A Multiscale Metric for 3D Mesh Visual Quality Assessment

Guillaume Lavoué

Université de Lyon, CNRS
Insa-Lyon, LIRIS UMR 5205, France

Abstract
Many processing operations are nowadays applied on 3D meshes like compression, watermarking, remeshing and so forth; these processes are mostly driven and/or evaluated using simple distortion measures like the Hausdorff distance and the root mean square error, however these measures do not correlate with the human visual perception while the visual quality of the processed meshes is a crucial issue. In that context we introduce a full-reference 3D mesh quality metric; this metric can compare two meshes with arbitrary connectivity or sampling density and produces a score that predicts the distortion visibility between them; a visual distortion map is also created. Our metric outperforms its counterparts from the state of the art, in term of correlation with mean opinion scores coming from subjective experiments on three existing databases. Additionally, we present an application of this new metric to the improvement of rate-distortion evaluation of recent progressive compression algorithms.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computer Graphics—Computational Geometry and Object Modeling I.3.m [Computer Graphics]: Computer Graphics—Perception

1. Introduction

Significant advances in the fields of computer graphics, telecommunications and hardware design over the last decade have boosted the use of 3D digital data. This 3D content, mostly represented by polygonal meshes, is subject to a wide variety of distortions during common operations such as compression, filtering, simplification, watermarking and so forth. These distortions may alter the visual quality of the 3D content which is critical since these processing operations are generally targeted at human centered applications. A main problem is that most of existing processing algorithms (e.g. simplification, watermarking, compression) are driven and/or evaluated by simple metrics like Hausdorff distance and root mean square error (RMS), which are not correlated with the human vision. For instance all distorted models presented in figure 1 are all associated with the same RMS with respect to the original model however their visual qualities vary from very good (top row) to very poor (bottom row). Hence, some *objective* quality metrics have been introduced, their goal is to produce a score that predicts the *subjective* visual quality (or the visual impact of the distortion) of a distorted 3D model with respect to a reference (distortion-free) model; these objective scores should be statistically consistent with those of human observers. Such metrics can play critical roles in computer graphics

by replacing standard geometric distances for (1) evaluating/benchmarking and (2) driving 3D mesh processing systems and algorithms. However objective quality assessment research for 3D objects is still in its very early stages; only a few metrics have been proposed and they have some heavy constraints (e.g. objects to compare have to share the same connectivity or the same sampling density), moreover they still do not correctly correlate with the human judgment as raised by a recent evaluation [LC10].

In that context we propose a multiscale metric for objective quality assessment of 3D mesh. This metric has no constraint on the meshes being compared, and demonstrates a very high correlation with the human judgment. Our approach first computes a fast asymmetric matching between the distorted object $\mathcal{M}_d$ and the original one $\mathcal{M}_r$, then for each vertex, Gaussian-weighted curvature statistics are computed at multiple scales over local windows to produce a local distortion map; local values are then pooled into a single Global Multiscale Distortion score (GMD). Our final metric is obtained by averaging forward ($\mathcal{M}_d \rightarrow \mathcal{M}_r$) and backward ($\mathcal{M}_r \rightarrow \mathcal{M}_d$) global distortion scores; we will refer to this new metric as MSDM2 as it shares the origins and the principles of *structural degradation* with the original MSDM (*Mesh Structural Distortion Measure*) from Lavoué et al. [LDD*06]. The source code of both

metrics is available within the MEsh Processing Platform (http://gforge.liris.cnrs.fr/projects/mepp).

The following section details the existing works on perception for computer graphics, including existing quality assessment metrics. Section 3 provides an overview of our method, while section 4 details the processing pipeline. Finally experiments and comparisons on several subjective databases are provided in section 5 while an application is presented in the last section.

2. Related work

2.1. 2D Image Quality Assessment

In the field of 2D image processing, the research on objective image quality assessment metrics is highly developed [WB06]. Existing algorithms can be classified according to the availability of a reference image: full reference (FR), no-reference (NR) and reduced-reference (RR). The following discussion only focuses on FR methods, where the original *distortion free* image is known as the reference image. Since the pioneer visual difference predictor (VDP) of Daly [Dal93], a lot of metrics have been introduced which aim at replacing the classical peak-signal-to-noise ratio (PSNR) which does not correlate with the human vision. Many techniques have tried to mimic the low-level mechanisms of the human visual system (e.g. the *contrast sensitivity* and the *visual masking* which defines the fact that one visual pattern can hide the visibility of another) like the visual signal-to-noise ratio (VSNR) [CH07]. Recently, a different class of techniques has appeared, which does not rely directly on these low level psychophysical properties, but instead proposes some signal fidelity criteria expected to correlate well with perceptual quality. Among them the structural similarity index (SSIM) from Wang et al. [WBSS04] focuses on the structural information of the image, considering that the structure of a good-quality image has to closely match that of the original. Their metric computes various spatial correlations between local windows from the original and distorted images, resulting in a local quality value for each pixel; a pooling algorithm then combines these local values into a single overall quality score. This approach has proven to offer very good performance in predicting image fidelity and was later improved and extended by many authors, e.g. a multiscale extension was proposed in [WSB03] and a better pooling strategy was introduced in [ZMZ11].

2.2. Perceptual issues in computer graphics and Visual Mesh Quality Assessment

In the field of computer graphics, the perception and human vision mechanisms have been studied for several applications [MMBC10], particularly *rendering* and *simplification*, however these works are mostly based on existing 2D image metrics. For instance Lindstrom and Turk [LT00] and more recently Qu and Meyer [QM08] drive their simplification algorithms using 2D perceptual models. For rendering,

2D perceptual metrics are used to determine, according to the location of the observer, the amount of accuracy to use, e.g. the best level of details [Red01] or the best ray sampling density [BM98]. Some interesting related works have been recently proposed: Zhu et al. [ZZDZ10] study the relationship between the viewing distance and the perceptibility of model details using 2D metrics (VDP and SSIM), Ramanarayanan et al. [RFWB07] propose a model that evaluates how object geometry, material and illumination influence the appearance or the rendered image and finally Aydin et al. [AvMS10] and Váša and Skala [VS11] introduce quality assessment metrics respectively designed for computer graphics video and dynamic meshes.

All the previously mentioned works are based on 2D image metrics. However, as observed in the subjective experiments of Rogowitz and Rushmeier [RR01], the perceived quality of a 3D model may not be correctly predicted by the quality of its 2D projections. Hence some authors have recently introduced perceptually-motivated algorithms operating directly on the mesh geometry. Lee et al. [LVJ05] introduce the notion of *saliency* for 3D meshes, it is computed by applying a difference of Gaussian operator at multiple scales on the curvature map; the Gaussian filters are applied on local spherical windows around each vertex. Similarly, Lavoué [Lav09] proposes a roughness estimator also based on curvature statistics over local windows. Our approach is quite related to these last works since it is based on differences of Gaussian-weighted curvature statistics, computed for all vertices over local spherical windows.

