perceptual mechanisms of the human visual system. They then serve as ground-truth to evaluate the performance of well-known geometric attributes and metrics for predicting the visibility of artifacts. Results show that curvature-based attributes demonstrate excellent performance. As expected, the Hausdorff distance is a poor predictor of the perceived local distortion while the recent perceptually-based metrics provide the best results.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—;

Keywords: Geometric artifact, visual quality metric, noticeable distortion, perceptual experiment.

*e-mails:{jinjiang.guo,vincent.vidal,atilla.baskurt,glavoue}@liris.cnrs.fr

1 Introduction

3D meshes are subject to a wide range of processes which include compression, simplification, filtering, watermarking and so on. These processes inevitably introduce distortions which alter the geometry of these 3D data and thus their final rendered appearance. Classical metrics such as Hausdorff distance and root mean square error have proven to be very poor predictor of the perceptual annoyance caused by geometric artifacts [Lavoué and Corsini 2010]. As a consequence, mesh visual quality (MVQ) metrics have been recently introduced by the scientific community [Váša and Rus 2012; Wang et al. 2012; Lavoué 2011], mostly inherited from image quality assessment metrics [Wang and Bovik 2006]. Their objective is to evaluate the visual impact of graphics artifacts (e.g. geometric quantization noise, smooth deformations due to watermarking, simplification artifacts and so on). A comprehensive review has been recently published about 3D mesh quality assessment [Corsini et al. 2013]. These quality metrics provide good results in term of correlation with the human opinion, however they are mostly specialized in outputting one global quality score which predicts a global level of annoyance caused by all artifacts present in the 3D model. Such global quality index is relevant for many computer graphics applications, for instance, to evaluate the rate-distortion performance or to reduce/augment the quantization precision in the case of a compression algorithm. However, beside this global information, it is also important in many cases to obtain an information about the local *visibility* of the artifacts (i.e. a per-vertex measure of the perceived distortion). Such local information is useful to drive locally any geometry processing algorithms (e.g. simplification, remeshing, filtering and so forth). While some metrics [Lavoué 2011; Wang et al. 2012] do produce such local distortion maps, they have never been evaluated quantitatively with regards to this task.

In that context we propose a new ground-truth dataset of localized distortion maps, obtained through a subjective experiment. For this purpose we gather a set of 3D models associated with several types of distortions (watermarking, quantization, non-uniform noise, smoothing and simplification). In the experiment, we ask the observers to mark vertices of the distorted models where they perceive noticeable differences as compared with the reference ones (using a brush painting interface). After an analysis of the inter-observer agreement, we show that the obtained results illustrate

some perceptual mechanisms of the human visual system (masking effect and frequency sensitivity). We then use this dataset to perform quantitative analysis and comparison of the performance of several geometric attributes, as well as recent perceptual metrics, for the task of local visible difference prediction.
The rest of this paper is organized as follows: section 2 describes the related work about visual quality assessment of 3D graphics and local artifact visibility metrics. Then, section 3 presents our subjective experiment. Finally, section 4 details the tested attributes and metrics while section 5 presents the results of our study.

2 Related work

In this section, we provide a brief survey of existing 3D mesh visual quality metrics; for more details we refer the reader to [Corsini et al. 2013]. We also present image quality metrics that focus on localized distortion prediction.

2.1 3D mesh visual quality metrics

Existing perceptually-motivated metrics are all *full reference*, i.e. they compare the distorted and original 3D models to compute a global quality score. The first authors who tried to incorporate some perceptual insights to improve the reliability of geometric distortion measures were Karni and Gotsman [2000] who proposed combining the Root Mean Square (RMS) distance between corresponding vertices with the RMS distance of their Laplacian coordinates (which reflect a degree of smoothness of the surface). Lavoué [2011] proposed a metric based on local differences of curvature statistics, while Vasa and Rus [2012] considered the dihedral angle differences. These metrics consider local variations of attribute values at vertex or edge level, which are then pooled into a global score. In contrast, Corsini et al. [2007] and Wang et al. [2012] compute global roughness values per model and then derive a simple global roughness difference. Similarly to recent image quality metrics, some of these latter algorithms [Váša and Rus 2012; Wang et al. 2012] integrate perceptually motivated mechanisms such as visual masking. While these methods have been primarily designed for global quality prediction, some of them [Lavoué 2011; Wang et al. 2012] are also able to output a localized distortion map of which accuracy has, however, never been evaluated objectively.
As can be seen above, existing metrics are based on different geometric attributes: pure geometric distance, Laplacian coordinates, different kinds of curvatures, dihedral angles and so forth, which are then subject to filtering, weighting and local or global comparisons. There is actually no consensus on the most perceptually important attributes. One of the contribution of the proposed work is precisely to determine the best set of attributes and filters for the task of visible difference prediction.

2.2 Image quality metrics for artifact localization

The idea of predicting localized artifacts has been already studied for 2D images. One of the pioneer works is the Visible Difference Predictor (VDP) from Daly [1993] which outputs a probability-of-detection map. This complex metric includes a number of processing stages related to the mechanisms of the human visual system (such as the contrast sensitivity function). This metric has been extended for high dynamic range images by Mantiuk et al. [2011]. The well known SSIM (Structural SIMilarity) index, introduced by Wang and Bovik [2004], also outputs a distortion map. Recently, several authors [Herzog et al. 2012; Cadik et al. 2013] have focused on detecting the visibility of rendering artifacts in synthetic images (e.g. structured noise from approximate global illumination or halo artifacts from tone mapping) after having observed than general-purpose metrics (VDP, HDR-VDP-2, SSIM) fail for this task [Cadík et al. 2012]. The subjective experiments of Čadík et al. [2012] are quite related to the proposed work. They conducted experiments where they ask observers to mark image pixels with noticeable/objectionable distortions in the presence/absence of a reference image. This study was inspiring for us and we elaborate on the same idea. However, we tackle the problem of 3D mesh quality at the vertex level, while they consider image quality at the pixel level. Note that we did not consider the no-reference task, since it would have been too difficult in the case of 3D models (as explained in the next section).

