Watermarked 3D Object Quality Assessment

Massimiliano Corsini, Elisa Drelie Gelasca and Touradj Ebrahimi
LTS-1 (Multimedia Signal Processing Lab)
Signal Processing Institute (ITS)
Swiss Federal Institute of Technology in Lausanne (EPFL)
LTS1-ITS-STI-EPFL, 1015 Lausanne, Switzerland
Technical Report No. 2004/29

Abstract

This work concerns the developing of new perceptual metrics for 3D watermarking quality assessment. Any watermarking algorithm, to be effective, requires that the distortions it inevitably introduces into the watermarked media is imperceptible. This requirements is particularly severe for watermarking of 3D objects where the visual quality of the original model has to be preserved, i.e. the visual aspect of the watermarked object have to be the same of the original one. Several methods based on the knowledge of Human Visual System (HVS) have been developed to achieve this goal for still images and video watermarking. Since now, no similar techniques for watermarking of 3D objects exist. Here, we propose a novel experimental methodology for subjective evaluations of 3D objects and two perceptual metrics for quality assessment of watermarked 3D objects. Such metrics have been developed combining roughness estimation of model surface with psychophysical data collected by subjective experiments based on the proposed methodology. The performances of the proposed metrics are deeply analyzed.

Index Terms

Visual Quality Assessment, Perceptual Metrics, 3D Watermarking, Mesh Watermarking.

Contents

I. Introduction

In this work, we propose two perceptual metrics for the quality assessment of watermarked 3D objects[1]. The reasons of proposing perceptual metrics are the evaluation and comparison of *perceptual artifacts* introduced by 3D watermarking algorithms. The final aim of evaluation is to minimize extraneous detail introduced by the watermarking by modulating the watermarking insertion in order to obtain little or no perceptual artifacts. The second is to use such metrics for comparing the performance of different 3D watermarking algorithms on the basis of the artifacts perceived on the 3D model.

Effectively, many perceptual metrics have been proposed in the image domain. A possible approach could be to simply apply such image-based perceptual metrics to the final rendered images of the 3D model. The main problem of this approach is that the perceived degradation of still images may not be adequate to evaluate the perceived degradation of the equivalent 3D model. The approach we chose is to evaluate in which way the human visual system perceives geometric defects directly on the the model's surface and to build an ad-hoc perceptual metric. In such a case, subjective experiments that deal directly with the 3D models are needed. An accurate choice of the 3D model rendering conditions has to be done. Then, through a suite of interactive subjective experiments the perceived amount of distortion has to be investigated.

We propose two subjective experiments with different purposes. The first experiment (Experiment I), is carried out to investigate the perception of artifacts caused by a single watermarking algorithm on 3D models and to find suitable metrics to measure artifacts' perceptual severity. Subsequently, two metrics based on roughness estimation of the model's surface has been devised to perceptually measure the amount of visual distortions introduced by the watermarking algorithm over the surface of the model. Another experiment (Experiment II) is conducted in order to validate the proposed metrics with other watermarking algorithms.

The report is organized as follows. Previous works related to perceptual image watermark insertion, to mesh simplification and perceptually-guided rendering are reviewed in Section II. In Section III we describe the artifacts introduced by common 3D watermarking algorithms. Our experimental methodology to carry out subjective experiments on 3D model quality evaluation is proposed in Section IV. Data analysis is performed in Section V. Section VI describes the proposed metric. Finally, results are presented and discussed in Section VII.

II. Related Work

The knowledge of human visual system has been widely applied by *perceptual image watermarking* to obtain high quality watermarked images, i.e. watermarked images indistinguishable from the original ones. Our investigation concerns the extension of this idea to 3D watermarking. The goal is to develop a perceptual metric to estimate the perception of visual artifacts introduced by watermarking. The evaluation of the visual impairment of a watermarking algorithm can be used to adjust its watermarking parameters in order to obtain a watermarked model that looks like the original one. Perceptual metrics are not limited to perceptual watermarking, but they have been used also in two other fields of Computer Graphics: *mesh simplification* and *perceptually-guided rendering*. The three issues related to our investigations, concerning perceptual image watermarking, mesh simplification and perceptually-guided rendering, will be discussed in the following.

A. Perceptual Image Watermarking

It is widely known among researchers working in digital watermarking that HVS characteristics have to be carefully considered in order to minimize the visual degradation introduced by the watermarking process while maximizing robustness [1], [2], [3]. Considering a noisy image, some aspects of the human visual perceptions that anyone can easily experienced are that *i)* disturbs in the uniform regions of the image are more visible than the ones in textured regions, *ii)* noise is more easily perceived around edges and *iii)* the human eye is less sensitive to disturb in very dark and very bright regions. These basic mechanisms of the human visual perception can be mathematically modeled considering two main concepts: the *contrast sensitivity function* (CSF) and the *contrast masking* model. CSF is a measure of the responsiveness to contrast for different spatial frequencies. Typically, CSF models estimate the capability of the human eye to perceive sinusoidal patterns on a uniform background. The contrast perception varies with the frequency of the sinusoidal pattern, with the orientation of the pattern, with the observer's viewing angle and with the luminance of the background where the stimulus is presented. Many analytical expressions of CFS can be found in literature, one of the most used is the Barten's model [4].

The *masking effect* concerns the visibility reduction of one image component due to the presence of other components. In other words, while CSF considers the visual perception of a sinusoidal pattern on an uniform background the *visual masking model* considers the perception of a sinusoidal pattern over spatially changing background. The non-uniform

[1]We assume that the 3D objects are represented as polygonal meshes.

background may be modelled with another sinusoidal pattern with different properties. Some models of visual masking has been developed by Watson [5], [6] and by Legge and Foley [7].

Many methods have been proposed so far to exploit the models of the HVS characteristics to improve the effectiveness of existing watermarking system [3], [8]. We can divide the approaches proposed so far into theoretical [8], [9], [10] and heuristic [11], [12]. Even if theoretically grounded approach to the problem would be clearly preferable, heuristic algorithms sometimes provide better results due to some problems with the HVS models currently in use [12], [13].

B. Mesh simplification

Mesh simplification regards the reduction of the number of vertices and triangles of a polygonal mesh model preserving its visual appearance. In general, the simplification process is driven by a similarity metric that measures the impact of the changes of the model after each simplification step. So, one of the most important characteristics of a simplification method is the *error metric* it uses. Two kind of metrics are considered for simplification: geometric metrics and (perceptual) image-based metrics.

1) Geometry-based metrics: Metrics for simplification are commonly used for two distinct purposes; evaluating the quality of the final model and determining where and how to simplify the model. The most used global geometry-based metrics for off-line quality evaluation of 3D models are based on the Hausdorff distance.

The *Hausdorff distance* is one of the most well-known metrics for making geometric comparisons between two point sets. Assuming that the shortest distance between a point x and a set of points Y (e.g. the vertices of the 3D model) is the minimum Euclidean distance, i.e:

$$d(x,Y) = \min_{y \in Y} d(x,y) \tag{1}$$

the asymmetric Hausdorff distance between two points sets can be defined as:

$$\vec{d}_\infty(X,Y) = \max_{x \in X} \min_{y \in Y} d(x,y) \tag{2}$$

Since $\vec{d}_\infty(.)$ is not symmetric , i.e. $\vec{d}_\infty(X,Y) \neq \vec{d}_\infty(Y,X)$, so this distance is not a metric in mathematical sense. To obtain symmetry it can be redefined as:

$$d_\infty(X,Y) = \max\left(\vec{d}_\infty(X,Y), \vec{d}_\infty(Y,X)\right) \tag{3}$$

The (3) is usually referred as the *maximum geometric error*. This metric is not able to catch well geometric similarity since a single point of the set X, or Y, can determines the Hausdorff error. One possible alternative based on the average deviation that best measures geometric similarity is:

$$\vec{d}_1(X,Y) = \frac{1}{\mathcal{A}_X} \int_{x \in X} d(x,Y) dX \tag{4}$$

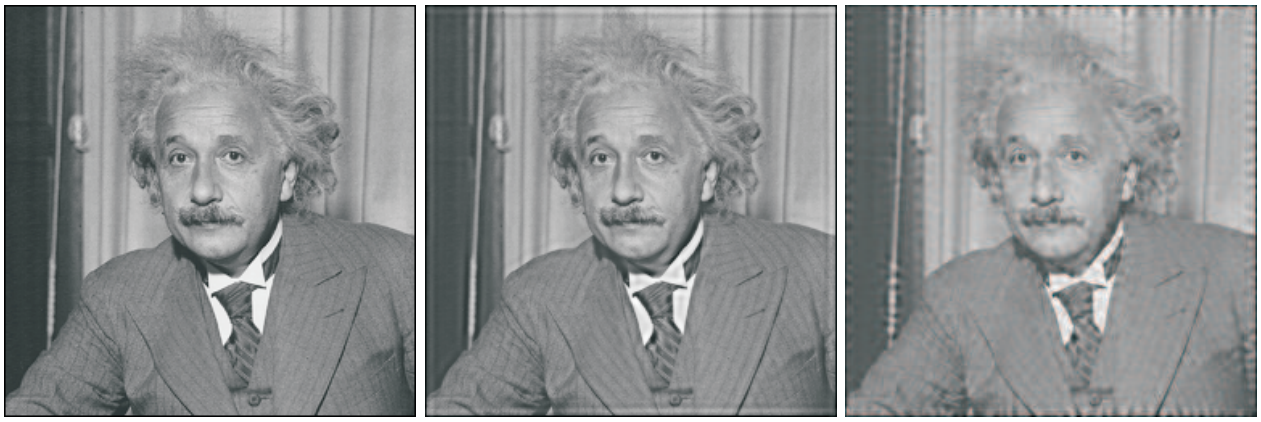
where $\mathcal{A}_X$ is the area of the surface X. Even this metric is asymmetric. The symmetric version of this metric assumes the following form:

$$d_1(X,Y) = \frac{\mathcal{A}_X}{\mathcal{A}_X + \mathcal{A}_Y} \vec{d}_1(X,Y) + \frac{\mathcal{A}_Y}{\mathcal{A}_X + \mathcal{A}_Y} \vec{d}_1(Y,X) \tag{5}$$

and it is usually referred to as *mean geometric error*. Two tools for geometric meshes comparison based on the maximum (3) and on the mean geometric error (5) are the Metro [14] and the Mesh [15] tool. Several researchers have proposed other geometry-based metrics to evaluate 3D model quality. Most of them are variations of the $d_\infty(.)$ and $d_1(.)$ metrics. In Section VII, we will analyze the performance of these two geometric metrics in the case of 3D watermarking quality evaluation.

2) Image-based metrics: Image metrics are adopted in several graphics applications. In fact, since most computer graphics algorithms produce images, it makes sense to evaluate their results using image differences instead of metrics based on geometry. A lot of simple images metrics such as the Root Mean Square (RMS) and the Peak Signal Noise Ratio (PSNR) have been widely used in the past, but, such metrics are not able to measure the differences between two images perceived by a human observer [16]. For example, figure 1 shows that the values or RMS do not correlate with the perception of image distortions. For this reason, nowadays, most applications move to perceptual-based image metrics. Two of the most perceptually accurate metrics for comparing images are the *Visual Difference Predictor* by Daly [17] and the *Sarnoff model* developed by Lubin [18]. Both of these metrics include models of different stages of the human visual system, such as, opponent colors, orientation decomposition, contrast sensitivity and visual masking.