At present only a few works exist for 3D object visual quality assessment, they all follow the full-reference principle. Karni and Gotsman [KG00], in order to evaluate the quality of a compressed 3D mesh with respect to the original one, propose a metric which combines the RMS geometric distance between corresponding vertices with the RMS distance of their Laplacian coordinates. Also in the context of the evaluation of compressed objects, Sorkine et al. [SCOT03] use this metric however they increase the weight associated to the Laplacian coordinates (which represent a kind of smoothness measure) guessing that the visual perception is more sensitive to smoothness/roughness variations than to pure geometric displacements. Drelie Gelasca et al. [GECB05] and Corsini et al. [CGEB07] propose perceptual metrics based on global roughness variation to measure the quality of watermarked meshes. They provide two methods for the computation of roughness: the variance of the difference between a 3D model and its smoothed version, and the variance of the dihedral angles between adjacent faces evaluated in a multi-resolution fashion. Lavoué et al. [LDD*06] propose a metric called Mesh Structural Distortion Measure (MSDM), it basically follows the framework of the SSIM index proposed for image quality assessment: differences of curvature statistics are computed over local windows from both meshes being compared. Lastly, Bian et al. [BHM09] introduce another perceptual metric based on the computation of the strain energy (i.e. the en-
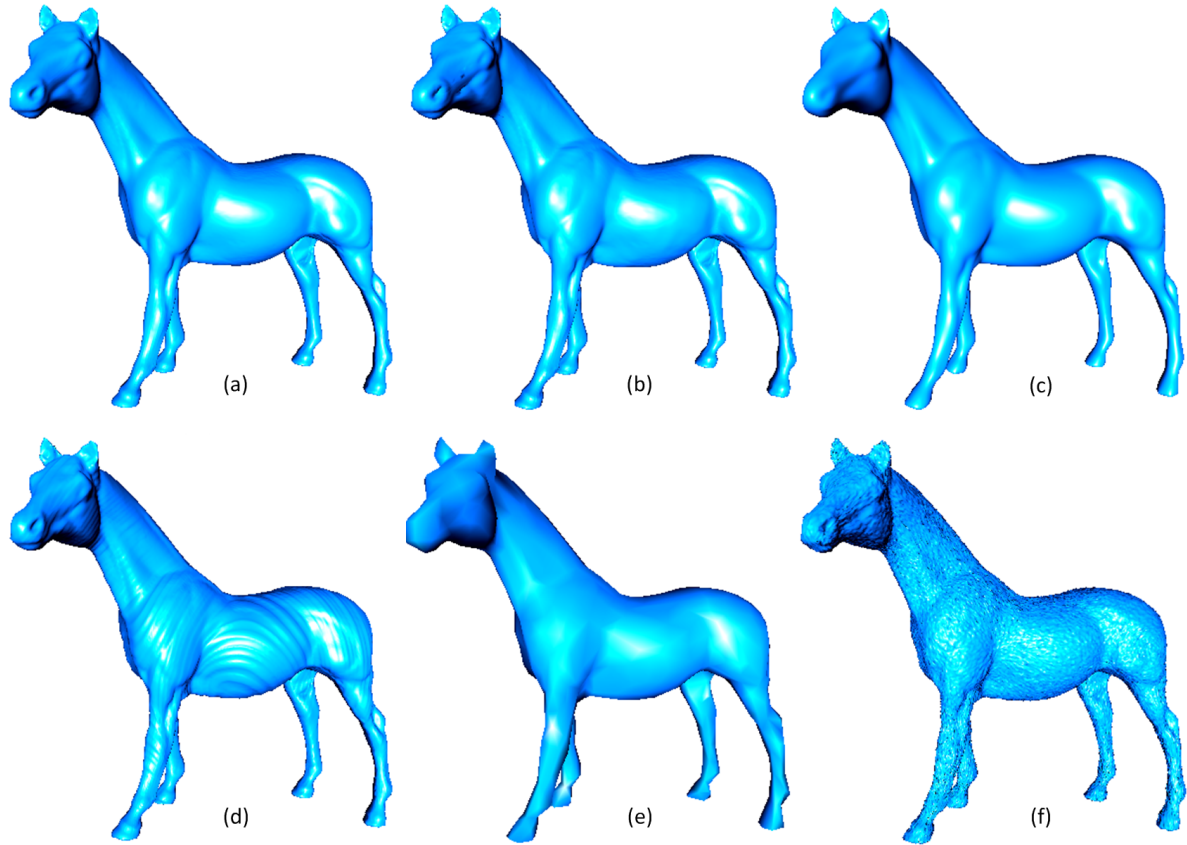
Figure 1: *Distorted versions of the Horse model, all associated with the same maximum root mean square error (MRMS = 1.05 × 10−3). (a) Original model. Results after (b) watermarking from Wang et al. [WLDB11] (MSDM2=0.14), (c) Laplacian smoothing [Tau00] (MSDM2=0.40), (d) watermarking from Cho et al. [CPJ06] (MSDM2=0.51), (e) simplification [LT99] from 113K vertices to 800 vertices (MSDM2=0.62), (f) Gaussian noise addition (MSDM2=0.84).*

ergy which causes the deformation) of each triangle. A very recent study done by Lavoué and Corsini [LC10] has provided a quantitative evaluation of these visual quality metrics, by computing their statistical consistency with mean opinion scores from human observers coming from two subjective experiments. Their conclusions is that whereas some metrics [CGEB07, LDD*06] provide correct results, there is still large room for improvement in performance; moreover most of these metrics are not able to compare meshes that do not share the same connectivity [LDD*06] or the same sampling density [CGEB07], that constitutes a heavy drawback since they cannot be used in a remeshing or simplification evaluation scenario for instance. The MSDM metric [LDD*06] can be seen as the ancestor of the proposed metric MSDM2, indeed they are both inspired by the SSIM index and share the principles of computation of local curvature statistics over spherical windows; however each step of the pipeline is quite different. Additionally our method is multiscale and has no constraint on the meshes being compared.

3. Overview of our approach

Our approach is largely inspired by the 2D image SSIM metric of Wang *et al.* [WBSS04], who consider that the human visual system is highly adapted for extracting structural information. Hence our measure relies on differences of structure (captured via curvature statistics) computed on local corresponding neighborhoods from the meshes being compared. Moreover we believe that a measure of visual distortion has to depend on a scale parameter, indeed as confirmed by Zhu et al. [ZZDZ10], the perceptability of a distortion on a 3D object depends on its level of details and its viewing conditions (e.g. display resolution and viewing distance). Accordingly a single-scale method may be appropriate only for specific settings, this constitutes a clear limitation of all existing 3D metrics. Therefore, similarly to what Lee et al. [LVJ05] did for their saliency model, we define a *multi-*

scale distortion measure aimed at capturing the distortions at all perceptually meaningful scales, hence increasing its efficiency and robustness. Additionally, in the 2D image field, the multiscale extension of SSIM proposed in [WSB03], has proven to produce better results than its single-scale counterpart.