3 Localized distortion experiment

The objective of this experiment is to study the visibility and annoyance of geometric artifacts localized on the surface of 3D objects. More specifically, we ask observers to mark areas of 3D objects that contain noticeable distortions, regarding a reference. Results of this study provide insights on the perceptibility of certain artifacts and will be used as well to evaluate the performance of existing metrics and geometric features to predict this visibility.

3.1 Experimental design

As raised in [Corsini et al. 2007], the design of subjective quality evaluation experiment involving 3D content requires the choice of many parameters (e.g. lighting, material, background, level of interaction). Several authors have made such 3D subjective quality experiments [Váša and Rus 2012; Corsini et al. 2007; Watson et al. 2001; Rogowitz and Rushmeier 2001]; whereas none of them tackled the task of local artifact marking, we inspired from these works to design our protocol as follows:

- We consider a *full-reference* experiment, i.e. the observers see the distorted 3D model to mark, together with the reference not-distorted one. A no-reference task would have been too difficult since an observer, even expert, cannot have a sufficient *a priori* on a 3D shape to be able to notice that a region presents an artifact without knowing the reference.
- We consider a *full interaction* scenario (the use of static viewpoints is not recommended [Rogowitz and Rushmeier 2001]).
- To minimize the influence of the light and the camera positions on the perception of artifacts, we simply consider a front directional lighting without specular reflections.
- As in [Corsini et al. 2007], we chose a non-uniform background in order to minimize the influence of the silhouette.
- We provide to the user a mesh painting tool, with a customizable brush size.

Our interface is built upon MeshLab [Cignoni et al. 2008].

3.2 Stimuli creation

The choice of an appropriate set of stimuli for a subjective quality experiment is critical. The objective of our study is to challenge the ability of existing metrics to detect *complex artifacts* (e.g. smooth structured patterns) in *complex scenarios* (e.g. near the visibility threshold). Indeed, even a basic Euclidean geometric distance is able to predict the visibility of a too obvious artifact like a sharp geometric crease added on a sphere. Besides this need of *complexity*, our dataset must also present a sufficient degree of *diversity* to allow us to draw significant conclusions regarding the performance of the metrics. Finally, given the time necessary for the painting task, the dataset has to remain of *limited size*. Indeed, it is not realistic to ask people to paint 80 models, which is the typical size of datasets used

for global quality assessment [Váša and Rus 2012; Lavoué 2011]. For all these reasons, we selected a small set of surface models exhibiting very different shapes, containing both smooth and rough parts, and associated with different sampling densities (see table 1 for details). For the distortions, we have considered several realistic processing operations:

- **Watermarking** - Three different algorithms are used, they respectively modify the volume moments [Wang et al. 2011], the low frequency spectral coefficients [Wang et al. 2009] and the vertex norms [Cho et al. 2007] (i.e. their distance to the centroid of the model). The distortions are respectively smooth deformations and structured ring patterns.
- **Compression** - We consider uniform geometric quantization, the most common lossy process of compression algorithms.
- **Simplification** - We use the QEM algorithm [Garland and Heckbert 1997].

On top of these realistic distortions we also consider **noise addition** and **smoothing** [Taubin 1995], in order to emulate further complex processes. These distortions are applied in a local manner on selected regions of the object, either rough or semi-rough (see objects #4 to #8). As stated above, the idea is to create a compact set of challenging and diversified stimuli (11 in total). We still have introduced a certain amount of consistency: we conducted the same attack (noise addition) on different models and different attacks on the same model (Egea), in order to be able to draw conclusions about these factors.

3.3 Participants and procedure

The participants were shown each distorted object next to its reference version, side by side on the same screen and with the same size. They were able to fully interact with the models (rotation, zoom and translation) and to change the size of the painting brush. They were instructed to mark the regions (on the distorted model) were they could see any visible difference with respect to the reference one, and were given no time limit. The 11 pairs of models were presented in random order to prevent any order bias. Before starting the test session, each participant was subject to a training phase to make him familiar with 3D object manipulation and with the mesh painting tool.
Three different sizes of screen were used: 14, 19 and 22 inches. Given the fact that each pair of stimuli (distorted and reference) spanned the full screen, each stimuli thus subtended at least 15 degree of visual angle horizontally. 20 participants took part to the experiment (3 females and 17 males), they were aged from 16 to 27 years and had normal or corrected vision. They were students and staffs from the university of Lyon in France and from the Northwestern Polytechnical University of Xi'an in China, with no expertise in geometry processing. No differences were observed due to gender or screen size.
The time to mark all 11 objects was 90 minutes on average (with a standard deviation of 30 minutes). This may seem long (e.g. 30 minutes is usually recommended for a rating experiment [Recommendation ITU-R BT. 500-11 2002]), however observing and painting a small number of models is far less tiring that rating hundreds of stimuli. It is interesting to observe that these painting times highly depend on the objects: 5 minutes on average were necessary to mark each simplified Bunny object (#10 and #11) while an average of 12 minutes was taken for the Vase Lion (#9). These times seem to depend both on the object shape and attack complexity, but interestingly these values remain consistent among the observers.
An example of distortion map obtained from the human subjects is shown in figure 2 (object #8). It results from averaging the 20 observer's binary maps and reflects the local probabilities of artifact detection. In this example, more than 50% of the observers have detected the geometric noise added on the neck and the ribs of the dinosaur.