Concerning *perceptually-based mesh simplification*, Lindstrom and Turk [19] propose an image-driven approach for guiding the simplification process: the model to be simplified is rendered by considering several viewpoints and an image quality metric, based on a simplified version of the Sarnoff Model [18], is used to evaluate the perceptual impact of the simplification operation. More recently, Luebke et al. [20] developed a view-dependent simplification algorithms based



Original image | Minimum Perceptual Distortion (RMS = 9.0) | Maximum Perceptual Distortion (RMS = 8.5)

Fig. 1. Image distortions and RMS metric (from Teo and Heeger [16]).

on a simple model of CSF that takes into account texture and lighting effects. This method provides also an accurate modelling of the scale of visual changes by using parametric texture deviation to bound the size (represented as spatial frequencies) of features altered by the simplification. Other studies related to perceptual issues in mesh simplification have been conducted by Rogowitz and Rushmeier [21] and by Yixin Pan et al. [22]. Rogowitz and Rushmeier analyze the quality of simplified models perceived by human observers in different lighting conditions by showing to the observers still images and animations of the simplified objects. From the experiments they draw a lot of interesting conclusions. The most important one is that the perceived degradation of the still images cannot be adequate to evaluate the perceived degradation of the equivalent animated objects. This result suggests that an experimental methodology to evaluate the perceived alterations of 3D objects should rely on interaction with the model.

C. Perceptually-Guided Rendering

The aim of perceptually-guided rendering is to accelerate photo-realistic rendering algorithms to avoiding computation for which the final result will be imperceptible.

One of the first works of this type has been done by Reddy [23]. Reddy analyzed the frequency content of 3D objects in several pre-rendered images and used these results to select the "best" version of the objects from a pool of models representing the same shape with different level of details in order to speed-up the visualization of a virtual environment. If the high-resolution version of the model differs only at frequencies beyond the modeled visual acuity or greatest perceptible spatial frequency, the system selects a low-resolution version of the model.

Other remarkable works in this field include Bolin and Meyer [24] that use a simplified Sarnoff Visual Discrimination Model [18] to speed-up rendering techniques based on sampling (e.g. Monte Carlo Ray Tracing), Myszkowski et al. [25] that incorporate spatio-temporal sensitivity in a variant of Daly Visual Difference Predictor [17] to create a perceptually based animation quality metric (AQM) to accelerate the generation of animation sequences and Ramasubramanian et al.[26] that use perceptual models to improve global illumination techniques used for realistic image synthesis.

Another excellent work related to study of human visual perception in rendering is the one of Ferwerda and Pattanaik [27]. In this work a sophisticated perceptual metric for the evaluation of how much a visual pattern, i.e. a texture, hides geometry artifacts is proposed. The visual masking effect caused by texturing is taken into account by analyzing the final rendered images.

D. Our Approach

Our goal is to develop a perceptual metric that measures the human perception of geometric artifacts introduced over a 3D surface by watermarking. Two approaches to develop a perceptual metrics for 3D watermarking are possible. The first one follows the (perceptual) image-based approach for simplification seen before [19], [20]. Instead of driving the simplification process, the perceptual image metric can be used to evaluate the visual effects of the watermark insertion by computing the perceptual differences between several images rendered from the original and the watermarked model. This approach presents two advantages. First, since it is rendering-dependent, complex lighting and texturing effects can be taken into account in a natural way. The second advantage is that all possible kinds of visual artifacts can be evaluated with the same approach. The main disadvantages is that no information about how a human observer perceives the geometric artifacts is provided. The other possible approach is to evaluate in which way the human visual system perceives geometric distortions on the model surface and build an ad-hoc perceptual metric for geometric

artifacts. This approach is more interesting from a research viewpoint, since no similar studies have been conducted at now. The potential field of application of this kind of studies is not limited to improve imperceptibility of 3D watermarking algorithms, but, other Computer Graphics applications can benefit from them. For these reasons we decide to follow the second approach, i.e. to work directly on the geometry of the 3D model.

III. 3D watermarking algorithms and artifacts

Digital watermarking algorithms can be classified in algorithms that work in the asset domain, in the hybrid domain, in the transformed domain and in the compressed domain. For a complete review of watermarking algorithms of polygonal meshes we remind to [28]. Here, for each class of algorithms, we describe the geometric artifacts that those algorithms introduce in the watermarked model.

First of all, we consider the algorithms working in the asset domain. The algorithms based on topological embedding [29], [30] produce small geometric distortions that can be described by the addition of a small amount of noise to the position of the mesh vertices. Those cases whose only the connectivity of the mesh is used to embed the watermark, such as the TSPS and the MDP algorithms [29], the amount of introduced distortions is imperceptible, since topology changes usually do not produce noticeable visual effects. Concerning geometric features embedding we have to distinguish between those algorithms that embed the watermark using vertices position and those algorithms that are based on shape-related features, such as vertex normals. Changes in the vertices position produce the same effect of topology-driven embedding, i.e. a "noisy" watermarked surface, but, in this case the amount of distortions may be considerably high, due to the vertices displacements needed to embed the watermark. For example the Vertex Flood Algorithm [31], may introduce moderate-to-strong distortions depending on the embedding parameters. In the same manner, the method of Harte and Bors [32] may produce moderate distortions depending on how many vertices are watermarked and on the dimension of the bounding volume used. Shape-related algorithms, like the Normal Bin Encoding (NBE) [33] algorithm and the method of Wagner [34], instead, introduce artifacts that look very different from the noisy effect of the other techniques. This kind of surface alterations are difficult to describe and, usually, less perceptible than other artifacts.

The algorithms that works in the hybrid domain are able to spread the distortions over the whole surfaces of the models by introducing the watermark in the low resolution of the model. Thus result in a low perception of the watermark [35]. The amount of distortions produced by the wavelet-based method of Uccheddu et al. [36], heavily depends on the level of resolution used to embed the watermark. High is the level of resolution used higher the amount of visual impairment in the watermark model is. In the same way the algorithm of Kanai et al. [37], that it is also based on wavelet decomposition, may introduce geometric artifacts since all the levels of resolution are used to embed the watermark. Kanai proposes a geometric tolerance threshold to limit the introduction of such visual artifacts.

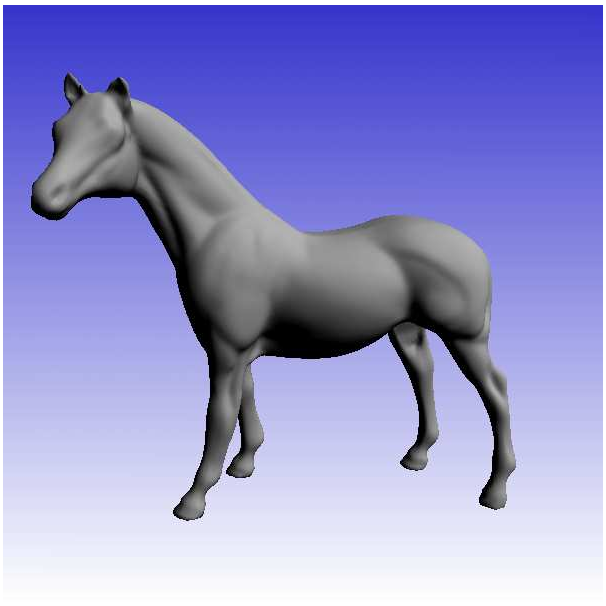
Concerning the transformed domain, the mesh spectral methods [38], [39], [40] cause a vertices perturbation due to the modifications of the mesh spectral coefficients, thus resulting in a moderate "noisy" watermarked surface. Ohbuchi [38] suggests to reduced this effect by watermarking those mesh spectral coefficients that are related to low frequencies content of the model.

Concluding, we can summarize what stated by saying that, in general 3D watermarking algorithms produce "noisy" surfaces. The characteristics of the noise depend on the specific algorithms; such noise can have different granularity and size, and may be uniform or not over the model surface. The watermarking techniques that do not introduce perceptible artifacts are typically those techniques that have relaxed robustness requirements. In our subjective experiments, that will be described in the next Section, we selected four different watermarking algorithms: the Vertex Flood Algorithm (VFA) [31], the Normal Bin Encoding (NBE) [33], the method of Kanai et al. [37], and the algorithm of Uccheddu et al. [36]. These last two algorithms are indicated in the following using the acronyms of the authors, i.e. KDK and UCB respectively. Figure 2 shows the artifacts introduced by these watermarking algorithms.

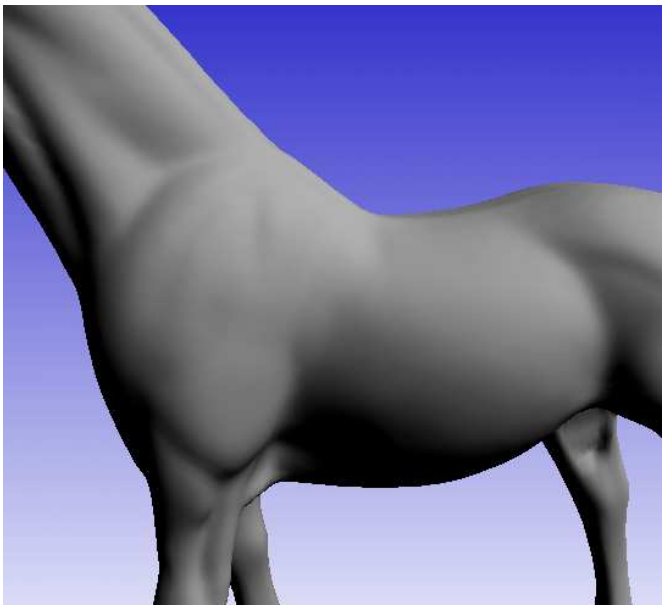
IV. Experimental Method

A set of standards and grading techniques to evaluate quality of video and multimedia content have been defined by ITU-R [41] and ITU-T [42]. However, there are no prescribed standards for the evaluation of 3D objects with impairments. In this chapter, we propose a method for subjective evaluation of 3D watermarked objects. This experimental methodology attempts to make subjective evaluations in this field more reliable, comparable and standardized.

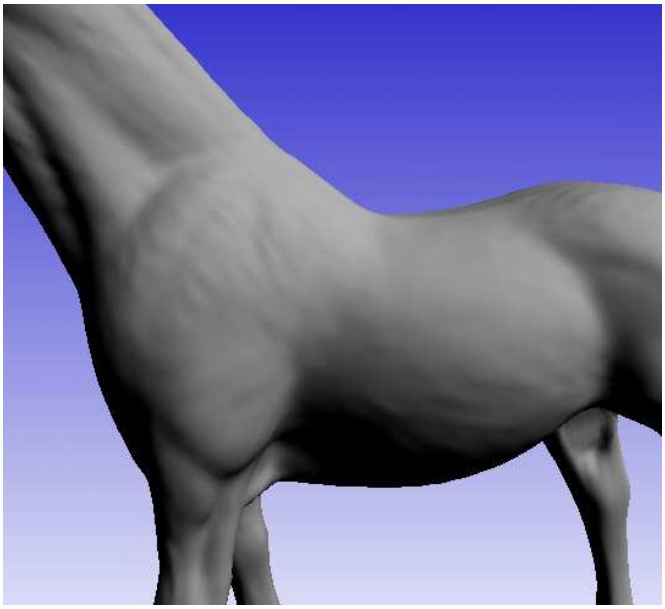
The starting point for the design of a subjective experiment for the evaluation of the quality of 3D objects is to define how to render the object under examination. By specifying appropriate rendering conditions, we aim at putting the human observer in favorable conditions to make a fair judgment on the three-dimensional object. The rendering conditions should not bias the human perception of the 3D model by privileging, for example, one view of the 3D object rather than another one.