In our work, each scale h_i defines the radius of the local neighborhoods and is also used for the curvature computation. Given a distorted mesh $\mathcal{M}_d$ and the corresponding reference (i.e. original) one $\mathcal{M}_r$, our visual distortion measure is computed as follows:

1. A scale-dependent curvature is computed on vertices from both meshes (see section 4.1).
2. Each vertex of the distorted mesh $\mathcal{M}_d$ is matched with its corresponding 3D point and curvature value from the reference mesh $\mathcal{M}_r$, using fast projection and barycentric interpolation (see section 4.2).
3. For each vertex of $\mathcal{M}_d$ a local distortion measure is computed as a difference of Gaussian-weighted statistics computed over a local spherical neighborhood of radius h_i (see figure 2 and section 4.3).
4. Steps (1-3) are repeated for multiple scales h_i, leading to several distortion maps.
5. The final distortion map is constructed by adding the local distortion maps at all scales (see section 4.4).
6. The global multiscale distortion score is then obtained by combining the local values using a Minkowski pooling (see figure 3 and section 4.4).

Our metric MSDM2 is finally computed as the average of forward ($\mathcal{M}_d \rightarrow \mathcal{M}_r$) and backward ($\mathcal{M}_r \rightarrow \mathcal{M}_d$) global distortion scores, and is therefore symmetric. For all the results in this paper, and as a compromise between performance and processing time, we have used three scales. $h_i \in \{2\varepsilon, 3\varepsilon, 4\varepsilon\}$, with $\varepsilon = 0.5\%$ of the max length of the bounding box of the model.

Here are the main differences between the proposed MSDM2 and its predecessor MSDM [LDD*06]:

- The curvature is scale-dependent (improving robustness).
- MSDM needs an implicit correspondence between vertices, while MSDM2 computes a fast projection and curvature interpolation. It implies no connectivity constraint and improves the matching quality.
- Curvature statistics have been slightly modified; moreover they are normalized by Gaussian weighting functions and their combination is different.
- The proposed approach is multiscale and symmetric.
- The correlation with the human judgement is greatly improved (see tables 1 and 2).
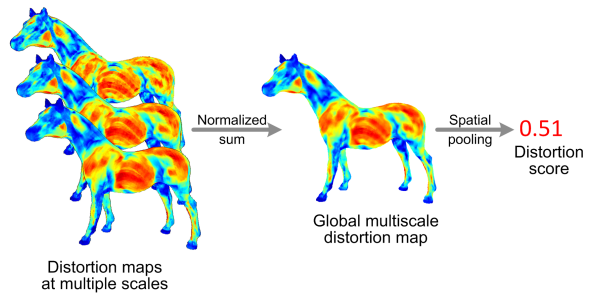
4. The visual distortion metric

4.1. Multiscale robust curvature

To capture the local structural information of the 3D objects our algorithm relies on curvature, more precisely the
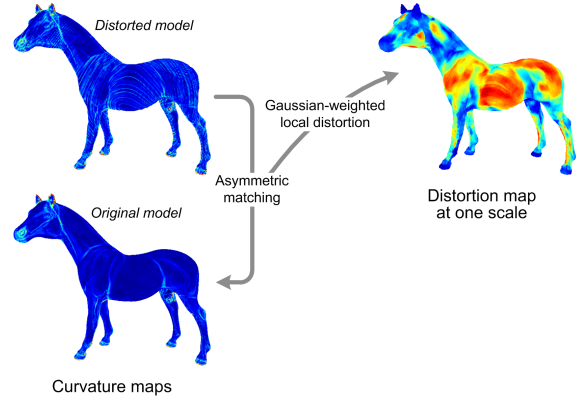

Figure 2: *Distortion map computation at one scale, for the distorted model (d) from figure 1.*

Figure 3: *Distortion maps at different scales, their combination into one multiscale map and the distortion score calculation.*

mean curvature, indeed this scalar field describes the visual characteristics of a 3D model well. Additionaly it was successfully used in several perceptually-motivated works [LVJ05, Lav09]. There exist plenty of methods to compute curvature (e.g the Normal Cycle algorithm [CSM03]), however many may produce unstable results in case of irregular tessellation and their range of values may depend on the sampling density. The main reason of these drawbacks is that the curvature tensors are usually computed on the one-ring neighborhoods of the vertices. This sensitivity to connectivity is critical in our application since our objective is to compare local curvature statistics (mean, variance, covariance) between two meshes that do not share the same connectivity nor the same level of details. To resolve this robustness issue, we adopt the solution proposed by Alliez et al. [ACSD*03] i.e. we evaluate the tensor on a larger neighborhood around each vertex: a geodesic disk approximated by the intersection of the surface with a sphere centered at the vertex. A very interesting side effect is that by varying the radius of the sphere we can compute the curvature at multiple scales. With a small radius r, tiny details are captured whereas a larger r leads to a kind of smoothing of the field, this behav-

ior seems directly linked to the scale h that drives our distortion measure calculation, hence we decided to directly link them by fixing for each scale h: $r = h/5$. Figure 4 illustrates two mean curvature maps for the Horse model, computed using two different radii, a small one (left) and a larger one (right). With this strategy, the range of values of the mean curvature field does not depend on the sampling of the model but rather on the scale used, this has been shown to result in significantly improve the performance; for instance the Pearson correlation between metric values and subjective scores drops from 79.6% to 36.6% on the Simplification database (see Table 3) when using a simple 1-ring curvature. Note that the curvature is also normalized by the size of the model to produce values invariant to this size.
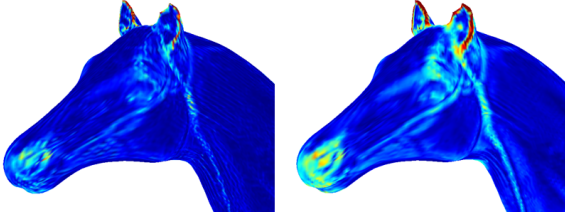
Figure 4: *Curvature map using different radius.* Left*: $r = 0.2\%$,* Right*: $r = 0.5\%$ of the the bounding box length.*

4.2. Fast matching

Once we have computed the curvature at a given radius $r = h/5$, the objective is now to establish a correspondence between the distorted mesh $\mathcal{M}_d$ and its corresponding original version $\mathcal{M}_r$. The general problem of shape correspondence is difficult and remain open for articulated meshes; however our task is much easier, indeed (1) we just look for an asymmetric correspondence $\mathcal{M}_d \rightarrow \mathcal{M}_r$ and (2) the meshes to match are aligned and they are very close to each other in term of shape. Hence, taking into account the fact that the computation time is also critical in our application, we simply perform a fast projection of the vertices from $\mathcal{M}_d$, onto the surface of $\mathcal{M}_r$.