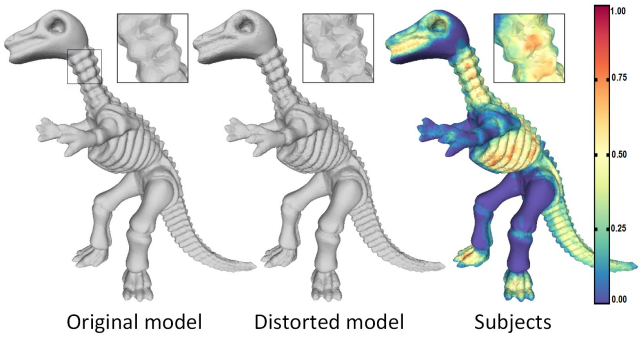

Figure 2: *Example reference and distorted models from our dataset (object #8), along with human-perceived distortions (color-coded probabilities of artifact detection).*

3.4 Analysis

3.4.1 Observer agreement

Before exploiting the results of our experiment, it is critical to analyze the agreement between participants. To assess this inter-observer agreement we first consider the Kendall's coefficient of agreement [Kendall and Babington Smith 1940]. This coefficient, noted as u, is usually used for assessing the consistency of votes in paired comparison studies. It ranges from 1 (perfect agreement) to $-1/(n-1)$, if n is even, and $-1/n$, if n is odd, where n is the number of observers. In their experiment related to image local distortion assessment, Čadík et al. [2012] computed this coefficient per pixel; following this idea we compute it per vertex by considering the number of participants that have chosen the *distorted* choice over the *not-distorted* one and vice versa. We can then derive an average coefficient $\overline{u}$ per object. Still as in [Cadík et al. 2012], and considering the fact that the high number of unmarked vertices will push the coefficient toward high values, we also compute $\overline{u_{mask}}$ for which we average only vertices marked as distorted by at least 5% of the observers. Results are detailed in table 2. The values of $\overline{u}$ and $\overline{u_{mask}}$ averaged over all the objects are respectively 0.60 and 0.43. That constitute good agreement values very close to those from Čadík et al. [2012] (0.78 and 0.41).
We could consider that these values sufficiently assess the agreement of the observers and thus the reliability of our experimental results. However, computing the agreement for each vertex separately may lose some important information about the randomness of the observer's choices. As an illustration, imagine that each observer votes for a different subset of the mesh vertices, then $\overline{u}$ (as well as $\overline{u_{mask}}$) will be high since all the vertices will be chosen as unmarked by n or $n-1$ observers. This high $\overline{u}$ value may be considered as correct because observers mostly agree on the fact that no vertex seems distorted. However, in that case we could also consider that their votes are random and thus that their agreement should be low. To face this issue we also computed a more global statistical agreement indicator: the Krippendorff's α coefficient [Krippendorff 2004]. This indicator is able to measure the agreement among an arbitrary number of observers assigning values to unstructured phenomena. It ranges from 0 (observers agree as if chance had produced the results) to 1 (observers agree perfectly). For each object, we do not compute a single α value but the distribution of α by bootstrapping (as recommended by the author

ID	Original Model (vertex number)	Distortion type	Method & ref.	Settings	MRMS (10^{-3})
#1	Egea (100K)	Watermarking	Volume moment [Wang et al. 2011]	75 bits inserted, α=0.7	0.85
#2	Egea (100K)	Watermarking	Spectral [Wang et al. 2009]	16 bits inserted	0.59
#3	Egea (100K)	Watermarking	Vertex norms [Cho et al. 2007]	64 bits inserted, method I	0.08
#4	Egea (50K)	Smoothing	Taubin [Taubin 1995]	30 iterations, on rough areas	0.33
#5	Egea (50K)	Noise	Uniform	$\alpha = 0.9\%$, on rough areas	0.16
#6	Egea (50K)	Noise	Uniform	$\alpha = 0.9\%$, on intermediate areas.	0.16
#7	Egea (50K)	Smoothing	Taubin [Taubin 1995]	30 iterations, on intermediate areas	0.26
#8	Dinosaur (42K)	Noise	Uniform	$\alpha = 0.75\%$, on intermediate areas	0.17
#9	Vase Lion (39K)	Quantization	Uniform	9 bits	0.33
#10	Bunny (25K)	Simplification	Garland & Eckbert [1997]	50% removed	0.15
#11	Bunny (25K)	Simplification	Garland & Eckbert [1997]	80% removed	0.29

Table 1: *Details of our dataset. MRMS stands for Maximum Root Mean Square error, calculated using the Metro tool [Cignoni et al. 1998].*

[Krippendorff 2004]) and thus we obtain a mean value and the associated confidence interval. Table 2 details these mean values as well as the 5^{th} and 95^{th} percentiles. The α averaged over all the object is 0.102. This is a relatively low value suggesting that the task was indeed difficult. However, the percentiles show that these values are significant so there really exists an agreement among the observers, assessing that their task has been reliably performed. Local distortion marking is still a very unusual task in subjective quality assessment, hence it is hard to judge what can be considered as a typical α value for this kind of data. Even if two observers totally agree, they will not exactly mark the same vertices. Hence, agreement values cannot meet the level of classical rating or paired comparison experiments. What we can assert is that our observer agreement is very similar to the one observed in the image localized distortion experiment from Čadík et al. [2012] (this is assessed by the similar $\overline{u}$ and $\overline{u_{mask}}$ values and illustrated, in the supplementary material, by the similarity in the obtained maps). We argue that the Krippendorff's α should become a gold standard to evaluate the observer agreement for localized distortion experiments.

When looking at the per-object agreement in table 2, we can observe that it varies from 0.011 to 0.218 (Krippendorff's α). These values directly reflect the difficulty for the observers to perceive the artifacts. For instance, the noise added on the Egea model (#5 and #6) is quite easy to detect ($\alpha > 0.20$). On the contrary, a smooth deformation due to watermarking (#1) is far more difficult to perceive ($\alpha = 0.029$). These agreement values allow us to detect two problematic cases: object #2 owns a very high $\overline{u}$ and a very low α, and object #11 owns both low $\overline{u}$ and α values. The artifacts of object #2 are of very low frequency (see figure 3) and almost invisible, thus observers have marked very few and almost random vertices. On the contrary object #11 is too much simplified, hence observers have painted almost all vertices. These two special cases will thus be removed in some of our evaluation experiments.

3.4.2 Observations

The subjective distortion maps constitute great data for further studies about the perceptual mechanisms involved in the visualization of 3D graphical content. While this investigation is not in the scope of this paper, we provide here a quick illustration of two well known features of the human visual system: the *visual masking effect* and the *frequency sensitivity*.