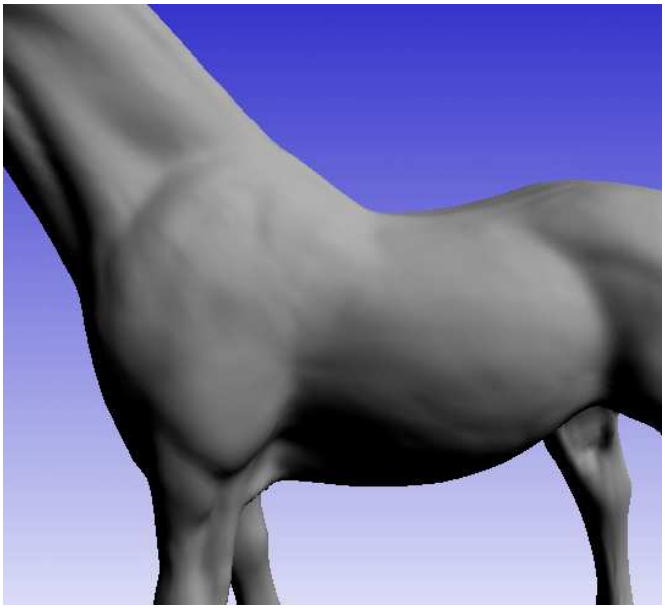
Original



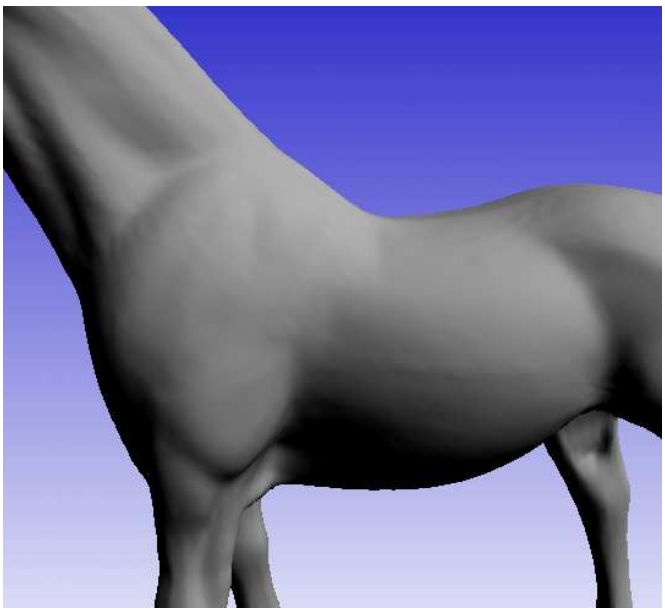
Original - detail



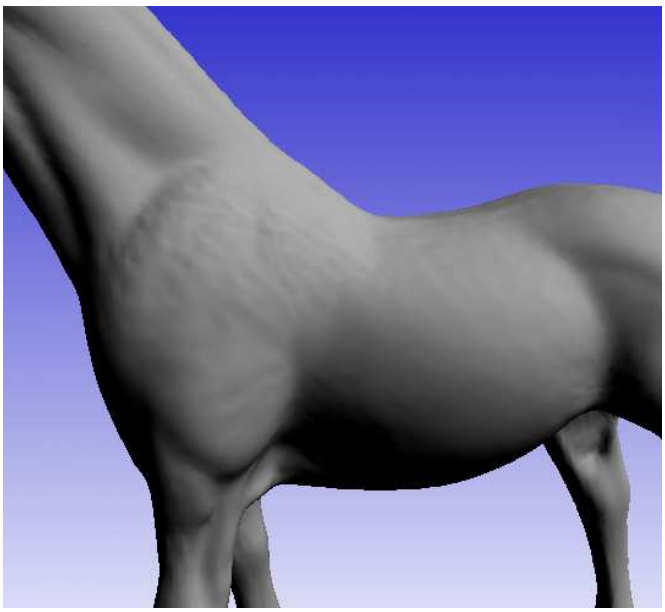
UCB algorithm



KDK algorithm



Normal Bin Encoding (NBE)



Vertex Flood Algorithm (VFA)

Fig. 2. Geometric defects introduced by 3D watermarking algorithms.

A. Rendering Conditions

The rendering of a three-dimensional model is accomplished via a combination of techniques such as associating a material to each surfaces of the model, applying various kind of light sources, choosing a lighting model, adding textures and so on. In our investigations we assumed that the rendering conditions have to be as simple as possible, because very few works have dealt with psychophysical tests of perceived 3D object quality as reported in Section II and no experimental data are available. Moreover, too many or complicated rendering effects would involve many and mutually linked aspects of spatial vision that have to be avoided to obtain more reliable results. Such results can be further extended taking into account more aspects of visualization techniques, such as the role of photorealism in the perception of impairments. In fact, by keeping plain but effective rendering conditions, we do not influence or bias the human perception and, then, the subjects' evaluation. The rendering conditions that we have chosen are described below.

- *Light sources.* Humans can see an object because photons are emitted from the surface of the object and reach the eyes of the viewer. These photons may come from other objects or from *light sources*. The common three different types of light sources are the *directional light*, the *point lights*, and the *spot lights* [43]. Point and spot light sources are also called *positional lights* because they are characterized by a location in space. Spotlights are not suitable for our purposes since this kind of light source could privilege some parts of the model better illuminated with respect to others. Multiple lights can cause effects that may confuse the human observer and provide contradictory results more complex to evaluate [44]. Additionally, the HVS tends to assume that the scene is illuminated by a single light source and that light illuminating the scene is coming from above. For all of these reasons, in our experiments, each model is illuminated with one white point light source located in a corner of the Object Bounding Box (OBB) of the 3D object. Each object we use is normalized to be put inside a cube (the OBB) of extreme corners of coordinates $\{-1,-1,-1\}$ and $\{1,1,1\}$ respectively. The light is positioned in the corner of coordinates $\{1,1,1\}$. Achromatic light is used in order to preserve the colors of the material.
- *Lighting and shading.* Lighting is the term that is used to designate interaction between material and light sources. Shading is the process of performing lighting computations and determining pixels' colors from them. A good choice of the lighting models and of the shading method succeeds in effectively communicating the scene to a human observer, such as the 3D shape of each object, the fine geometric details, and the spatial relationships between the objects in the scene. The influence of lighting model is very important to consider and it may affect the perceived quality of the 3D model considerably. Ideal lighting and shading conditions are very hard to find and some methods have been proposed to optimize rendering conditions in order to improve the perceptual quality of the rendered model [44]. To narrow the scope, we use a simple local illumination lighting model where only the diffusive component of the reflected light is considered. In fact, the diffusive component is well-connected to physical reality since it is based on Lambert's Law [45] which states that for surfaces that are ideally diffuse (totally matte, without shininess) the reflected light is determined by the cosine between the surface normal and the light vector. For this reason the diffuse component does not depend on the viewer's position making them suitable to unbias the human perception of the 3D object under examination. The specular component of the reflected light is not considered even if it would improve the photorealism of the objects. In fact, while the diffuse component catches the behavior of matte surfaces, the specular component models the shininess of the surfaces. The highlights created by the specular component help the viewer to better perceive the surface's curvature. Moreover, other more complex photorealistic issues such as self-shadowing are not introduced in our model, as they would unnecessarily complicate the experimental method and introduce too many variables to evaluate during results analysis. So, the implemented lighting model is:
$$I_r = I_{\mathtt{amb}} K_a + I_i K_d \min(0\,,\, \vec{N} \cdot \vec{L}) \tag{6}$$
where the vectors $\vec{N}$ is the surface normal at the considered point, $\vec{L}$ is the incident light direction vector and the constant K_a and K_d depend on the material properties. About shading methods that deal with triangular meshes, we can choose among flat, Gouraud and Phong shading [43]. Flat shading is not suitable for our purposes since it produces the well-known unnatural faceting effect. Since Gouraud and Phong shading produce almost the same visual effects if the models resolution, i.e. the number of triangles of the model, is high, we decide to use the Gouraud shading that it is more common and less computationally expensive than the Phong method. Finally, we have decided to show the model on a non-uniform background since an uniform background highlights too much the countour edges of 3D objects.
- *Texturing.* We want to evaluate the perception of artifacts on the surface of the 3D objects, hence textures or other effects are avoided as they usually produce a masking effects on the perceived geometry [27]. In fact, image texture mapping, bump mapping, and other kind of texturing may hide the watermarking artifacts. This is partially due to a perceptual effect called *visual masking*, in which frequency content in certain channels suppresses the perceptibly of other frequencies in that channel [27]. We do not account for visual masking, leaving that as an important and interesting area for future researches.

- *Material properties.* The color of a surface is determined by the parameters of the light source that illuminate the surface, by the lighting model used and by the properties of the surface's material. We consider only gray, *granite-like* objects. This choice is made for different reasons: first, if all models are seen as "statues" the subjects perceived all the models at the same manner and enough naturally; second, in this way we avoid the *memory color* phenomenon experimented in psychology studies [46]. This phenomenon regards the fact that an object's characteristic color influences the human perception of that object's color, e.g. shape such as heart and apple are characteristically red. A particular choice of a specific color for all the models could have mis-leaded the perceived quality of the object and introducing too many colors for different objects would have made the experimental method less general by introducing too many degrees of freedom.
- *Screen and Models Resolution.* The monitor resolution used in the experiments is 1280×600 and each model is displayed in a window of 600×600 pixels. The model occupies around 80% of such window and the resolution of all models ranges between 50.000 and 100.000 triangles. This screen resolution and the level of details of the models used in the experiments allow a good visualization of the model details, and hence of the model distortions. In particular the blurring effect of the Gouraud shading interpolation is negligible. Obviously such blurring effect increases when the subject observes the model closely. Also, the complexity of these 3D objects allows us to render them quickly (using a powerful graphics accelerator board) making fluid user-interaction possible. A minimum frame rate of 50 fps for each of the visualized models is guaranteed.
- *Interaction.* One essential feature of an interactive application is that objects are observed in motion. In our experimental method, we decide to allow the subject to interact with the model by rotation and zoom operations. The user interacts with the model using a mouse. Three-dimensional interaction is achieved by *ARCBALL rotation control* [47]. The motion of the 3D object is then interactively driven by the subject and not pre-registered like in other subjective experiments in literature [21], [22]. This avoid to detect less details in frames that pass quickly. It has to be mentioned that in previous work [21], [48], 2D images of 3D objects have been used for subjective experiments. The problem is that different static views of 3D objects can have significantly different perceived quality depending on the direction of illumination. Subjective experiments that were conducted to address this question suggest that judgments on still images do not provide a good predictor of 3D model quality [21]. Rogowitz's work confirmed that the use of still images produce results that are view-dependent and not correlate well with the real perception of the 3D object.