For each vertex v of the distorted mesh $\mathcal{M}_d$, we compute its nearest 3D point $\hat{v}$ on the surface of the reference model $\mathcal{M}_r$ using the efficient AABB tree structure [ATW09] from the CGAL library:

$$\hat{v} = arg \min_{p_i \in \mathcal{M}_r} \|p_i - v\| \tag{1}$$

Then, for each 3D point $\hat{v}$, its curvature $\mathcal{C}(\hat{v})$ is interpolated using barycentric coordinates. More precisely, considering that $\hat{v}$ belongs to the triangle $\{v_1^r, v_2^r, v_3^r\}$ of $\mathcal{M}_r$, then its curvature is defined as follows:

$$\mathcal{C}(\hat{v}) = \sum_{j=1}^{3} b_j(\hat{v}).\mathcal{C}(v_j^r) \tag{2}$$

with $\left\{ b_j(\hat{v}) \right\}_{j=1..3}$ the barycentric coordinates of $\hat{v}$ in the triangle $\{v_1^r, v_2^r, v_3^r\}$, i.e. $b_j(\hat{v})$ is the ratio of the area of the triangle formed by $\hat{v}$ and the opposite edge to vertex v_j^r over the area of the whole triangle. Hence we now have for all vertices $\{v_i\}$ from $\mathcal{M}_d$, their corresponding 3D points $\{\hat{v}_i\}$ on the surface of $\mathcal{M}_r$ associated with curvature values.

Note that this matching using the AABB tree is obviously not done for each scale but only once at the beginning of the algorithm: for each vertex v_i of $\mathcal{M}_d$, we record its projection $\hat{v}_i$ and the associated triangle. Then for each scale h, once the curvature is computed for vertices of both meshes (see step 1 in section 3) we update the curvature values $\{\mathcal{C}(\hat{v}_i)\}$.

4.3. Local distortion measurement

As discussed above our metric is mostly inspired by the work from Wang *et al.* [WBSS04] who define a quality measure between two images, based on the visual degradation of the structural information. Their structural similarity index $SSIM(\mathbf{x}, \mathbf{y})$ between two image signals $\mathbf{x}$ and $\mathbf{y}$ is defined as follows: $SSIM(\mathbf{x}, \mathbf{y}) = L(\mathbf{x}, \mathbf{y}) \cdot C(\mathbf{x}, \mathbf{y}) \cdot S(\mathbf{x}, \mathbf{y})$, where L is a luminance comparison function (means of luminance are compared), C is a contrast comparison function (variances of luminance are compared) and S is a structure comparison function (which studies the covariance of the signals). These statistics are not computed globally on the entire images but locally on 11×11 circular neighborhoods around each pixel. In our case, we apply the same strategy: we consider a spherical neighborhood around each vertex v of $\mathcal{M}_d$ and we compute differences of curvature statistics between the set of vertices from the neighborhood and their corresponding 3D points on $\mathcal{M}_r$. For a given scale h, we define the neighborhood $\mathcal{N}(v, h)$ at each vertex v as the connected set of vertices belonging to the sphere with center v and radius h; we also compute and add to $\mathcal{N}(v, h)$, the intersections between this sphere and edges of the mesh, as in [Lav09] (their curvature values are interpolated).

Functions L, C and S have already been adapted for 3D mesh quality assessment in the MSDM metric [LDD*06]; we have slightly modified them by optimizing the weights and adding a Gaussian normalization; furthermore the matching mechanism between vertices of the meshes being compared is completely different. In our case, we have already computed the correspondence between the distorted mesh $\mathcal{M}_d$ and the reference one $\mathcal{M}_r$; for each vertex v of $\mathcal{M}_d$, we introduce the following functions, defined for each scale h:

$$L^h(v) = \frac{\left\|\mu_v^h - \mu_{\hat{v}}^h\right\|}{max(\mu_v^h, \mu_{\hat{v}}^h) + K} \tag{3}$$

$$C^h(v) = \frac{\left\|\sigma_v^h - \sigma_{\hat{v}}^h\right\|}{max(\sigma_v^h, \sigma_{\hat{v}}^h) + K} \tag{4}$$

$$S^h(v) = \frac{\left\|\sigma_v^h \sigma_{\hat{v}}^h - \sigma_{v\hat{v}}^h\right\|}{\sigma_v^h \sigma_{\hat{v}}^h + K} \tag{5}$$

where K is a constant to avoid instability when denominators are close to zero. μ_v^h, $\mu_{\hat{v}}^h$ are respectively Gaussian-weighted averages of curvature over the vertices of the h-scale neighborhood of v and over their projections onto $\mathcal{M}_r$. Similarly σ_v^h, $\sigma_{\hat{v}}^h$ and $\sigma_{v\hat{v}}^h$ are Gaussian-weighted standard deviations and covariance values:

$$\mu_v^h = \sum_{v_i \in N(v,h)} w_v^h(v_i) \mathcal{C}(v_i) \tag{6}$$

$$\mu_{\hat{v}}^h = \sum_{v_i \in N(v,h)} w_{\hat{v}}^h(\hat{v}_i) \mathcal{C}(\hat{v}_i) \tag{7}$$

$$\sigma_v^h = \sqrt{\sum_{v_i \in N(v,h)} w_v^h(v_i)(\mathcal{C}(v_i) - \mu_v^h)^2} \tag{8}$$

$$\sigma_{\hat{v}}^h = \sqrt{\sum_{v_i \in N(v,h)} w_{\hat{v}}^h(\hat{v}_i)(\mathcal{C}(\hat{v}_i) - \mu_{\hat{v}}^h)^2} \tag{9}$$

$$\sigma_{v\hat{v}}^h = \sum_{v_i \in N(v,h)} w_v^h(v_i)(\mathcal{C}(v_i) - \mu_v^h)(\mathcal{C}(\hat{v}_i) - \mu_{\hat{v}}^h) \tag{10}$$

$w_v^h()$ is a Gaussian weighting function centered on v with standard deviation of $h/2$ (i.e. the size h of the neighborhood is twice the standard deviation of the Gaussian filter, similar to what Lee et al. [LVJ05] did for their saliency estimator). It is defined as follows:

$$w_v^h(v_i) = \frac{e^{-2\|v_i - v\|^2/h^2}}{\sum_{v_j \in N(v,h)} e^{-2\|v_j - v\|^2/h^2}} \tag{11}$$

Our local distortion measure is then defined, at a given scale h, for each vertex v from $\mathcal{M}_d$ as follows:

$$LD^h(v) = \frac{\alpha\, L^h(v) + \beta\, C^h(v) + \gamma\, S^h(v)}{\alpha + \beta + \gamma} \tag{12}$$

α, β and γ were set respectively to 1, 1 and 0.5. We argue that γ must be smaller since $S^h(v)$ has a larger disparity than both other functions.