Visual masking defines the reduction in the visibility of one stimulus due to the simultaneous presence of another. For 3D surfaces, this phenomenon points out the maximized visibility of geometric distortions on smooth regions rather than on rough ones. This effect is perfectly illustrated by object #9: whereas the same uniform 9-bits quantization is applied for all vertices, we can observe that the subjective artifact probability is much higher on the face of the Lion which is rather smooth, than on the mane which is rough (see figure 1).

ID	$\overline{u}$	$\overline{u_{mask}}$	Kripp. α [5^{th},95^{th}]
#1	0.901	0.719	0.029 [0.029;0.030]
#2	0.952	0.761	0.011 [0.010;0.012]
#3	0.806	0.604	0.057 [0.056;0.058]
#4	0.705	0.435	0.127 [0.126;0.128]
#5	0.546	0.292	0.218 [0.217;0.220]
#6	0.537	0.282	0.206 [0.205;0.208]
#7	0.649	0.366	0.177 [0.175;0.178]
#8	0.483	0.337	0.162 [0.161;0.164]
#9	0.320	0.310	0.082 [0.081;0.083]
#10	0.537	0.498	0.032 [0.030;0.034]
#11	0.181	0.132	0.022 [0.020;0.025]
Mean	0.602	0.430	0.102 [0.101;0.104]

Table 2: *Inter-observer agreement statistics, in terms of Kendall's u and Krippendorff's α coefficients.*

The human visual system is also sensitive to the frequency of a visual stimulus. Studies of this phenomenon have led to define the *contrast sensitivity function (CSF)*, characterized by a band-pass filter with a peak frequency between 4 and 6 cpd (cycles by degree) and a quick drop on each side of this peak. For 3D shapes, this generally leads to observe that high-frequency distortions have a much higher probability to be visually noticeable than low-frequency ones. This *frequency sensitivity* is particularly well illustrated when looking at objects #2 and #3. Their distortions are respectively of very low and very high frequency (see figure 3). Whereas the *power* of the distortion is much higher in object #2 (as reflected by the MRMS values - 0.59 against 0.08), the number of vertices marked as *distorted* is much lower: on average observers have marked 1.2% of the vertices against 5.4% for object #3. This is also illustrated by the observer maps in figure 3.

4 Selection of geometric attributes and metrics

Our subjective dataset provides the opportunity to analyze the performance of existing metrics and well-known geometric attributes for the task of visual difference prediction. We implemented the 12 features detailed below.

- **Curvatures** - We compute the minimum (K_1), maximum (K_2), mean ($(K_1 + K_2)/2$) and Gaussian ($K_1 \times K_2$) curvatures. To ensure the stability of the four curvature fields, as in [Lavoué 2011], we evaluate the curvature tensor using the method from

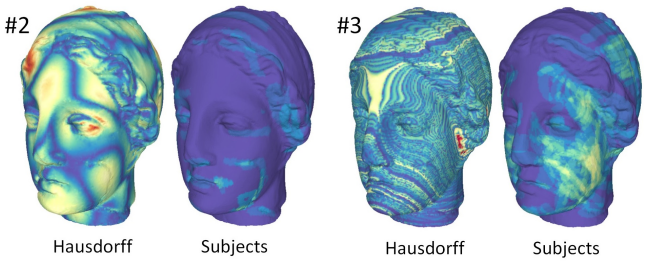

Figure 3: *Distorted models #2 (low frequency watermark) and #3 (high frequency watermark). We present their Hausdorff and observer maps. High frequency distortions have been much more perceived by the observers.*

[Cohen-Steiner and Morvan 2003] on a fixed-size neighborhood around each vertex, *i.e.* a geodesic disk approximated by the intersection of the surface with a sphere centered at the vertex.

- **Shape Index and Curvedness** introduced by Koenderink and van Doorn [1992] - These fields are computed as follows:
  $SI = 2/\pi * arctan[(K_2 + K_1)/(K_2 - K_1)]$
  $Curv = \sqrt{(K_1^2 + K_2^2)/2}$.
- **Geometric position**.
- **Dihedral angle** - It is computed for each edge and corresponds to the angle between the normals of its adjacent faces. The computed values are then averaged for each vertex. This attribute was used in the perceptual metric from Váša and Rus [2012].
- **Normal** - This attribute was considered in several simplification algorithms such as VSA [Cohen-Steiner et al. 2004].
- **Mesh Saliency** - We have implemented the method from Lee et al. [2005], based on a difference of Gaussian operator applied on the mean curvature field.
- **Geometric Laplacian** - The relevance of the geometric Laplacian regarding the perceived artifact visibility was raised by Karni and Gotsman [2000]. As the authors, we consider the graph Laplacian operator.
- **Laplacian of Gaussian curvature** - It corresponds to the Laplacian operator applied on the Gaussian curvature field. This attribute was used in the perceptual metric from Wang et al. [2012]. As the authors, we consider here the *cotangent weights* discretization of the Laplace operator.

Geometric positions, normals and geometric Laplacian are 3D vectors, while the others are scalar values. Besides most simple geometric features (positions and normals), we selected attributes either commonly used in mesh quality metrics [Váša and Rus 2012; Wang et al. 2012; Lavoué 2011], mesh analysis tools [Silva et al. 2009] or pointed as *relevant* by previous works related to perception [Lee et al. 2005; Chen et al. 2012].
Our objective is to predict the local visual distortion in a *full reference* scenario. Hence, for this task, we do not consider the attributes themselves but *differences* of attributes. To compute these differences, we first establish a correspondence between the distorted mesh $\mathcal{D}$ and the reference one $\mathcal{R}$. For that purpose, we simply perform a fast projection of the vertices from $\mathcal{D}$ onto the surface of $\mathcal{R}$. For each vertex v of the distorted mesh $\mathcal{D}$ (associated with given attribute a_v), we compute its nearest 3D point $\hat{v}$ on the surface of the reference model $\mathcal{R}$ using an efficient AABB tree structure. For each 3D point $\hat{v}$, its attribute $a_{\hat{v}}$ is interpolated from the triangle it belongs to using barycentric coordinates. The features that we consider are thus the local per-vertex differences of each attributes and are computed as follows:

$$f_v = |a_v - a_{\hat{v}}| \tag{1}$$

where $|\cdot|$ is the L2 norm in case of vector attributes. Additionally to this *simple difference*, we also consider a *normalized difference*, defined as follows:

$$f_v^N = \frac{|a_v - a_{\hat{v}}|}{max(a_v, a_{\hat{v}}) + K}. \tag{2}$$

where K is a constant to avoid instability when denominators are close to zero. The denominator has two goals, (1) it normalizes the features f_v (between 0 and 1), which may be necessary for further machine learning algorithms and (2) it acts as a visual masking filter (e.g. it gives more weight to a same curvature variation in a flat region than in rough or highly curved ones). Of course, for vector attributes (geometric position, normal and geometric Laplacian) this denominator is discarded. The geometric position difference is normalized by the maximum length R of the bounding box.
To strengthen the robustness of the attribute differences, we consider each attribute at several scales; indeed on top of each *per vertex* attribute values a_v, we consider also *Gaussian-weighted averages* a_v^h computed as follow:

$$a_v^h = \sum_{v_i \in N(v,h)} w_v^h(v_i)\, a_{v_i} \tag{3}$$

$\mathcal{N}(v, h)$ is the connected set of vertices belonging to the sphere with center v and radius h (including the intersections of edges with the sphere) and $w_v^h()$ is a Gaussian weighting function centered on v with standard deviation of $h/2$ (i.e. the radius size h of the neighborhood is twice the standard deviation of the Gaussian filter). For each attribute, we consider two scales $h = \{0.012R, 0.036R\}$ on top of the *per vertex* version (with R the maximum length of the bounding box). These scales correspond to those used in [Lee et al. 2005].

Apart from these attribute differences, we also consider three other metrics which produce distortion maps:

- The **Hausdorff distance** - Each vertex of the distorted mesh is associated with its Hausdorff distance to the reference surface. It actually corresponds to the *geometric position* attribute without averaging and normalization. This measure is commonly used to drive or evaluate geometry processing operations.
- **MSDM2** [Lavoué 2011] - This metric computes differences of curvature statistics between the vertices of the distorted mesh and their projections on the reference one, at multiple scales. These local differences are then pooled into a single global score. Hence, this metric natively produces a distortion map. However, while it has shown to provide excellent results for global quality prediction, this metric has never been evaluated for the task of localized distortion visibility prediction.
- **FMPD** [Wang et al. 2012] - Like MSDM2, this perceptual metric has proven to be an excellent predictor of the global quality of 3D meshes but has never been evaluated for the task of visible difference prediction. The original FMPD computes a single difference of global roughness value between the meshes to compare. Here, we consider local differences of their local roughness estimator, based on the Laplacian of Gaussian curvature modulated by filters modeling the masking effect. This local roughness difference is also subject to normalization and averaging (see eq. 2 and 3).

The whole set of tested features thus includes 12 attributes $\times$ 2 kinds of differences (simple and normalized) $\times$ 3 averaging filters

(no averaging, small scale and large scale) + 3 metrics (including FMPD also subject to normalization and averaging), leading to 77 features.

5 Evaluation of geometric attributes and metrics

5.1 Performance measure

As in [Cadík et al. 2012], we measure the performance of the attributes and metrics using the receiver-operator-characteristic (ROC) that represents the relation between probability of false positives (vertices classified as *distorted* while they are not) and probability of true positives (vertices correctly classified as *distorted*) by varying a decision threshold on the metric output (resp. attribute difference). Binary ground-truth data are obtained by considering as *distorted* vertices marked by more than 25% of the observers (this threshold was also used in [Cadík et al. 2012]). The area under the ROC curve (AUC) can then be used as a direct indicator of the performance (1 corresponds to a perfect classification while 0.5 corresponds to a random one).
For each attribute/metric we compute the ROC performance separately for each object; the AUC values are then averaged over the 11 models. To test how the features can adapt to several objects and distortion types together, we also compute the ROC curve for all the objects together. For this latter experiment, we duplicate vertices of smallest models in order to balance the respective importance of each object in this global ROC calculation.

5.2 Results and comparisons

5.2.1 Performance of the attributes and metrics

Figures 1 and 4 illustrate some visual results, while table 3 details the quantitative performance in terms of AUC values for each feature and for each model from the dataset. Average values are presented in figure 5 (according to agreement statistics we excluded models #2 and #11 from this averaging). Additional results, including #2 and #11, are available in the supplementary materials.
For each attribute, we present the results obtained by the best filter: (i, S) and (i, N) respectively refer to simple (eq. 1) and normalized (eq. 2) attribute difference, while i refers to the averaging (eq. 3): 0 means per-vertex attribute and 1, 2 respectively refer to small and large averaging.
Table 3 shows that curvature-based features (including MSDM2 and FMPD) significantly outperform other ones (position, normal, dihedral angle, saliency, geometric Laplacian), whatever the type of distortion. In particular, curvedness produces among the best results on every models, however the best performing filters may be different depending on the model (e.g. (2,N) for #3 and (1,S) for #8). As expected, recent metrics MSDM2 and FMPD provide also excellent results, and the Hausdorff distance provides the worst ones. These results are confirmed in figure 5-*top row* which presents the average performances and standard deviations for all the features (excluding objects #2 and #11). Note that for this analysis, the filters are fixed for all the dataset (the best performing one is selected for each attribute). MSDM2 provides the best performance, as well as the smallest standard deviation which reflects an excellent stability whatever the model and/or the distortion. This stability is due to the fact that this metric is multi-scale and thus able to integrate artifacts at different levels of details. Curvedness and maximum curvature also provide very good results but a higher standard deviation.
When computing the classification performance on the whole dataset (all objects together, excluding #2 and #11), we obtain the AUC values detailed in figure 5-*bottom row*. For this particular scenario the FMPD metric as well as its raw attribute (Laplacian of Gaussian curvature) are ranked first, then MSDM2 and curvedness. This means that in scenarii where one single threshold has to be determined for detecting artifacts for several objects and/or distortion types, then the FMPD metric is the most suited. This may be due to the fact that this metric explicitly incorporates global and local normalization pre-processes (linked to the masking effect).