B. Experimental Procedure

Our test subjects were drawn from a pool of students from the École Polytechnique Fédérale de Lausanne (EPFL). They were tested one at a time and not specifically tested for visual acuity or color blindness. The 3D models were displayed on a 17-inch LCD monitor, with participants sitting approximately 0.4 meters from the display. The experiment followed a five-stage procedure [49]. The stages were: (1) oral instructions, (2) training, (3) practice trials, (4) experimental trials, (5) interview.

1) Oral Instructions: Test subjects have to be told how to perform the experiment. Prior to each experiment, the instructions, also known as *experiment scripts*, are elaborated to help the experimenter to define the task. The script contains details of what the experimenter should do at each step of the experiment. More importantly, it contains oral instructions that should be given to the subject to make sure she/he understands the task to be performed. An introductory explanation about what are 3D models and what is watermarking is given. Different sections of the instructions actually apply to all the various stages of the experiment. However, the most important part of the instructions comes before the training stage. After the subject is properly seated and made comfortable, the main task is explained. The instructions for both experiments can be found in Appendix.

2) Training: In each experiment the subject is asked to perform a task which consists of entering a judgment about an impairment detected in the 3D object. In order to complete this task subjects need to have an idea of how the original 3D objects with no impairments look. Therefore, a training session is included in the procedure which consists of displaying the four original models used to embed the watermark. In this phase, only the experimenter interacts with the graphical interface, the subject is asked to look carefully to the models displayed on the screen. In the next phase, a set of 3D models with the typical distortions introduced by watermarking is shown.

Another set of 3D objects is required to set a point on the scale of judgements. The end of the scale is set by 3D objects with the strongest defect (in this case the perceptually strongest watermarking). A total of 12 and 16 models were shown as worst examples respectively for the Experiment I and II. Therefore, the test subjects are instructed to pick the strongest stimulus in the training set and assign that stimulus a number from the upper end of the scale. In our experiments, the subject is asked to assign 10 to the worst example (on a discrete scale ranging from 0, which means that no distortions are perceived, to 10). However, due to the visual masking effect and the variety of the originals, it is not possible to anticipate which defects the subjects will consider worst or strongest. Therefore, the subjects are asked to record values greater than the nominal upper value if 3D objects are observed to exceed the expectations

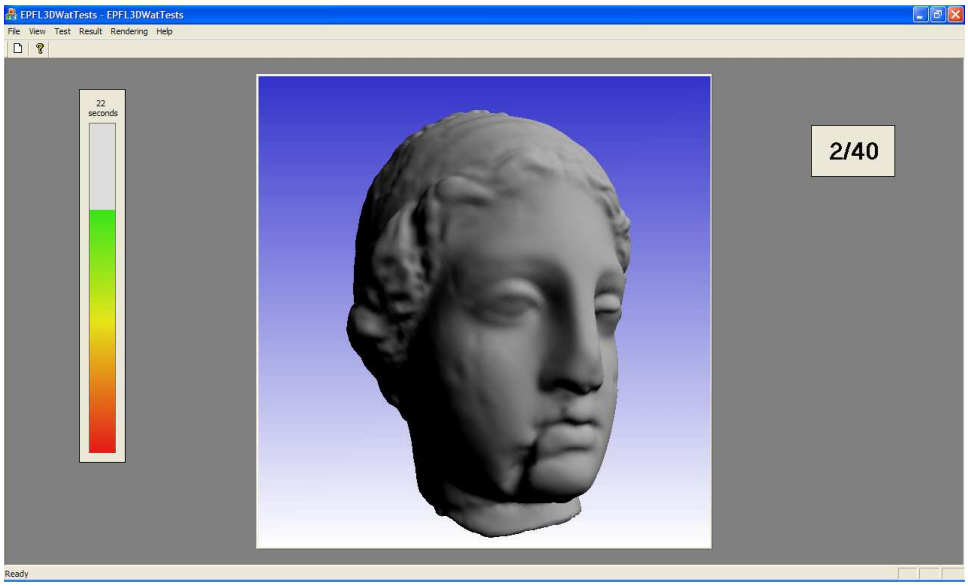


Fig. 3. The subject during the interaction with the model.

established by the training phase. For example, if a test subject perceives a defect twice as bad as the worst in the training set, he/she is asked to give it a value of 20. Finally, in the last phase of the training stage, subjects are told how to interact with the 3D models and learn how to use the graphical interface to visualize them.

3) Practice Trials: In practice trial stage, subjects are asked to make judgments for the first time. Because of the initial erratic responses of the subjects, ITU Recommendation [42] suggests to throw away the first five to ten trials of an experiment. In this methodology, instead of discarding the first trials, we include practice trials. It also gives other benefits. It exposes the test subject to 3D models throughout the impairment stage. It gives the test subject a chance to try the data entry and above all the chance to get familiar with the graphical interface for the virtual interaction with the 3D object (rotation and zooming). The number of practice trials is 6. The subject has to perform 3 tasks at most. The first one is to detect the distortion and he/she has to answer to the question *did you notice any distortion*. In case of positive answer, the subject has to give a score to indicate *how much the distortions are evident*. The subjects have 30 seconds at their disposal to interact with the model and to make their judgement. Then, they have to input the score in a dialog box. Figure 3 shows the interface for the subjective tests. On the left a time bar advices the user of the remaining interaction time. The box indicates the progression of the test by showing the number of the model under examination. The model is displayed in the center of the screen. Finally, the third question is *where* he/she noticed the distortion on the 3D model. To answer to this question he/she has to indicate, by selection, the part of the model where the distortions are the most evident (figure 4).

4) Experimental Trials: The subjective data is gathered during the experimental trials. In this stage, a complete set of 3D objects is presented in random order. To be more specific, for each experiment, several random-ordered lists of test watermarked 3D objects are generated. In this way the results of the test are made independent of the order in which the models are presented. All the test subjects see all the 3D objects. The number of test 3D objects is limited so that the whole experiment lasts no more than one hour. This limit translates to 40 models in Experiment I and 48 in Experiment II.

5) Interview: After the trials are complete, the test subjects are asked a few questions before they leave. The kind of questions depends on the experiment, mainly test subjects are asked for qualitative descriptions of the impairment. The questions asked in our case are:

1) Did you experience any problem correlated to a specific model in identifying the distortion?
2) How would you describe the distortions that you saw?
3) Have you general comments or remarks about the test?

These questions gather interesting information, for example, they are useful for categorizing the distortion features seen in each experiment and to help in the design of next experiments.

C. Experiment I

The goal of this first experiment is to make an initial study on the perception of artifacts caused by watermarking on 3D models and try to find metrics to measure their perceptual severity. The artifacts varying in strength and resolution were generated using UCB watermarking algorithm. The most important goal was to collect subjective data for a set of

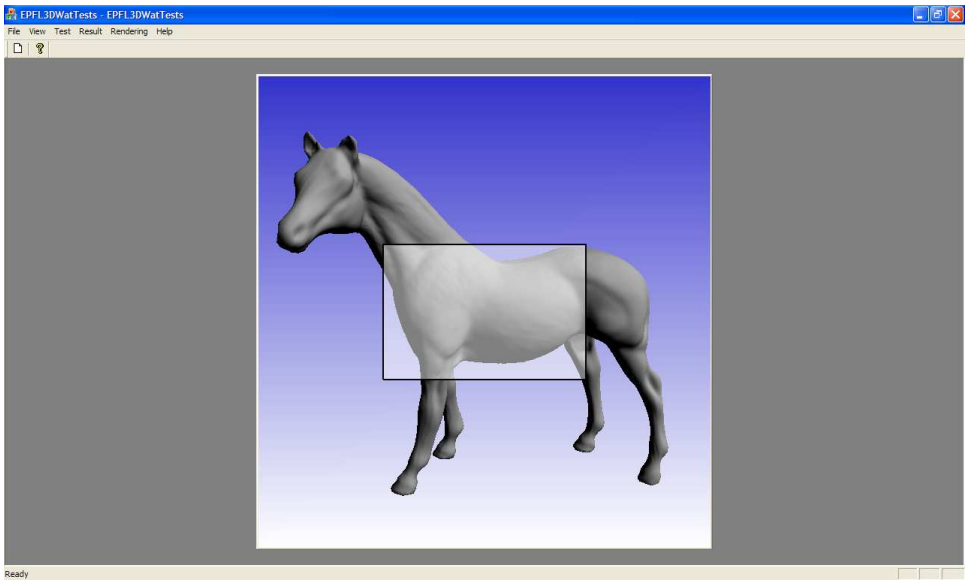


Fig. 4. The subject during the selection of the part where the distortions are more evident.

Variable	Values
Peak luminance	≤ 0.04
Maximum observation angle	10 degrees
Monitor resolution	1280×600
Interaction window resolution	600×600
Viewing Distance	$35 - 45$ cm
Monitor Size	17”

TABLE I
VIEWING CONDITIONS DURING SUBJECTIVE TEST.

3D objects of varying quality. This data set allowed us to confirm some basic findings such as geometric-based metric are not a good measure for subjective quality of 3D objects and compare existing metrics based on geometric distances between the original and the impaired model as already mentioned in Section II. Finally, we were interested in how the test subjects would describe the watermarking defects that were produced for this experiment. Technically, the worst test models exhibited strong noise effect. The appearance related questions included in the interview provided us with some directions to design a perceptually driven objective metric. With this purpose, 11 subjects performed detection and evaluation tasks. The methodology for the experiment have been just described in Section IV-B.

D. Generation of Stimuli

The test models for this experiment are generated by applying the UCB algorithm to four models: *“Bunny”*, *“Horse”*, *“Venus”* and *“Feline”* (Figure 5). These models are commonly used for algorithms comparison in Computer Graphics research community and can be found in several versions. Here we use the ones made publicly available by the Caltech Multi-resolution Modeling Group [50], which are the semi-regular version of the original irregular ones. This is made necessary since the UCB algorithm is specific for semi-regular meshes. These models are very suitable for perceptual studies thanks to the wide range of characteristics presented by their surface. For example, the Bunny model surface is full of bumps, all parts of the Horse model are smooth, Feline model presents a wide range of characteristics such as parts with high curvature, parts with low curvature, moderate bumps, smoothed parts and several protrusions, and the Venus Head model has the same range of characteristics of the Feline but without consistent protrusions.