The multiscale local distortion measure MLD is finally obtained simply by averaging single-scale values. The normalization is very simple since the LD^h are all in $[0,1]$:

$$MLD(v) = \frac{\sum_{i=1}^n LD^{h_i}(v)}{n} \tag{13}$$

where n is the number of scales used (3 in our experiments). Figures 5 and 6 illustrate forward MLD maps for distorted versions of the Lion and Horse models, Hausdorff distance maps are also given for comparison (in this latter case, each vertex of the distorted mesh is associated with its Hausdorff distance to the reference surface). In figure 5 the Lion model is distorted using a random noise addition on every vertex, however when looking at the distorted mesh the distortion is much more visible when applying on smooth parts (i.e. the face) than on rough parts (i.e. the mane); this observation is related to the visual masking effect, indeed this perceptual concept tells us that a textured (i.e. *rough*) region is able

to hide geometric distortions much better than a smooth one; we can observe that this phenomenon is captured by our metric which provides higher distortion values when the noise is applied on the smooth parts. In figure 6, the
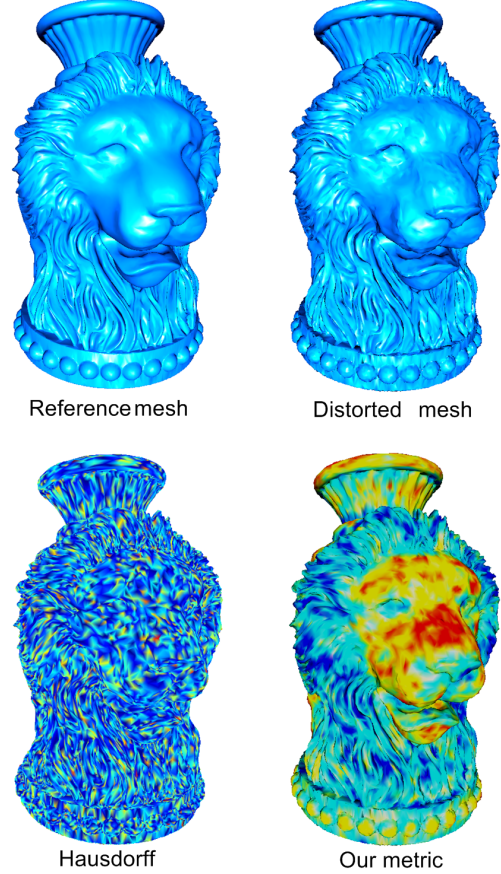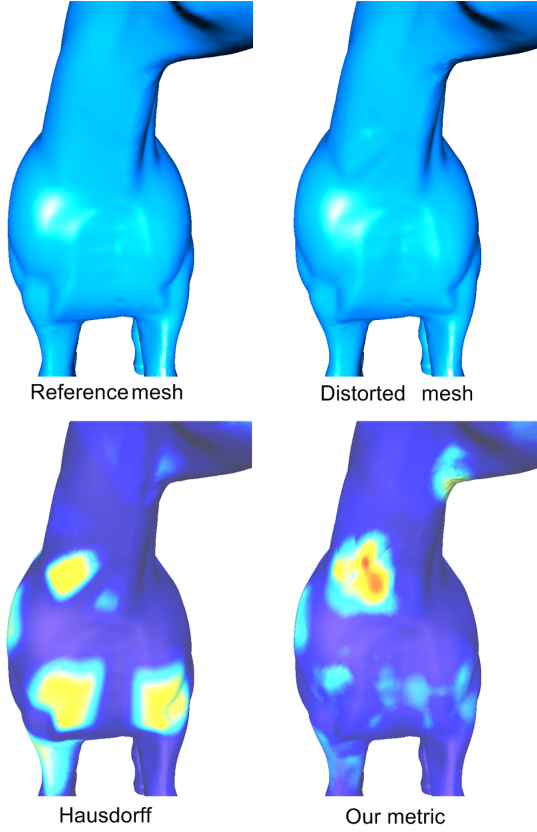

Figure 5: Top*: The Lion model and a distorted version after random noise addition,* Bottom*: distortion maps computed using the Hausdorff distance and our metric. Warmer colors represent higher values.*

Horse model has been watermarked using the algorithm from Wang et al. [WLDB11]; this algorithm creates some smooth bumps on the surface, as can be observed on the Hausdorff distance map. However, when looking at the model, all these bumps are not equally visible; once again the distortion visibility is very efficiently estimated by our metric, indeed higher values are on the most visible bump on the top left of the chest (others are quite invisible) and on the high frequency noise on the throat.

4.4. Global distortion score

Once we have created the multiscale distortion map, the objective is then to compute a single score assessing the global

Figure 6: Top*: The Horse model and a distorted version after watermarking (algorithm from Wang et al. [WLDB11]),* Bottom*: distortion maps computed using the Hausdorff distance and our metric. Warmer colors represent higher values.*

visual distortion. In the original 2D metric from Wang et al. [WBSS04], a simple sum over the pixels is calculated. In our case, we follow one of the most popular solutions used in many perceptual metrics: a *Minkowski* pooling defined as follows:

$$GMD_{\mathcal{M}_d \to \mathcal{M}_r} = \left(\frac{1}{|\mathcal{M}_d|} \sum_{v \in \mathcal{M}_d} MLD(v)^p \right)^{\frac{1}{p}} \quad (14)$$

As p increases, more emphasis is given to the high distortion regions, typical values are between 2 and 4. This is intuitively closer to the human judgment than a simple averaging, indeed humans tend to pay more attention to high distortion regions when they establish their visual judgment; in our metric, we consider $p = 3$.

$GMD_{\mathcal{M}_d \to \mathcal{M}_r}$ evaluates the structural distortion of the distorted model regarding the reference one. In order to strengthen the robustness of our method and to obtain a symmetric measure we also compute $GMD_{\mathcal{M}_r \to \mathcal{M}_d}$ and we retain the average as our final distortion measure MSDM2.

This metric is within the range $[0, 1]$, a value of 0 means that the two objects are identical while values near 1 mean that they are visually very different. Some examples of MSDM2 values are given in figure 1. Distortions of different types are applied on the Horse model (they are all associated with the same maximum root mean square distance). We can observe that our distortion metric provides a low value when the object seem nearly identical to the original (see fig. 1.b) and high values when distortions are very visible (see figure 1.d to 1.f). Hence this metric seems a good predictor of the human opinion, these results are confirmed quantitatively in the next section.

5. Experiments and comparisons

The classical method to measure the performance of a quality assessment metric is to evaluate its correlation with the human judgment; practically and whatever the type of media (image, video or 3D models), this evaluation is usually done as follows:

1. A database is constructed containing different objects (reference objects and distorted versions).
2. A subjective experiment is conducted where human observers directly give their opinion or some ratings about the perceived distortions of the database objects. A mean opinion score (MOS) is then computed for each distorted object of the corpus: $MOS_i = \frac{1}{n} \sum_{j=1}^{n} m_{ij}$ where MOS_i is the mean opinion score of the ith object, n is the number of test subjects, and m_{ij} is the score (in a given range) given by the jth subject to the ith object.
3. A normalization of the MOS values and a filtering of possible outlier subjects is performed according to recommendation of the I.T.U. (International Telecommunication Union) [I.T00]
4. The correlation is computed between the mean opinion scores of the objects and their associated metric's values; usually two correlation coefficients are considered: the Spearman Rank Order Correlation which measures the monotonic association between the MOS and the metric values and the Pearson Linear Correlation Coefficient, which measures the prediction accuracy. The Pearson correlation is computed after performing a non-linear regression on the metric values, usually using a logistic or a cumulative Gaussian function. This serves to optimize the matching between the values given by the objective metric and the subjective opinion scores provided by the subjects. This step allows the evaluation to take into account the saturation effects typical of human senses.