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
MinCurv	0.88 (1,N)	0.92 (2,S)	0.71 (2,N)	0.83 (1,S)	0.87 (1,S)	0.84 (0,N)	0.79 (0,S)	0.77 (0,N)	0.58 (1,S)	0.60 (0,S)	0.54 (1,S)
MaxCurv	**0.95** (2,S)	**0.97** (2,S)	**0.86** (2,N)	**0.90** (2,S)	**0.91** (2,S)	**0.89** (2,N)	**0.86** (2,S)	0.74 (0,S)	0.77 (2,N)	**0.73** (2,N)	0.57 (2,N)
MeanCurv	0.93 (1,N)	**0.96** (2,S)	**0.85** (2,N)	**0.89** (2,S)	**0.90** (2,S)	0.87 (2,N)	**0.86** (2,S)	0.78 (1,S)	0.77 (2,N)	0.70 (2,N)	0.56 (1,N)
GaussCurv	0.86 (1,S)	0.91 (2,S)	0.68 (1,N)	0.84 (2,S)	**0.90** (2,S)	0.87 (1,N)	0.81 (2,S)	0.79 (1,S)	0.54 (1,S)	0.62 (2,N)	0.58 (0,S)
ShapeInd	0.87 (1,N)	0.88 (2,S)	0.74 (2,S)	0.79 (0,S)	0.86 (1,S)	0.82 (1,S)	0.78 (1,S)	0.75 (1,S)	0.71 (1,S)	0.63 (1,S)	0.53 (1,S)
Curved.	**0.95** (2,S)	**0.97** (2,S)	**0.87** (2,N)	**0.89** (2,S)	**0.91** (2,S)	**0.90** (2,S)	**0.85** (2,S)	**0.80** (1,S)	**0.78** (2,N)	**0.73** (2,N)	0.58 (2,N)
Position	0.90 (2,N)	0.83 (2,S)	0.68 (2,N)	0.83 (1,N)	0.88 (2,N)	0.86 (2,N)	0.81 (1,N)	0.79 (1,N)	0.59 (1,N)	0.50 (0,N)	0.56 (0,N)
Dihedral	0.89 (1,N)	0.59 (2,N)	0.68 (0,N)	0.82 (1,N)	0.86 (1,N)	0.84 (0,N)	0.79 (1,N)	0.76 (0,N)	0.58 (2,N)	0.65 (2,N)	0.57 (0,N)
Normal	0.91 (2,N)	0.93 (2,N)	0.61 (0,N)	0.87 (2,N)	0.88 (1,N)	0.85 (1,N)	0.84 (1,N)	0.77 (1,N)	0.61 (1,N)	0.58 (1,N)	0.60 (2,N)
Saliency	0.85 (2,N)	0.91 (2,N)	0.64 (1,N)	0.79 (1,S)	0.83 (1,S)	0.79 (1,S)	0.76 (1,S)	0.67 (2,N)	0.56 (1,S)	0.63 (1,N)	0.57 (2,S)
Laplacian	0.90 (2,S)	0.65 (0,N)	0.60 (0,N)	0.82 (1,N)	0.87 (1,N)	0.85 (0,N)	0.78 (1,N)	0.77 (0,N)	0.54 (1,N)	0.69 (2,N)	0.63 (2,N)
LapGauss	0.88 (2,S)	0.89 (1,S)	0.71 (2,S)	0.86 (2,S)	**0.92** (2,S)	**0.91** (2,N)	0.83 (1,N)	0.78 (1,N)	0.72 (2,N)	0.56 (0,S)	**0.66** (0,S)
FMPD	**0.96** (2,S)	**0.95** (2,S)	**0.87** (2,S)	0.83 (2,S)	0.83 (1,S)	0.82 (0,S)	0.79 (0,S)	0.74 (0,S)	**0.82** (2,N)	0.63 (0,N)	**0.66** (0,N)
MSDM2	**0.97**	**0.97**	**0.85**	0.84	**0.90**	**0.88**	0.81	0.77	**0.79**	**0.71**	0.54
Hausdorff	0.76	0.83	0.60	0.79	0.83	0.80	0.76	0.73	0.50	0.50	0.56

Table 3: *AUC values for all the tested features (higher is better). For each attribute, we present the best filter: (i, S) and (i, N) respectively refer to simple (eq. 1) and normalized (eq. 2) attribute difference, while i refers to the averaging (0 means per-vertex attribute and 1, 2 respectively refer to small and large averaging). Best features are highlighted. Details about each model are given in table 1.*

5.2.2 Influence of the filters

Table 4 illustrates the AUC values averaged over all features and objects, for each filter separately. We run paired t-test with confidence level 0.95 to evaluate the effect of each filter. The normalization does not have a significant impact (p-value = 0.11). However Scale 1 is significantly better than Scale 0 (p-value = 6×10^{-8}) and Scale 2 is significantly better than Scale 1 (p-value = 0.02). The conclusion to draw here is that averaging the features has a significant impact on their ability to predict the visible distortion. The main reason is that such averaging (on geodesic neighborhoods) makes the computation reasonably independent of the sampling density.