The watermarking was uniformly distributed on all the surface of the 3D object. The amount of distortions introduced by the watermarking varied according to two parameters: *i)* the resolution level l that hosts the watermark and *ii)* the coefficient γ determining the strength of the watermark. Table II shows the values of the UCB algorithm parameters used for this experiment. Three amount of watermarking strength (low, medium and high) and three levels of resolution (low, medium and high) were applied to each model. In addition to the watermarked models, the 4 original models were included in the complete model set. In fact, the original may present impairments unrelated to the watermarking ones deliberately inserted into the test models. To separate the effects of the deliberate and the pre-existing defects,

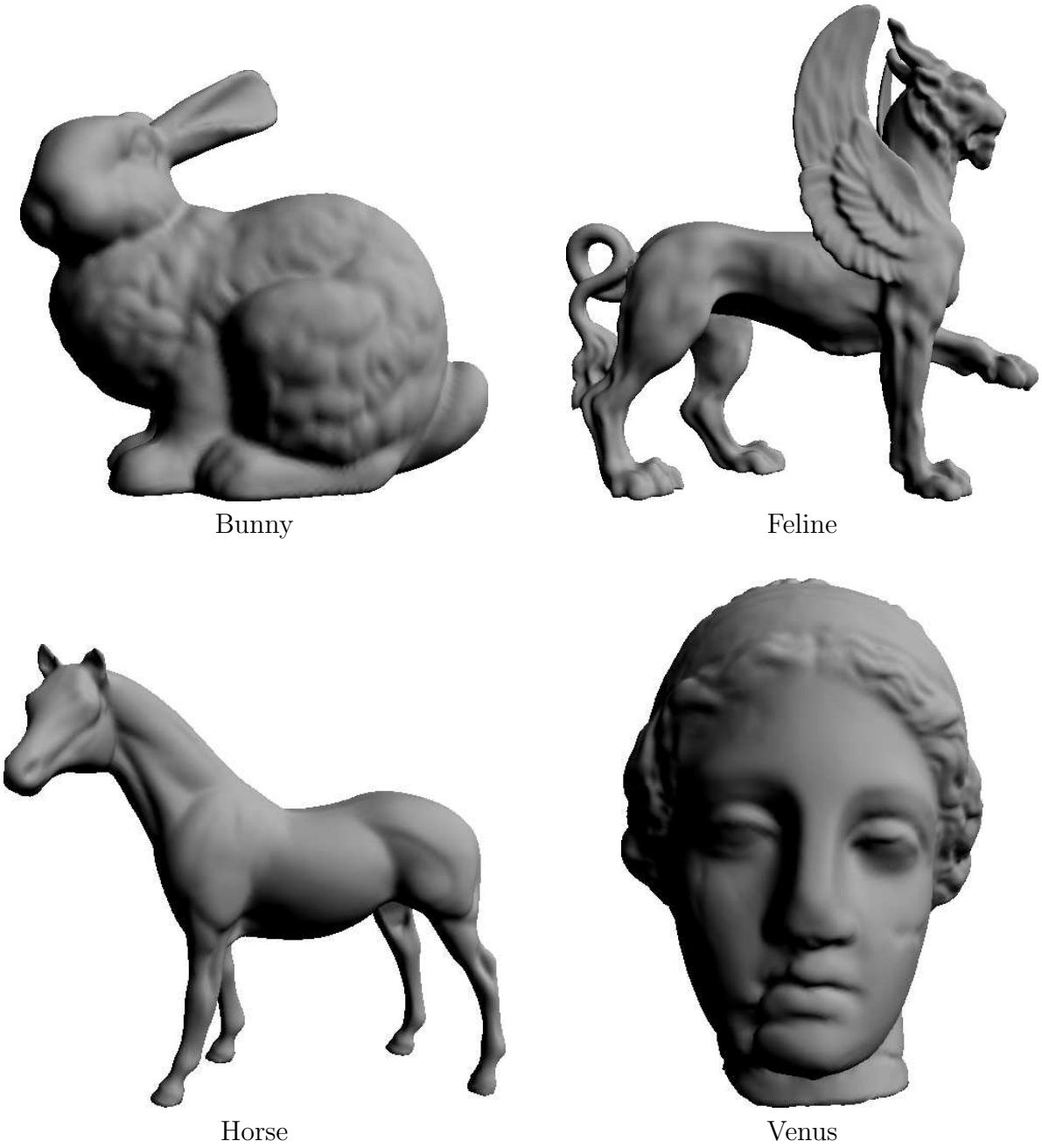


Fig. 5. Original models used for stimuli generations.

the originals had to be inserted. A total of 40 (4 originals × 3 watermarking strength × 3 resolution level + 4 originals) test models were used in the experiment.

Level of Resolution (l)	**Corresponding Value**	**Watermarking Power (γ)**	**Corresponding Value**
Low	4	Low	0.0003
Medium	3	Medium	0.0015
High	2	High	0.003

TABLE II
EXPERIMENT I: WATERMARKING PARAMETERS VALUES USED FOR EACH MODEL.

E. Experiment II

In Experiment I, test subjects evaluated differently watermarked models ranging from severely down to weakly visual impairments. Those different distortions' strengths were generated using a specific watermarking algorithm, i.e. the UCB algorithm. There are several major goals for Experiment II. First, we wanted to find whether watermarking algorithms generally produce impairments perceptually similar to those given by watermarking algorithm tested in Experiment I. Therefore, as previously stated, we chose three different watermarking algorithms: NBE,VFA and KDK. Second, with this experiment we wanted by means of subjective validation to test the perceptually-based objective metric for the quality assessment we inferred from Experiment I. Technically, the defect inserted are slightly different from the ones studied in the Experiment I (see figure 2). In fact, while the UCB algorithm produces an uniform kind of noise that can be described as an increase of the roughness of the watermarked surface, VFA produces a kind of noise that looks like marble streak, depending by the viewpoint. The artifacts of the KDK algorithm are the same of the UCB algorithm but due to the geometric tolerance introduced by Kanai to limit the visual impact of the watermarking the final visual effects of such distortions is not uniformly distributed over the model's surface. Concerning NBE, the visual aspect of its artifacts is very different from the one of the UCB, VFA and KDK and more difficult to perceive. The methodology for this experiment is practically the same of the Experiment I. The only difference is that no location information was gathered since the metric developed on the basis of the data collected in the Experiment I does not take into account the location information.

F. Generation of Stimuli

The test models for this experiment are generated using the same four original models of the Experiment I, i.e. the *"Bunny"*, *"Horse"*, *"Venus"* and *"Feline"* models. The three watermarking algorithms used to generate the watermarked models are the VFA, the NBE and the KDK algorithms. Each watermarking algorithm is characterized by its own embedding parameters that are qualitatively and quantitatively different. For an exact description of each parameters we remind to the literature. The watermarking parameters of the VFA are the number of clusters used to embed the bits (one bit for each cluster) and the maximum allowable distance (D_{MAX}) from the starting application points. NBE is characterized by the feature type used, by the number of bins (N_B), and by the search range (Δ_R) and the number of iterations (n_I) of the optimization process. KDK parameters are the selection threshold (δ_1), the geometric tolerance threshold (δ_2) and the least significant decimal digits used to embed the watermark (d_{w}); if $d_{\mathrm{w}} = 2$ this means that the second least significant decimal digits of the wavelet coefficients is modified to embed the watermark, $d_{\mathrm{w}} = 3$ indicates the third least significant decimal digits, and so on. The watermarking parameters for the three algorithms are reported in Table III. In our test set, we have tried to range from severely to weakly watermarked model as in Experiment I. The 11 level of impairment are also reported in Table III. A total of 48 (4 models $\times$ 11 watermarking settings + 4 originals) test models were used in this experiment.

Algorithm	**Watermarking Parameters**	**Impairment**
KDK1	$\delta_1 = 0.001$, $\delta_2 = 0.005$, $d_{\mathrm{w}} = 2$	medium-strong
KDK2	$\delta_1 = 0.001$, $\delta_2 = 0.008$, $d_{\mathrm{w}} = 3$	medium
KDK3	$\delta_1 = 0.001$, $\delta_2 = 0.02$, $d_{\mathrm{w}} = 3$	medium-strong
NBE1a	Feature Type I, $N_B = 80$, $\Delta_R = 0.0015$, $n_I = 3$	medium
NBE1b	Feature Type I, $N_B = 80$, $\Delta_R = 0.0008$, $n_I = 3$	weak-medium
NBE2a	Feature Type II, $N_B = 80$, $\Delta_R = 0.0001$, $n_I = 1$	weak
NBE2b	Feature Type II, $N_B = 20$, $\Delta_R = 0.0004$, $n_I = 1$	medium
VFA1	600 clusters, $D_{\mathrm{MAX}} = 1.8$	strong
VFA2	960 clusters, $D_{\mathrm{MAX}} = 1.8$	medium
VFA3	1320 clusters, $D_{\mathrm{MAX}} = 1.8$	weak
VFA4	200 clusters, $D_{\mathrm{MAX}} = 0.6$	medium

TABLE III
EXPERIMENT II: WATERMARKING PARAMETERS VALUES.

V. DATA ANALYSIS

During the experiments, if a subject notices surface defects, he/she is supposed to enter a value proportional to the amount of distortions perceived on the model surface. In the following we refer to these values as *subjective scores*. The subjective scores have to be condensed by statistical techniques used in standard methods [41], [49] to yield results which summarize the performance of the system under test. The averaged score values, called *Mean Opinion Score* or *MOS*, are considered as the amount of distortions that anyone can perceive on a particular watermarked 3D object. However, impairment is measured according to some scale, and such scale may vary from person to person. In this section, we report the methods used for matching the scales of our test subjects. Then, we describe the methods used

to combine the subjective data and evaluate the precision of the estimates. The subjects are screened and outliers are discarded. Finally, the results are checked for error due to the methodology of the experiment.

A. Normalization

As a measurement device, a test subject may be susceptible to both systematic and random errors. The purpose of normalizing procedure is to compensate for any systematic errors. The procedure must be applied to the measurement gathered from each test subject prior to the combination of measurements across subjects. The unscaled annoyance value, m_{ij} obtained from subject i after viewing the test object j can be represented by the following model:

$$m_{ij} = g_i a_j + b_i + n_{ij} \tag{7}$$

where let a_j be the true annoyance value for test object j in the absence of any error, g_j is a gain factor, b_i is an offset, and n_{ij} is generally assumed to be a sample from a zero-mean, white Gaussian noise.

In this model, the gain and offset could vary from subject to subject. If the variations are large across the subjects or the number of subjects is small, a normalizing procedure can be used to reduce the gain and the offset variations among test subjects. In order to check if the offset and gain factors vary significantly from subject to subject, a two-way analysis of variance (ANOVA) approach was used [51]. A two-way ANOVA divides the total variations of a table of data into two parts: the variation that is attributed to the columns of the data table, and the variation that is attributed to the rows of the data table. Specifically, the F-test can be used to determine the likelihood that the means of the columns, or the means of the rows, are different. The experimental data, m_{ij} is arranged so that each row represents data for one test object and each column represents all the data for one test subject. The analysis assumes that the data can be modeled as [51]:

$$m_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{8}$$

where μ is the overall mean, α_i stands for the subject effects, β_j stands for the model effect, and ϵ_{ij} are the experiment errors. In Table IV and Table V the ANOVA results are indicated for objects and subjects from Experiment I and Experiment II respectively. The F-values for both subjects (F = 12.22 and F = 18.3 for Experiment I and Experiment II respectively) and objects (F = 23.01 and F = 21.16) were large. The right-most column of the table contains the probabilities that the subject effect and the object effect are constant and that is no differences among the subjects or among the test objects. It is important to underline that the object effect is not expected to be constant since it depends on the particular 3D object and the 3D objects were deliberately varied.