To the best of our knowledge, there are currently three publicly-available subject-rated 3D model databases, we consider all of them in our experiments and comparisons:

---

† http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html

The LIRIS/EPFL General-Purpose Database[†] [LDD*06] was created at the EPFL, Switzerland. It contains 88 models between 40K and 50K vertices generated from 4 reference objects (Armadillo, Dyno, Venus and RockerArm). Two types of distortion (noise addition and smoothing) are applied with different strengths and at four locations: uniformly (on the whole object), on smooth areas, on rough areas and on intermediate areas. 12 observers participated to the subjective evaluation.

The LIRIS Masking Database[†] [Lav09] was created at the Université of Lyon, France. It contains 26 models between 9K and 40K vertices generated from 4 reference objects (Armadillo, Bimba, Dyno and Lion) specifically chosen because they contain significantly smooth and rough areas. The only distortion is noise addition applied with three strengths however it is applied either on smooth or rough regions. The specific objective of this database was to evaluate the behavior of the metrics regarding the visual masking effect. Indeed the noise is far less visible on rough regions hence the metrics should follow this perceptual mechanism. 11 observers participated to the subjective evaluation.

The IEETA Simplification Database[‡] [SSFM09] was created at the University of Aveiro, Portugal. It contains 30 models generated from 5 reference objects (Bunny, Foot, Head, Lung and Strange) from 2K to 25K vertices. The reference models have been simplified using three different methods and two levels (20% and 50% of the original number of faces). 65 observers participated to the subjective evaluation.

For each of these databases, we compare the performance of our metric in term of correlation with the MOS, against the following existing algorithms:

- Simple geometric distances: Hausdorff and root mean square error. Note that for the first two databases, distorted models have the same connectivity than the reference ones, hence in that case we use the RMS between corresponding vertices. In contrast, for the simplification database, we consider the MRMS error which is the maximum of the two asymmetric RMS errors computed using METRO [CRS98].
- The combinations, GL_1 and GL_2, of the RMS with Laplacian coordinates introduced respectively by Karni and Gotsman [KG00] and Sorkine et al. [SCOT03].
- The roughness-based measures, $3DWPM_1$ and $3DWPM_2$, from Drelie Gelasca et al. [GECB05] and Corsini et al. [CGEB07].
- The Mesh Structural Distortion Measure, $MSDM$, from Lavoué et al. [LDD*06].

Note that for the simplification database, we consider only Hausdorff and MRMS since others cannot apply on meshes

with different connectivity and/or sampling densities. Tables 1, 2 and 3 present the Spearman and Pearson correlations of these metrics and our algorithm. Values of state-of-the-art metrics from tables 1, 2 have been taken from [LC10].

For the general-purpose and masking databases, we have several major observations:

Firstly, simple metrics (Hausdorff, RMS, GL) provide quite poor results; the main reason is that it is very difficult for these metrics to correctly merge the visual effects coming from different types of distortion (noise and smoothing in the case of the general-purpose corpus) or to take into account some difficult psychophysical mechanisms like the masking effect.

Secondly, when considering distorted versions from each reference object separately, most recent metrics like 3DWPM and MSDM provide correct results, however our method MSDM2 shows a significant improvement, particularly in term of Pearson correlation (which characterizes the strength of the relationship).

Thirdly, when we consider the correlation over the whole set of models of the databases, our method clearly outperforms the others: respectively 66% and 76% Pearson correlation against 56% and 48% for the second best (MSDM), for general-purpose and masking databases respectively. The fact is that even simple metrics are able to behave consistently with human judgment when applied to distorted models originating from the same reference one with a single type of distortion, however when applied to a set of models generated from different reference objects and including different types of distortions then the task becomes much more difficult, however our metric still provides very good results in that case.

For the simplification database (see table 3), when considering the correlation for each reference model separately MRMS provides correct results, while unstable (for the Head model, results are very poor); these good results are due to the fact that this corpus is easier than the others because it considers only one type of distortion (simplification) which is applied uniformly. However our method demonstrates a significant improvement in term of Pearson correlation and a very good stability of the results over the models. When considering the whole corpus, once again our method clearly outperforms its counterparts (80% Pearson correlation against 34% for MRMS).

Table 4 provides the processing time for our algorithm (forward and backward distance processing) and the Metro tool [CRS98] for comparison. The Horse is compared with itself at different resolutions. The overall time remains reasonable with the most time consuming part being the computation of spherical neighborhoods at highest scale ($h = 4\varepsilon$). In case of high timing constraints the user can choose to compute only the forward distance and only at the finest scale ($h = 2\varepsilon$), a lot of computation time will be saved, for instance the whole comparison with respectively 50K and 113K vertices

Table 1: *Spearman (r_S) and Pearson (r_P) correlation values (%) between Mean Opinion Scores and values from the metrics for the General-purpose corpus.*

	Armadillo		Dyno		Venus		Rocker		**Whole Corpus**	
	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P
Hausdorff	69.5	30.2	30.9	22.6	1.6	0.8	18.1	5.5	13.8	1.3
RMS	62.7	32.2	0.3	0.0	90.1	77.3	7.3	3.0	26.8	7.9
GL_1 [KG00]	70.2	43.7	15.5	3.2	**92.0**	**80.2**	14.2	8.4	33.1	12.6
GL_2 [SCOT03]	77.8	55.5	30.6	12.5	91.0	77.6	29.0	17.1	39.3	18.0
$3DWPM_1$ [CGEB07]	65.8	35.7	62.7	35.7	71.6	46.6	87.5	53.2	69.3	38.3
$3DWPM_2$ [CGEB07]	74.1	43.1	52.4	19.9	34.8	16.4	37.8	29.9	49.0	24.6
MSDM [LDD*06]	**84.8**	70.0	73.0	56.8	87.6	72.3	**89.8**	75.0	73.9	56.4
*MSDM*2 (**Our metric**)	81.6	**72.8**	**85.9**	**73.5**	89.3	76.5	89.6	**76.1**	**80.4**	**66.2**

Table 2: *Spearman (r_S) and Pearson (r_P) correlation values (%) between Mean Opinion Scores and values from the metrics for the Masking corpus.*

	Armadillo		Lion		Bimba		Dyno		**Whole Corpus**	
	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P
Hausdorff	48.6	37.7	71.4	25.1	25.7	7.5	48.6	31.1	26.6	4.1
RMS	65.7	44.6	71.4	23.8	71.4	21.8	71.4	50.3	48.8	17.0
GL_1 [KG00]	65.7	44.4	37.1	22.4	20.0	19.8	71.4	50.0	42.0	15.7
GL_2 [SCOT03]	65.7	44.2	20.0	21.6	20.0	18.0	60.0	49.8	40.1	14.7
$3DWPM_1$ [CGEB07]	58.0	41.8	20.0	9.7	20.0	8.4	66.7	45.3	29.4	10.2
$3DWPM_2$ [CGEB07]	48.6	37.9	38.3	22.0	37.1	14.4	71.4	50.1	37.4	18.2
MSDM [LDD*06]	**88.6**	**72.2**	**94.3**	78.0	42.9	33.9	**100**	**91.7**	65.2	47.9
*MSDM*2 (**Our metric**)	**88.6**	65.8	**94.3**	**87.5**	**100**	**93.7**	**100**	91.5	**89.6**	**76.2**

for $\mathcal{M}_d$ and $\mathcal{M}_r$ takes only 9.3 seconds (less than METRO), furthermore this *fast* configuration still provides correct results: respectively 70.7%, 71.7% and 68.5% Pearson correlations over the whole corpus respectively for the general-purpose, masking and simplification databases; nevertheless, results are less stable among the models compared with the *standard* configuration (i.e. multiscale and symmetric).