	Scale 0	Scale 1	Scale 2
Attr. dif. S	0.70	0.73	0.75
Attr. dif. N	0.71	0.74	0.74

Table 4: *AUC values averaged over all features, for each filter separately.*

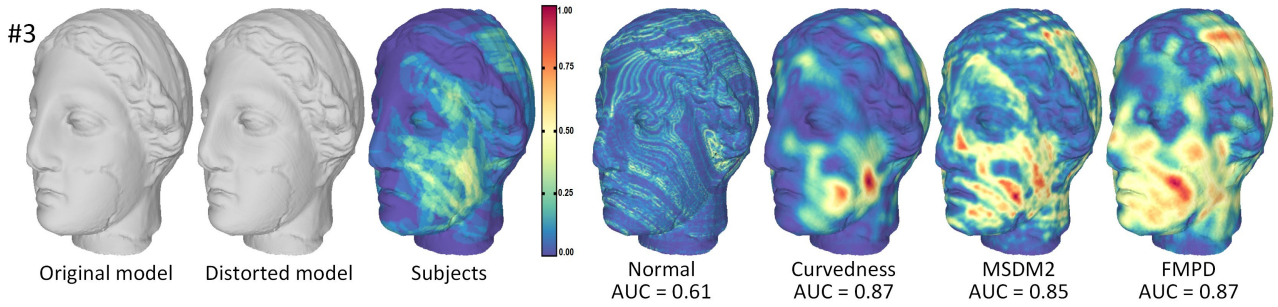

Figure 4: *Reference and distorted models from our dataset (object #3), observer data and distortion maps from different attributes and metrics (colors are mapped in the min-max range).*

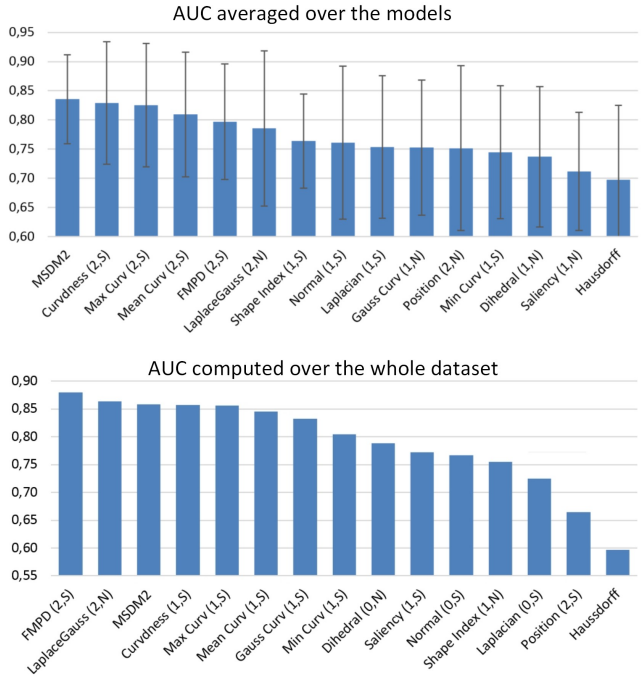

Figure 5: *Top: Mean AUC values for all the tested features. Error bars denote standard deviations. Bottom: AUC values when computing ROC curves over the whole dataset.*

5.3 Discussion

Several conclusions can be drawn from the results presented above:

- Previous studies [Lavoué and Corsini 2010; Corsini et al. 2013] showed that geometric distances (e.g. Hausdorff and root mean square error) are not good predictors of the perceived *global* visual quality. Our study completes this analysis by demonstrating that the geometric distance (Hausdorff as well as averaged position difference) is not a good predictor of the *local* visible difference.
- Recent metrics like MSDM2 [Lavoué 2011] and FMPD [Wang et al. 2012] provide the best results for this task of visible difference prediction; these good performances are mainly due to two key features: (1) being multi-scale does improve the robustness, and (2) smart masking filters improve the stability across different models and distortions.
- Simple curvature measures (e.g. the curvedness difference) may produce excellent results. However, we noted unstable performance with a high and unpredictable effect of the normalization. For instance, AUC values for model #9 are respectively 0.78 and 0.67 for curvedness (2,N) and curvedness (2,S) (see supplementary materials).
- Predicting the perceived local distortions on the surface of a 3D mesh appears to be a very challenging task. Figures 1 and 4 illustrate the fact that even best metrics are still far from perfectly predicting the ground-truth distortion maps. Hence, much works remain to be done in this area and our dataset is an excellent basis for that.

The provided insights into the visual perception of local artifacts are of broad interest for the computer graphics community. The attributes and metrics detected as good predictors by our study (i.e. MSDM2, FMPD and curvedness) may be used to design quality metrics as well as Just Noticeable Difference (JND) models. These features may also be useful to guide geometry processing algorithms.

6 Conclusion and future work

In this work, we have designed a novel subjective experiment to obtain information about the *local visibility* of complex artifacts on a set of 3D objects. We use the obtained distortion probability maps to (1) illustrate some properties of the human visual system and to (2) quantitatively evaluate a large set of 3D mesh attributes as well as recent perceptual metrics for the task of predicting perceived local distortions. From the results, we observe that curvature-based attributes (in particular curvedness) demonstrate a much better performance than others (e.g. dihedral angles, normals, Laplacian, saliency). As expected, Hausdorff geometric distance is a poor predictor of the perceived local distortion. Another interesting result is that recent perceptual metrics, originally designed for global quality evaluation, also provide the best and most stable results for this local task. We also introduce some recommendations highlighting the influence of attribute averaging and normalization for the design of new metrics.

These publicly available dataset and results[1] constitute the very first steps toward the efficient prediction of the perceivability of geometric artifacts, and more broadly toward the understanding of local perceptual mechanisms involving 3D geometry. In particular, this dataset will be very useful to evaluate and/or to train future metrics. Machine-learning is appearing as a powerful tool to design full reference image quality metrics [Cadik et al. 2013; Narwaria and Lin 2012; Charrier et al. 2012] and could indeed provide good results for 3D mesh artifact localization. However, we believe that such top-down approach should also integrate low-level human vision processes (such as a model of contrast sensitivity function) to be really efficient.

Our study focused on the geometry and thus considered a simple rendering style. We now plan to investigate how complex textures

[1] http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html

and shaders, used in modern rendering pipelines, influence the visibility of artifacts. Recent studies about the influence of material [Fleming 2014] and lighting [Faisman and Langer 2013] on the shape perception, should bring highly relevant cues for this goal.

References

CADÍK, M., HERZOG, R., MANTIUK, R. K., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. New Measurements Reveal Weaknesses of Image Quality Metrics in Evaluating Graphics Artifacts. *ACM Transactions on Graphics 31*, 6, 147.