Source	*SoS*	*df*	*MS*	*F*	*p*
Subjects	372.22	10	37.24	12.22	0
Models	2735.54	39	70.14	23.01	0
Error	1188.83	390	3.04		
Total	4296.82	439			

TABLE IV
EXPERIMENT I: ANOVA ANALYSIS RESULTS.

Source	*SoS*	*df*	*MS*	*F*	*p*
Subjects	382.84	10	38.28	18.3	0
Models	2080.63	47	44.26	21.16	0
Error	983.34	470	2.09		
Total	3446.82	527			

TABLE V
EXPERIMENT II: ANOVA ANALYSIS RESULTS.

Source	*SoS*	*df*	*MS*	*F*	*p*
Subjects	21.79	10	2.18	8.92	0
Sequences	151.454	39	3.88	15.89	0
Error	95.33	390	0.24		
Total	268.58	439			

TABLE VI
EXPERIMENT I: ANOVA ANALYSIS RESULTS FOR LN($m_{ij} + 1$).

Source	*SoS*	*df*	*MS*	*F*	*p*
Subjects	21.18	10	2.11	14.54	0
Sequences	132.57	47	2.82	19.35	0
Error	68.50	470	0.14		
Total	222.26	527			

TABLE VII
EXPERIMENT II: ANOVA ANALYSIS RESULTS FOR LN($m_{ij}+1$).

For the F-values in Tables IV and V, the probabilities were zero. This means that there is a significant variation in the subjective value means from subject to subject. To check if the variation is caused by variations in the gain factor g_i, an ANOVA is also done for the natural logarithm of m_{ij}. In fact, by taking the logarithm of the Equation 7 we obtain:

$$\ln(m_{ij}) = \ln(g_i a_j + b_i + n_{ij}) \approx \ln(g_i) + \ln(a_j) + b_i/g_i a_j + n_{ij}/g_i a_j \tag{9}$$

In this equation $\alpha_i = \ln(g_i)$, $\mu + \beta_j = \ln(a_j)$ and $\epsilon_{ij} \approx b_i/g_i a_j + n_{ij}/g_i a_j$ (see equation 8). ϵ_{ij} is no longer independent of the other factors, however if b_i and n_{ij} are small, this will not matter much in the analysis. Table VI contains the ANOVA results for $\ln(m_{ij}+1)$. Here, we decide to use the $\ln(m_{ij}+1)$ instead of the $\ln(m_{ij})$ to avoid numerical problems to the presence of zero scores. The F-values are large and the probabilities near zero. This means that there were significant subject-to-subject variations in the gain factors and then some form of subject-to-subject correction was required. Several methods for estimating the offsets and the gain are possible. The probability of the null hypothesis (no variation in the subject means) for each correction method and the ANOVA results are summarized in Table VIII. The measurements are adjusted prior to combination in the following way:

$$\hat{m_{ij}} = \frac{1}{\hat{g}_i}(m_{ij} - \hat{b}_i) \tag{10}$$

where $\hat{g}_i$ is the corrected gain, $\hat{b}_i$ is the corrected offset and $\hat{m_{ij}}$ is the normalized score.

In the first correction method the offsets are estimated using the mean of all measurements made by each subject:

$$\hat{b}_i = \frac{1}{J}\sum_{j=1}^{J} m_{ij} - \mu \tag{11}$$

Since the mean is not a robust estimator of the center of a distribution the median is also tried as an estimate of the offset.

$$\hat{b}_i = \text{median}\,\{m_{ij}\ ,\ \forall j \in J\} - \mu \tag{12}$$

where J is the set of test models. The results are shown in Table VIII. The mean estimate remove the subject to subject variations for both Experiments.

To adjust also the gain, two gain estimation methods has been considered. The first gain estimation evaluates the measurements in terms of the experiment instructions and corrected the gain if the instructions were not followed exactly. In fact, the test subjects are told to assign a value of 10 to the worst of the test models seen during the training session. The corrected gain is set to make this true:

$$\hat{g_i} = \frac{1}{K}\max_{j\in J}(m_{ij}) \tag{13}$$

where K is equal to the upper end of the scale (10) in our testing procedure. The second method for correcting the gain variation relies on a statistical estimate of each test subject's range. The standard deviation of the values is used to estimate the range. In this case, the gain factor becomes:

$$\hat{g_i} = \frac{4\delta_i}{N} \tag{14}$$

where δ_i is the standard deviation of all the values recorded by the i-th subject. The results are summarized in Table VIII. The gain correction (14) combined to the mean offset correction provides the best results. Hence, after the normalization the collected data depend on the model but do not depend on the subject. This fact indicates that *the experiment is well-designed*, i.e. the experimental methodology is not affected by any systematic error.

	Experiment I		Experiment II	
Correction	**Offset**	**Gain**	**Offset**	**Gain**
None	0	0	0	0
Mean	1	0.5557	1	0.6903
Median	0.0006	0.0001	0.0006	0
Mean + max	0.998	0.7962	1	0.8286
Mean + std	1	0.9996	1	0.999

TABLE VIII
ANOVA analysis results after normalization (F-Test values for subject Offset and Gain dependency). Probabilities near one means that there is no difference across subjects.

B. Overall Scores and Screening

After the normalization process, the subjective scores values are combined into a single overall score for each test model using the sample mean. The subjective Mean Opinion Score (MOS) is given by:

$$\mu_j = \frac{1}{N_j} \sum_{i=1}^{N} m_{ij} \tag{15}$$

where N_j is the number of test subjects in the experiment. This measure represents the subjective scores of the non-detected and detected defects. A score of zero is assigned to the non-detected defects. Confidence intervals for the mean subjective scores were calculating using Student's T-distribution. The T-distribution is appropriate when only a small number of sample is available [51]. The sample standard deviation s_j was calculated for each sequence j using μ_j and the limits were calculated by:

$$l_j = t_{(0.05,N_j)} s_j \tag{16}$$

where $t_{(0.05,N_j)}$ is the t-value associated with a probability of 0.95 and N_j degrees of freedom. As the number of observation N_j increases, the confidence interval decreases. The final result is j Mean Opinion Score values with associated 95% confidence interval, $\mu_j \pm l_j$. The MOS values and sample standard deviations were first used in the *subject screening procedure*. In the screening procedure used (Annex 2 of ITU BT.500 Recommendation [41]) an expected range of values is calculated for each model. A subject is not rejected for always being above the expected range or always being below the expected range but for being erratic on both sides of the expected range. This procedure is appropriate to reduce the variability of the data in small sample sets. If necessary, after the screening procedures, the MOS and their relative confidence intervals were recalculated without the data from rejected test subject. After the normalization and screening, for the Experiment I the data of one subject were discarded, and for the Experiment II the data of three subjects were rejected.

C. Data Evaluation

The subjective data after normalization, averaging and screening procedures will be used to test the proposed objective model of the watermarked 3D object defects. In this section we check the validity of the obtained data and evaluate if the methodology can be improved. Figures 6 and 7 shows the overall data spread for Experiment I and II. Note that the data spread was good for all these experiments. For Experiment I the MOS values ranged from 0.36 to 10.0 before the normalization and screening and ranged from 0.45 to 8.76 after (figure 6 (a)). For Experiment II the ranges before and after are (0.45 , 9.17) and (0.64 , 9.44) respectively (figure 7 (a)). The large number of data points (test 3D models) compensated for the lack of precision of individual points.

In summary, the experiments provided good data for most of the test models. The confidence intervals were reduced after the normalization and screening procedure (figures 6 (b) and 7 (b)). The confidence intervals were large for the test models that were hard to examine. This situation could be improved in future experiments by ensuring that the weakly watermarked models are closer to perceptual threshold or using many more test subjects. In conclusion, the data were good for the overall fits.

VI. Proposed Perceptual Metric

Perceptual metrics that compute predictions of human visual task performance from input images, are usually based on a vision model. For any application in which a vision model produces reliable performance predictions, its use is almost always preferable to psychophysical data collection. One reason for this preference is that running a model generally costs much less than running a psychophysical experiment to generate the same system performance information, especially when the system evaluation in question is still in design phase as in our case. There are two approaches [52] to model psychophysical quantities: *performance modeling* and *mechanistic modeling*. Although the

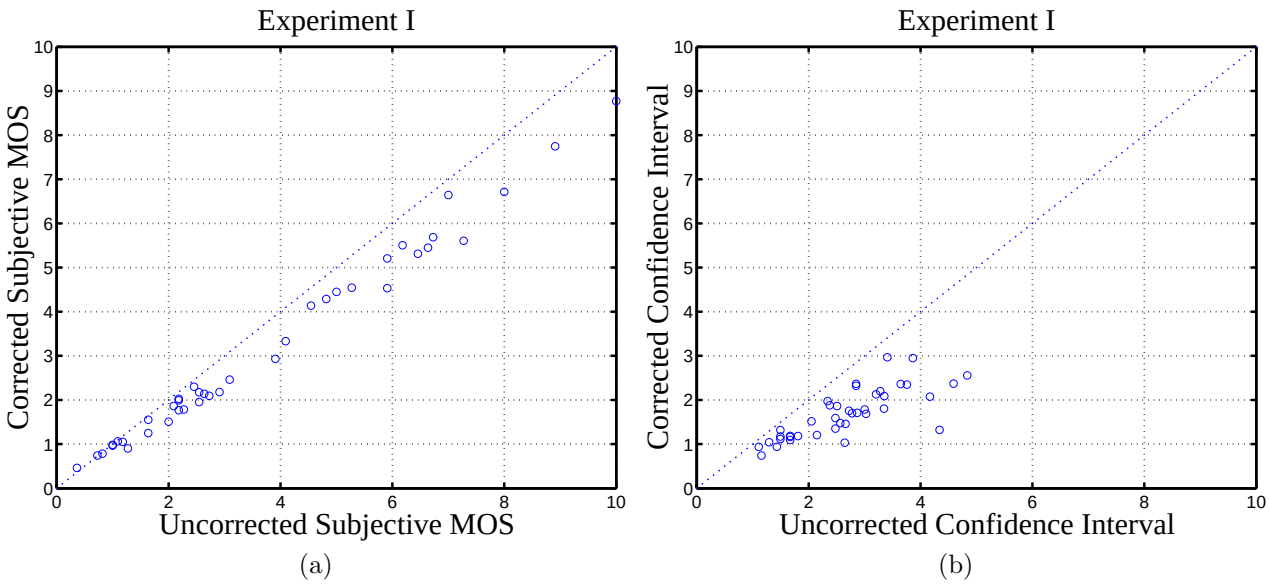


Fig. 6. Subject data correction results for Experiment I: (a) the mean scores values and (b) the confidence intervals