Table 4: *Processing times (in seconds, for a 2GHz laptop) of our metric and METRO for objects of different sizes.*

$\mathcal{M}_d/\mathcal{M}_r$	$GMD_{\mathcal{M}_d \to \mathcal{M}_r}$	$GMD_{\mathcal{M}_r \to \mathcal{M}_d}$	METRO
113K/113K	50.0	50.0	11.5
50K/113K	14.7	44.9	10.1
1K/113K	4.7	42.6	10.5
50K/50K	11.5	11.5	4.6
1K/50K	1.7	9.7	4.1

6. Application to the evaluation of progressive compression algorithms

In previous sections we have presented and compared a new algorithm for visual distortion assessment of a 3D model with respect to a reference one, without any constraint on the connectivity. This new metric provides an alternative, closer to the human visual perception, to traditional distortion metrics like RMS or Hausdorff distance for evaluating or driving any kind of processing operations. In the following section we demonstrate that our MSDM2 metric is more accurate than the widely used MRMS for assessing the rate-distortion performance of progressive compression algorithms. The main idea of progressive compression is to transmit a simple coarse mesh (low-resolution), and a refinement sequence allowing the viewer to update incrementally the level of details of the mesh during the transmission; most of existing techniques are evaluated with the rate-distortion curve, the distortion being computed usually by the MRMS error. Figure 7 illustrates these performance curves for three of the most recent and efficient algorithms: the Wavemesh method from Valette and Prost [VP04], the Octree method from Peng and Kuo [PK05] and the Iterative Parametric Refinement (IPR) approach from Valette et al. [VCP09]. In the left figure, where the distortion is computed using MRMS error, the three methods seem close in performance; more precisely at 6 bits per vertex the Wavemesh method seems to deliver a better level of details than the IPR method. However when actually visualizing these levels of details on figure 8, we see that the best visual quality is reached by the IPR, moreover it appears also that the 8 *bpv* Octree version

Table 3: *Spearman (r_S) and Pearson (r_P) correlation values (%) between Mean Opinion Scores and values from the metrics for the Simplification corpus.*

	Bunny		Foot		Head		Lung		Strange		Whole Corpus	
	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P	r_S	r_P
Hausdorff	39.5	14.3	**94.3**	84.8	**88.6**	53.0	88.6	64.9	37.1	27.4	49.4	25.5
MRMS	77.1	79.2	**94.3**	71.1	42.9	23.1	**94.3**	71.3	94.3	92.4	64.3	34.4
MSDM2 **(Our metric)**	**94.3**	**96.3**	77.1	**96.7**	**88.6**	**79.0**	65.7	**85.3**	**100**	**98.1**	**86.7**	**79.6**

is of worst quality; all these observations are in contradiction with the rate-MRMS distortion curves. However when computing the distortion using our MSDM2 metric (see figure 8 on the right) then all these visual observations appear on the curves: the 6 *bpv* version of the IPR is better than its Wavemesh counterpart and both are better than the 8 *bpv* version of the Octree.
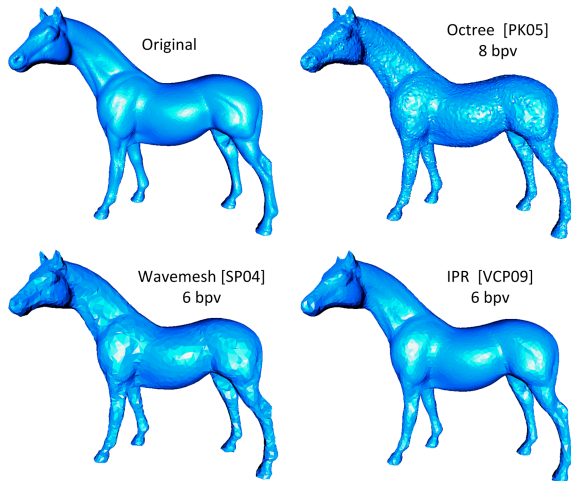
Figure 8: *Different levels of details of the Horse model (20K vertices), created using different progressive compression algorithms.*

6.1. Conclusions

We have presented a new multiscale metric for visual distortion assessment of 3D meshes; this metric has proven to outperform its counterparts in term of correlation with the human judgment, on three subjective databases. This metric has been applied for rate-distortion performance evaluation of recent progressive compression algorithms and has shown to deliver more relevant results than classical root mean square distance regarding the visual quality of the levels of details. More generally such visual distortion/quality assessment metrics provide a new paradigm for the evaluation, control and optimization of many kinds of processing operations.

Our future works will focus on integrating photometric data which may be associated with vertices and which participate to a large extent in the visual perception of the 3D models.

References

[ACSD*03] ALLIEZ P., COHEN-STEINER D., DEVILLERS O., LÉVY B., DESBRUN M.: Anisotropic polygonal remeshing. *ACM Transactions on Graphics 22*, 3 (2003), 485. 4

[ATW09] ALLIEZ P., TAYEB S., WORMSER C.: AABB Tree. CGAL 3.5 edition, 2009. 5

[AvMS10] AYDIN T. O., ČADÍK M., MYSZKOWSKI K., SEIDEL H.-P.: Video quality assessment for computer graphics applications. *ACM Transactions on Graphics 29*, 6 (2010), 1. 2

[BHM09] BIAN Z., HU S., MARTIN R.: Evaluation for small visual difference between conforming meshes on strain field. *Journal of Computer Science and Technology 24* (2009), 65–75. 2

[BM98] BOLIN M., MEYER G.: A perceptually based adaptive sampling algorithm. In *ACM Siggraph* (1998), pp. 299–309. 2

[CGEB07] CORSINI M., GELASCA E. D., EBRAHIMI T., BARNI M.: Watermarked 3-D Mesh Quality Assessment. *IEEE Transactions on Multimedia 9*, 2 (2007), 247–256. 2, 3, 8, 9

[CH07] CHANDLER D., HEMAMI S.: VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing 16*, 9 (2007), 2284–2298. 2

[CPJ06] CHO J., PROST R., JUNG H.: An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms. *IEEE Transactions on Signal Processing 55* (2006), 142–155. 3

[CRS98] CIGNONI P., ROCCHINI C., SCOPIGNO R.: Metro: Measuring Error on Simplified Surfaces. *Computer Graphics Forum 17*, 2 (1998), 167–174. 8