CADIK, M., HERZOG, R., MANTIUK, R., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2013. Learning to Predict Localized Distortions in Rendered Images. *Computer Graphics Forum 32*, 7, 401–410.

CHARRIER, C., LÉZORAY, O., AND LEBRUN, G. 2012. Machine learning to design full-reference image quality assessment algorithm. *Signal Processing: Image Communication 27*, 3 (Mar.), 209–219.

CHEN, X., SAPAROV, A., PANG, B., AND FUNKHOUSER, T. 2012. Schelling points on 3D surface meshes. *ACM Transactions on Graphics 31*, 4 (July), 1–12.

CHO, J., PROST, R., AND JUNG, H. 2007. An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms. *IEEE Transactions on Signal Processing 55*, 1 (Jan.), 142–155.

CIGNONI, P., ROCCHINI, C., AND SCOPIGNO, R. 1998. Metro: Measuring Error on Simplified Surfaces. *Computer Graphics Forum 17*, 2, 167–174.

CIGNONI, P., CALLIERI, M., AND CORSINI, M. 2008. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*.

COHEN-STEINER, D., AND MORVAN, J. 2003. Restricted delaunay triangulations and normal cycle. In *19th Annu. ACM Sympos. Comput. Geom.*

COHEN-STEINER, D., ALLIEZ, P., AND DESBRUN, M. 2004. Variational shape approximation. In *ACM Siggraph*, 905–914.

CORSINI, M., GELASCA, E. D., EBRAHIMI, T., AND BARNI, M. 2007. Watermarked 3-D Mesh Quality Assessment. *IEEE Transactions on Multimedia 9*, 2 (Feb.), 247–256.

CORSINI, M., LARABI, M. C., LAVOUÉ, G., PETRIK, O., VÁŠA, L., AND WANG, K. 2013. Perceptual Metrics for Static and Dynamic Triangle Meshes. *Computer Graphics Forum 32*, 1 (Feb.), 101–125.

DALY, S. 1993. The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital images and human vision*, A. Watson, Ed. MIT Press, Cambridge, Oct., 179–206.

FAISMAN, A., AND LANGER, M. S. 2013. How does lighting direction affect shape perception of glossy and matte surfaces? *ACM Symposium on Applied Perception*.

FLEMING, R. W. 2014. Visual perception of materials and their properties. *Vision research 94* (Jan.), 62–75.

GARLAND, M., AND HECKBERT, P.-S. 1997. Surface simplification using quadric error metrics. *ACM Siggraph*, 209–216.

HERZOG, R., CADIK, M., AYDIN, T. O., KIM, K. I., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. NoRM : No-Reference Image Quality Metric for Realistic Image Synthesis. *Computer Graphics Forum 31*, 2.

KARNI, Z., AND GOTSMAN, C. 2000. Spectral compression of mesh geometry. In *ACM Siggraph*, 279–286.

KENDALL, M. G., AND BABINGTON SMITH, B. 1940. On the method of paired comparisons. *Biometrika 31*, 3/4, 324–345.

KOENDERINK, J. J., AND VAN DOORN, A. J. 1992. Surface shape and curvature scales. *Image and Vision Computing 10*, 8 (Oct.), 557–564.

KRIPPENDORFF, K. 2004. *Content analysis: An introduction to its methodology*. Thousand Oaks.

LAVOUÉ, G., AND CORSINI, M. 2010. A comparison of perceptually-based metrics for objective evaluation of geometry processing. *IEEE Transactions on Multimedia 12*, 7, 636–649.

LAVOUÉ, G. 2011. A Multiscale Metric for 3D Mesh Visual Quality Assessment. *Computer Graphics Forum 30*, 5, 1427–1437.

LEE, C., VARSHNEY, A., AND JACOBS, D. 2005. Mesh saliency. In *ACM Siggraph*, 659–666.

MANTIUK, R., KIM, K. J., REMPEL, A. G., AND HEIDRICH, W. 2011. HDR-VDP-2. *ACM Transactions on Graphics 30*, 4 (July), 1.

NARWARIA, M., AND LIN, W. 2012. SVD-Based Quality Metric for Image and Video Using Machine Learning. *IEEE Transactions on Systems, Man, and Cybernetics 42*, 2, 347–364.

RECOMMENDATION ITU-R BT. 500-11, . 2002. Methodology for subjective assessment of the quality of television pictures.

ROGOWITZ, B. E., AND RUSHMEIER, H. 2001. Are image quality metrics adequate to evaluate the quality of geometric objects? *Proceedings of SPIE*, 340–348.

SILVA, S., MADEIRA, J., AND SOUSA SANTOS, B. 2009. PolyMeCo-An integrated environment for polygonal mesh analysis and comparison. *Computers and Graphics 33*, 2, 181–191.

TAUBIN, G. 1995. A Signal Processing Approach to Fair Surface Design. In *ACM Siggraph*, 351–358.

VÁŠA, L., AND RUS, J. 2012. Dihedral Angle Mesh Error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. *Computer Graphics Forum 31*, 5.

WANG, Z., AND BOVIK, A. 2006. *Modern image quality assessment*. No. 1. Morgan & Claypool Publishers.

WANG, Z., BOVIK, A., SHEIKH, H., AND SIMONCELLI, E. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4, 600–612.

WANG, K., LUO, M., BORS, A., AND DENIS, F. 2009. Blind and robust mesh watermarking using manifold harmonics. In *IEEE International Conference on Image Processing*, no. 3, 211–226.

WANG, K., LAVOUÉ, G., DENIS, F., AND BASKURT, A. 2011. Robust and blind mesh watermarking based on volume moments. *Computers & Graphics 35*, 1 (Feb.), 1–19.

WANG, K., TORKHANI, F., AND MONTANVERT, A. 2012. A Fast Roughness-Based Approach to the Assessment of 3D Mesh Visual Quality. *Computers & Graphics*.

WATSON, B., FRIEDMAN, A., AND MCGAFFEY, A. 2001. Measuring and predicting visual fidelity. *ACM Siggraph*, 213–220.