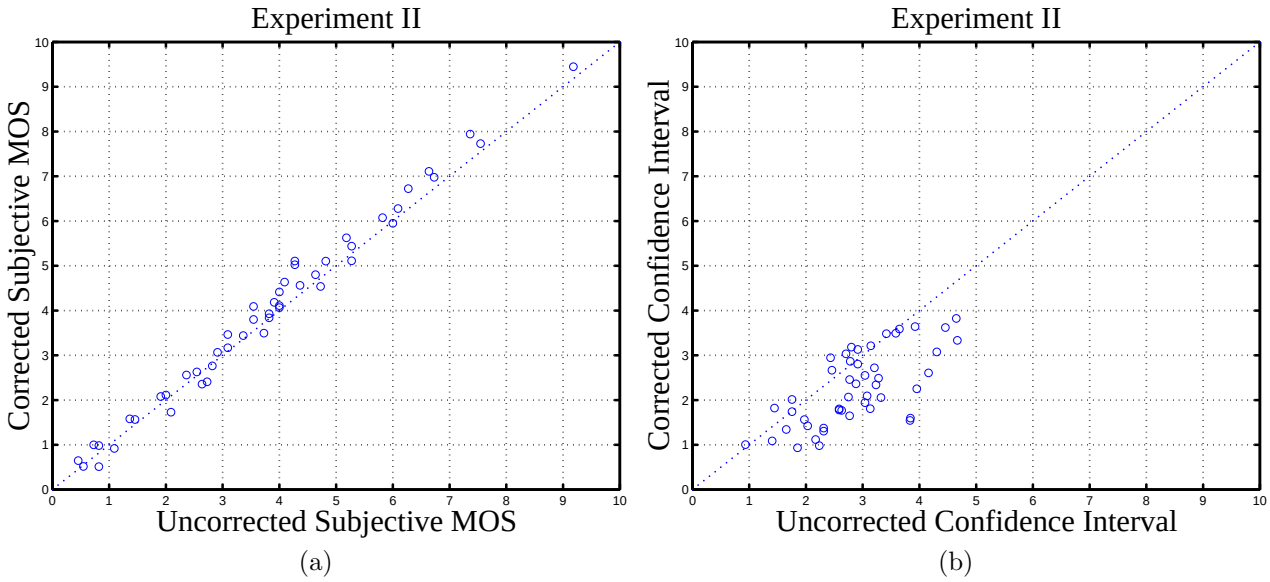


Fig. 7. Subject data correction results for Experiment II: (a) the mean scores values and (b) the confidence intervals

distinction is more a continuum that a strict dichotomy, the performance models tends to treat the entire visual system as a "black box" for which input/output functions need to be specified. For the mechanistic model, physiological and psychophysical data are used to open the black box. As a result, input/output functions are needed not only for the system as a whole but for a number of component mechanisms within. These components of the model have the same functional response as physiological mechanisms of different stages of the HVS. It is important to underline that, at this time, it does not appear feasible to build a sophisticated model of visual perception of geometric artifacts since such models could become too complex to be handle in practice. In fact, one essential feature of an interactive application is that 3D objects are observed in motion, so the classical visual models used for still images should be integrated with other perceptual models that take into account the behavior of the human perceptions in dynamic scene. Additionally, this model should take into account a lot of parameters that depend on the rendering, such as the lighting model, the texturing, and so on. For all of these reasons we opt for the "black box" approach. In particular, we use an objective measure based on surface roughness estimation combined with a standard psychometric function to model the black box. Typical a psychometric curve is a logistic, a cumulative gaussian or a Weibull function. The only function of the psychometric curve is to transform the defect strengths into subjective scores values and vice versa. In particular, we

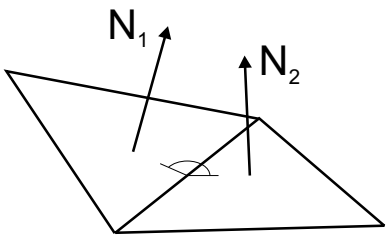


Fig. 8. Dihedral angle.

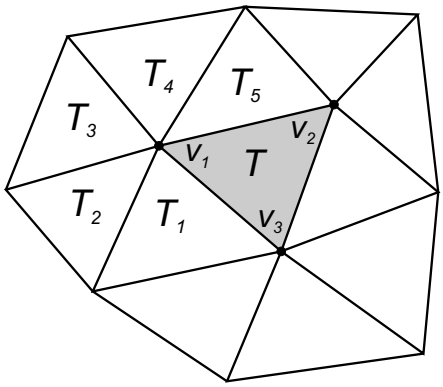


Fig. 9. $G(.)$ and $V(.)$ descriptions.

use a *gaussian psychometric function*:

$$g(a,b,t) = \frac{1}{2\pi}\int_{a+bx}^{\infty} e^{-\frac{t^2}{2}}dt \tag{17}$$

where a and b are the parameters to be estimate by fitting the objective metrics values versus the subjective data. To estimate such parameters we use a nonlinear least-squares data fitting by the Gauss-Newton method.

The intuition and the questions in Experiment I and II confirm us that the watermarking artifacts that producing different kind of noise on the surfaces can be described essentially with *roughness*. In fact, in Experiment II defects were described the defects in the same way that in the Experiment I, i.e. *"... In few words the roughness of the surface of the model is increased in some way."*. So, the objective metric which we choose to measure the strength of the defects is based on a measure of the surface roughness.

A. Roughness Estimation

Thanks to previous studies about 3D watermarking, we have realized that a good measure of the visual artifacts produced by watermarking should be based on the amount of roughness introduced on the surface [53]. As just said, the interview phase of these two experiments confirm us that "roughness" is a good term to describe, in a general way, the defects introduced over the surface of the model. Hence, two objective metrics to estimate the roughness of the surface have been developed. In the following we give a detailed description of those metrics.

1) Multi-scale Roughness Estimation: The first roughness measure we propose is a variant of the method of Wu et al. [54]. This metric measures the per-face roughness by making statistical considerations about the dihedral angles associated to each face. Wu et al. developed this measure in order to preserve significative shape features in mesh simplification algorithm.

The *dihedral angle* is the angle between two planes. For a polygonal mesh, the dihedral angle is the angle between the normals of two adjacent faces (figure 8). The basic idea of this method is that the dihedral angle is related to the surface roughness. In fact, the faces normals of a smoothed surface varying slowly over the surface, consequently the dihedral angles between each adjacent face are close to zero. To be more specific Wu et al. associate to each dihedral angle an amount of roughness given by the quantity $1-(N_1 \cdot N_2)$. Given a triangle T with vertices v_1, v_2 and v_3, its roughness is computed as:

$$\mathcal{R}_1(T) = \frac{G(v_1)V(v_1) + G(v_2)V(v_2) + G(v_3)V(v_3)}{V(v_1) + V(v_2) + V(v_3)} \tag{18}$$

Referring to figure 9, $G(v_1)$ is the average of the roughness associated the dihedral angles $T - T_1$, $T_1 - T_2$, $T_2 - T_3$, $T_3 - T_4$, $T_4 - T_5$ and $T_5 - T$. In the same way $G(v_2)$ and $G(v_3)$ are the mean roughness associated to the dihedral angles of the faces adjacent to the vertices v_2 and v_3. Instead, $V(v_1)$, $V(v_2)$ and $V(v_3)$ are the variance of the roughness associated to the dihedral angles of the faces adjacent to the vertex v_1, v_2 and v_3.

A roughly surface can be considered as a surface with a high concentrations of bumps of different size over it. This metric is able to measure 'bumpiness' of the surfaces at face level, but, if the granularity of the surface roughness, i.e. the size of the bumps, is higher than the medium dimension of one face this metric fails to measure them correctly. In other words this measure does not take into account the *scale* of the roughness. For this reason we modified it in order to account for different bumps scale. The first step to achieve this goal is to transform this per-face roughness estimation in a per-vertex roughness estimation in the following way:

$$\mathcal{R}_1^N(v) = \frac{1}{|S_T^N|}\sum_{i \in S_T^N} \mathcal{R}_1(T_i)\mathcal{A}_{T_i} \tag{19}$$

where S_T^N is the set of the faces of the *N-ring*[2] of the vertex v, $|.|$ is the usual cardinality operator and $\mathcal{A}_{T_i}$ is the area of the face T_i. The fact to consider the N-ring in the roughness evaluation accounts for different scale of bumpiness.

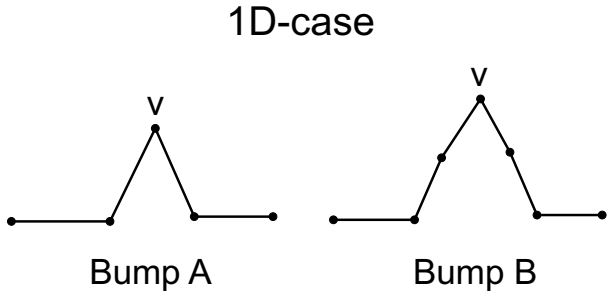


Fig. 10. Bumps with different scale.

Referring to figure 10; the bump of size equivalent to the 1-ring (A) is well measured by $\mathcal{R}_1^1(v)$, a correct value of roughness for the vertex v in the case (B) is provided by $\mathcal{R}_1^2(v)$. Approximatively, we can state that the roughness of a vertex v centered on a bump of area close to the area of the faces that form the N-ring is well measured by $\mathcal{R}_1^N(v)$. This approximation could not be valid in certain cases, for example for a surface that presents high curvature of for high value of N. Hence, a real multi-scale measure of bumpiness would require further developments but we assume this approximation sufficient. In order to obtain for each vertex a single value of roughness that depend on the roughness evaluated at several scales we take the maximum value produced by N-ring of different size. In particular, in our objective metric we have chosen 3 scales of roughness:

$$\mathcal{R}_1(v) = \max\{\mathcal{R}_1^1(v), \mathcal{R}_1^2(v), \mathcal{R}_1^4(v)\} \tag{20}$$

The total roughness of the 3D object is the sum of the roughness of all vertices:

$$\mathcal{R}_1(M) = \sum_{i=1}^{N_v} \mathcal{R}_1(v_i) \tag{21}$$

where N_v is the total number of mesh vertices. In the following we will see how to transform this multi-scale roughness estimation in an objective metric that correlates well to the human perception of geometric defects.

2) Smoothing-based Roughness Estimation: The second method we develop to measure surface roughness is based on considerations arise during the subjective tests. Since most of the subjects have said, during the interview, that the defects are perceived better on smooth surface we decide to develop a smoothing-based roughness estimation. The basic idea of this approach is to apply to the model a smoothing algorithm and then to measure the roughness of the surface as the variance of the differences between the smoothed version of the model and the original one. A sketch of the smoothing-based roughness is depicted in Figure 11.