[CSM03] COHEN-STEINER D., MORVAN J.: Restricted delaunay triangulations and normal cycle. In *19th Annu. ACM Sympos. Comput. Geom.* (2003). 4

[Dal93] DALY S.: *The visible differences predictor: an algorithm for the assessment of image fidelity*. MIT Press, Cambridge, 1993, pp. 179–206. 2
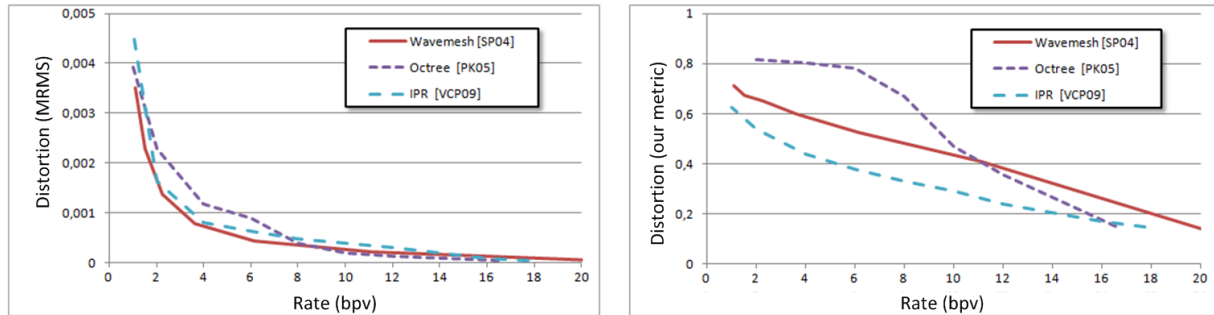
Figure 7: *Rate-distortion curves for three progressive compression algorithms.* Left*: the distortion is measured using the MRMS geometric distance,* Right*: The distortion is evaluated using MSDM2.*

[GECB05] GELASCA E., EBRAHIMI T., CORSINI M., BARNI M.: Objective evaluation of the perceptual quality of 3D watermarking. *IEEE International Conference on Image Processing 2005* (2005), I–241. 2, 8

[I.T00] I.T.U: BT.500-10: Methodology for subjective assessment of the quality of television pictures, 2000. 7

[KG00] KARNI Z., GOTSMAN C.: Spectral compression of mesh geometry. In *ACM Siggraph* (2000), pp. 279–286. 2, 8, 9

[Lav09] LAVOUÉ G.: A local roughness measure for 3D meshes and its application to visual masking. *ACM Transactions on Applied Perception (TAP) 5*, 4 (2009). 2, 4, 5, 8

[LC10] LAVOUÉ G., CORSINI M.: A comparison of perceptually-based metrics for objective evaluation of geometry processing. *IEEE Transactions on Multimedia 12*, 7 (2010), 636–649. 1, 3, 8

[LDD*06] LAVOUE G., DRELIE GELASCA E., DUPONT F., BASKURT A., EBRAHIMI T.: Perceptually driven 3D distance metrics with application to watermarking. In *Proceedings of SPIE* (2006), vol. 6312, pp. 63120L–63120L–12. 1, 2, 3, 4, 5, 8, 9

[LT99] LINDSTROM P., TURK G.: Evaluation of memoryless simplification. *IEEE Transactions on Visualization and Computer Graphics 5*, 2 (1999), 98–115. 3

[LT00] LINDSTROM P., TURK G.: Image Driven Simplification. *ACM Transactions on Graphics 19*, 3 (2000), 204–241. 2

[LVJ05] LEE C., VARSHNEY A., JACOBS D.: Mesh saliency. In *ACM Siggraph* (2005), pp. 659–666. 2, 3, 4, 6

[MMBC10] MCNAMARA A., MANIA K., BANKS M., C: Perceptually-motivated graphics, visualization and 3D displays. *ACM SIGGRAPH Courses* (2010). 2

[PK05] PENG J., KUO C.-C. J.: Geometry-guided progressive lossless 3D mesh coding with octree (OT) decomposition. *ACM Transactions on Graphics 24*, 3 (2005). 9

[QM08] QU L., MEYER G.: Perceptually guided polygon reduction. *IEEE Transactions on Visualization and Computer Graphics 14*, 5 (2008), 1015–1029. 2

[Red01] REDDY M.: Perceptually optimized 3D graphics. *IEEE Computer Graphics and Applications 21*, 5 (2001), 68–75. 2

[RFWB07] RAMANARAYANAN G., FERWERDA J., WALTER B., BALA K.: Visual equivalence: towards a new standard for image fidelity. In *ACM Siggraph* (2007), ACM, pp. 76–es. 2

[RR01] ROGOWITZ B. E., RUSHMEIER H.: Are image quality metrics adequate to evaluate the quality of geometric objects? *Proceedings of SPIE* (2001), 340–348. 2

[SCOT03] SORKINE O., COHEN-OR D., TOLDEO S.: High-pass quantization for mesh encoding. In *Eurographics Symposium on Geometry Processing* (2003), pp. 42–51. 2, 8, 9

[SSFM09] SILVA S., SANTOS B. S., FERREIRA C., MADEIRA J.: A Perceptual Data Repository for Polygonal Meshes. *International Conference in Visualisation* (2009), 207–212. 8

[Tau00] TAUBIN G.: Geometric signal processing on polygonal meshes. *Eurographics State of the Art Reports 4* (2000). 3

[VCP09] VALETTE S., CHAINE R., PROST R.: Progressive lossless mesh compression via incremental parametric refinement. In *Eurographics Symposium on Geometry Processing* (2009), vol. 28, pp. 1301–1310. 9

[VP04] VALETTE S., PROST R.: A wavelet-based progressive compression scheme for triangle meshes : Wavemesh. *IEEE Transactions on Visualization and Computer Graphics 10*, 2 (2004), 123–129. 9

[VS11] VÁŠA L., SKALA V.: A perception correlated comparison method for dynamic meshes. *IEEE transactions on visualization and computer graphics 17*, 2 (2011), 220–30. 2

[WB06] WANG Z., BOVIK A. C.: *Modern Image Quality Assessment*, vol. 2. Morgan & Claypool Publishers, 2006. 2

[WBSS04] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (2004), 600–612. 2, 3, 5, 7

[WLDB11] WANG K., LAVOUÉ G., DENIS F., BASKURT A.: Robust and blind mesh watermarking based on volume moments. *Computers & Graphics 35*, 1 (2011), 1–19. 3, 6, 7

[WSB03] WANG Z., SIMONCELLI E., BOVIK A.: Multiscale structural similarity for image quality assessment. *IEEE Asilomar Conference on Signals, Systems and Computers 2*, 1 (2003), 1398–1402. 2, 4

[ZMZ11] ZHANG L., MOU X., ZHANG D.: FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing 99* (2011), 1–1. 2

[ZZDZ10] ZHU Q., ZHAO J., DU Z., ZHANG Y.: Quantitative analysis of discrete 3D geometrical detail levels based on perceptual metric. *Computers & Graphics 34*, 1 (2010), 55–65. 2, 3