The first step of this approach is to build a smoothed version of the model (M^S) by applying a *smoothing* algorithm to the input model (M). Several possibilities for smoothing exist [55], [56], [57], [58]. Here, we decide to use the *Taubin filter* [55] for its simplicity of implementation. The parameters of the Taubin filter used are $\lambda = 0.6307$, $\mu = -0.6352$, number of iterations $= 5$. Such parameters provide a medium smoothing effect. When the smoothed model is obtained, the distance between each vertex of M and M^S is computed in the following way:

$$d_{OS}(v, v^S) = \mathtt{proj}_{n_v^S}(v - v^S) \tag{22}$$

where $\mathtt{proj}(.)$ indicates the projection of the vector $(v - v^S)$ on the vertex normals of the smoothed surface (n_v^S). At this point the per-vertex roughness is computed by evaluating the local variance of the distances $d_{OS}(.)$ around each vertex. To be more specific, for each vertex v, the set of distances associated to its *2-ring* ($S_d^2(v)$) is build and the variance of this set evaluated. Then, the per-vertex smooth-based roughness is computed by:

$$\mathcal{R}_2(v) = \frac{V(S_d^2(v))}{\mathcal{A}_{S^2}} \tag{23}$$

where A_{S^2} is the area of the faces that form the 2-ring of v. This area is used at the denominator since surface with the same local variance of the distances but smaller area are assumed more roughly. The roughness of the input model

[2]The N-ring neighborhood vertices of a vertex v is an extension of the 1-ring neighborhood. A 2-ring neighborhood is created from the 1-ring by adding all of the vertices of any face containing at least one vertex of the 1-ring. Additional rings can be added in the same way to form the 3-ring, the 4-ring and so on.

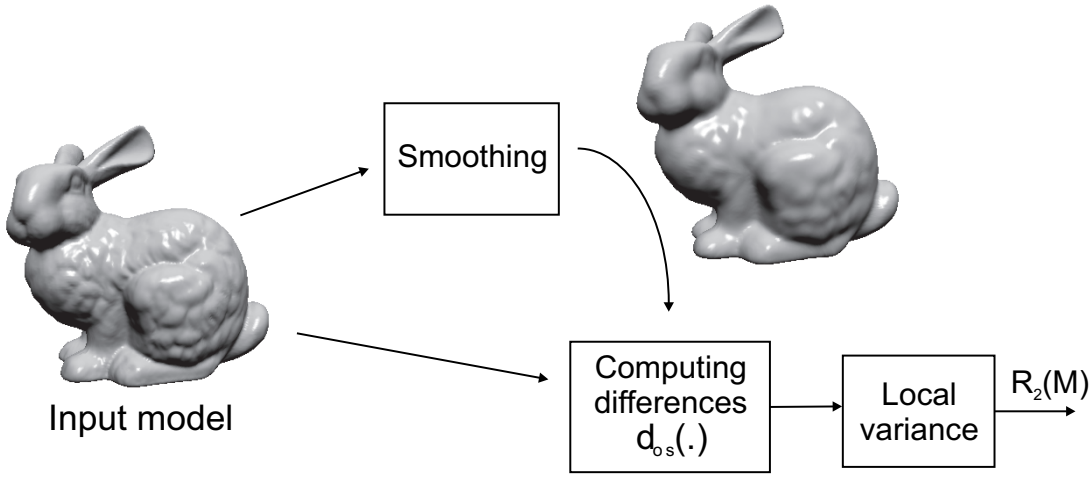


Fig. 11. Smoothing-based Roughness Estimation.

is the sum of the roughness of all vertices of the model:

$$\mathcal{R}_2(M) = \sum_{i=1}^{N_v} \mathcal{R}_2(v_i) \tag{24}$$

where N_v is the number of vertices of the model.

3) Objective Metric: Now, we describe how to use the roughness estimation to predict the visual distortions produced by a certain 3D watermarking algorithm. On the basis of several evaluations we decide to define our objective metric as the increment of roughness between the original and the watermarked model. This increment is normalized with respect to the roughness of the original model. In formula:

$$\mathcal{R}(M, M^w) = \log\left(\frac{\mathcal{R}(M) - \mathcal{R}(M^w)}{\mathcal{R}(M)} + k\right) - \log(k) \tag{25}$$

where $\mathcal{R}(M)$ is the total roughness of the original model and $\mathcal{R}(M^w)$ is the total roughness of the watermarked model. Both $\mathcal{R}_1(.)$ and $\mathcal{R}_2(.)$ can be used to obtain two different objective metrics for 3D watermarking. The logarithm is employed to better discriminate low value of relative roughness increments. The constant k is used to avoid numerical instability of the (25) since the logarithm tends to $-\infty$ for M^w very close to M. In particular the value of k has been adjusted to normalize the values provided by the metric between 0 and 10, that is the same range of values used by the subjects during the experiments. In the following, we indicate with $\mathcal{R}_1(M, M^w)$ the objective metric based on the multi-scale roughness and with $\mathcal{R}_2(M, M^w)$ the objective metric based on the smoothing-based roughness estimation.

VII. Experimental Results

In this section we analyze the performances of the two proposed objective metrics and we compare them with geometric metrics usually adopted in literature for model quality evaluation. First, the correlation between the subjective Mean Opinion Score (MOS) collected in Experiment I and the distances given by two geometry metrics based on Hausdorff distance for model similarity is evaluated. In this way we obtain a term of comparison for the evaluation of our metrics. Then, the objective metrics are fitted with a gaussian psychometric curve to match the subjective data collected in the first experiment. The performances of the *perceptual metrics* so obtained, are evaluated using the subjective MOS provided by the Experiment II. In other words the subjective data of the Experiment II are used to *validate* the developed perceptual metrics. The results obtained will be discussed at the end of the Section.

A. Hausdorff distances

As previously stated (Section II) two of the common geometry metrics used to measure the similarity between two 3D objects are the Maximum (3) and the Mean geometric error (5). These two metrics are based on the Hausdorff distance between models' surface. Here, we want to evaluate if the distance between the original and the watermarked model could be a reliable metric for perceptual watermarking impairments prediction. To do this, the Hausdorff distances of each watermarked models from the original are plotted versus the MOS provided by the Experiment I. At this point, the linear correlation coefficient of Pearson (r_P) and the non-linear (rank) correlation coefficient of Spearman (r_S) are calculated in order to evaluate the global performances of the metric obtained by fitting these geometric data with

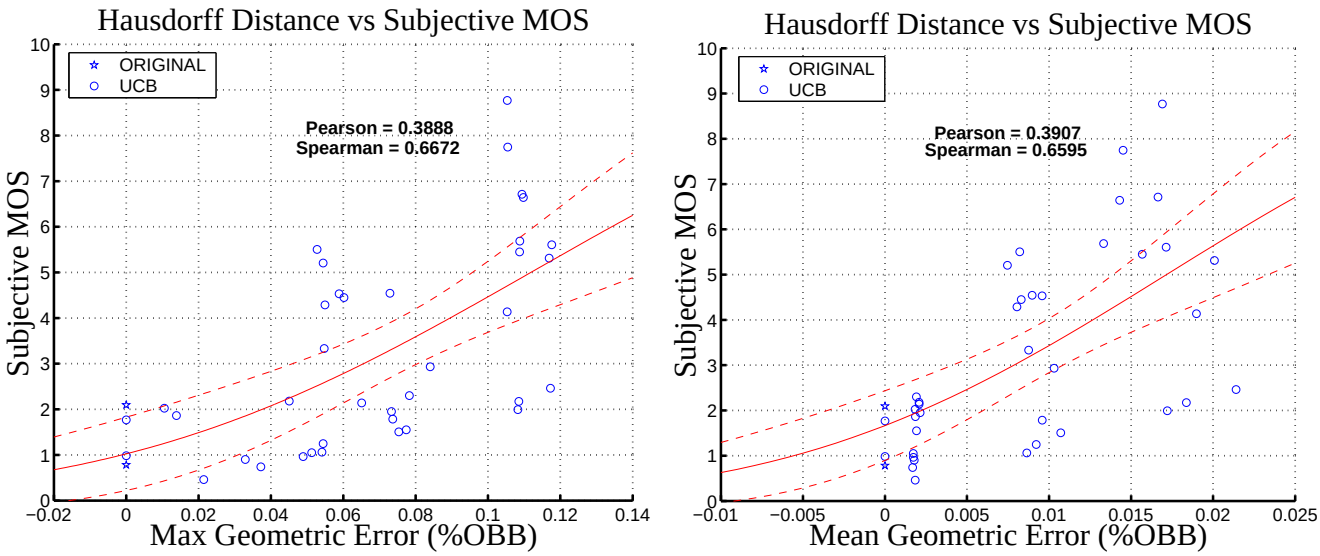


Fig. 12. Geometric Hausdorff distance vs Subjective MOS.

a cumulative gaussian (17). Even if a psychometric curve is used to fit the geometric measures the results do not correlate well with subjective MOS. This underline the fact that $d_\infty(.)$ and $d_1(.)$ are not designed on the basis of how humans perceive geometric defects. The results are summarized in Figure (12). Such results will be used as a reference to compare the performances of the perceptual metrics based on roughness estimation.

B. Roughness-based metrics results for Experiment I

As we stated in Section V the goal of the first experiment was to make an initial study on the perception of the geometric defects caused by watermarking algorithms. The experimental data confirm that the subjective perception of the impairments is well-described by measure of roughness. The subjective data of this experiment are used to obtain two perceptual metrics ($\mathcal{R}_1^*(M, M^w)$ and $\mathcal{R}_2^*(M, M^w)$) from the two proposed objective metrics ($\mathcal{R}_1(M, M^w)$ and $\mathcal{R}_2(M, M^w)$) by fitting them with a gaussian psychometric curve (17). In this way two kind of perceptual metrics are obtained, one based on multi-scale roughness measure and another one based on smoothing-based roughness estimation. The parameters of the gaussian psychometric curve after the fitting are $(a = 1.9428, b = -0.2571)$ for $\mathcal{R}_1(M, M^w)$ and $(a = 2.0636, b = -0.2981)$ for $\mathcal{R}_2(M, M^w)$. The smoothing-based one provides a better fit ($r_P = 0.6730$, $r_S = 0.8680$) than the multi-scale one ($r_P = 0.8286$, $r_S = 0.8919$) as depicted in Figure 13. The 95% confidence intervals of the subjective MOS versus the roughness metric are depicted in Figure 13 (Top). Few confidence intervals are large, approximatively the 20% of the maximum range scale. Note that the width of the intervals would have been reduced if the experiment had been carried out with more subjects. On the right top of the graphs it is possible to notice some points outside ones of the fitting curve. Most of these outliers correspond to the Venus model. This is due to the fact that the Venus model represents a human face. Human face images are well-known in subjective experiments as an high level factor attracting human attention, i.e. people are more able to deal with human face, so the distortions on the Venus head result more visible and annoying with respect to the other models.

C. Objective Metrics Performances

As discussed in previous section, the two proposed objective metrics have been transformed into two corresponding perceptual metrics using the data from the Experiment I. In order to evaluate such metrics Experiment II was carried out with three other watermarking algorithms: KDK, NBE and VFA. The validation is very simple: the perceptual metrics obtained in Experiment I are used to predict the MOS obtained in the second experiment and their correlation coefficients are computed. The correlation coefficients r_P and r_S are reported in Table IX. The rows indicate the watermarking algorithm groups. The first two columns of this table report the Spearman correlation coefficient of the Maximum and Mean geometric error for comparison. The third and the fourth column shown the values of r_P and r_S for $\mathcal{R}_1^*(M, M^w)$, while the last two columns are the r_P and r_S values for $\mathcal{R}_2^*(M, M^w)$. Referring to this table we can make the following important considerations:

- Overall, both geometric metrics based on the Hausdorff distance do not correlate well with the subjective data. On the other hand the developed metrics exhibit strong correlation with subjective data, in particular concerning the Spearman's coefficient.

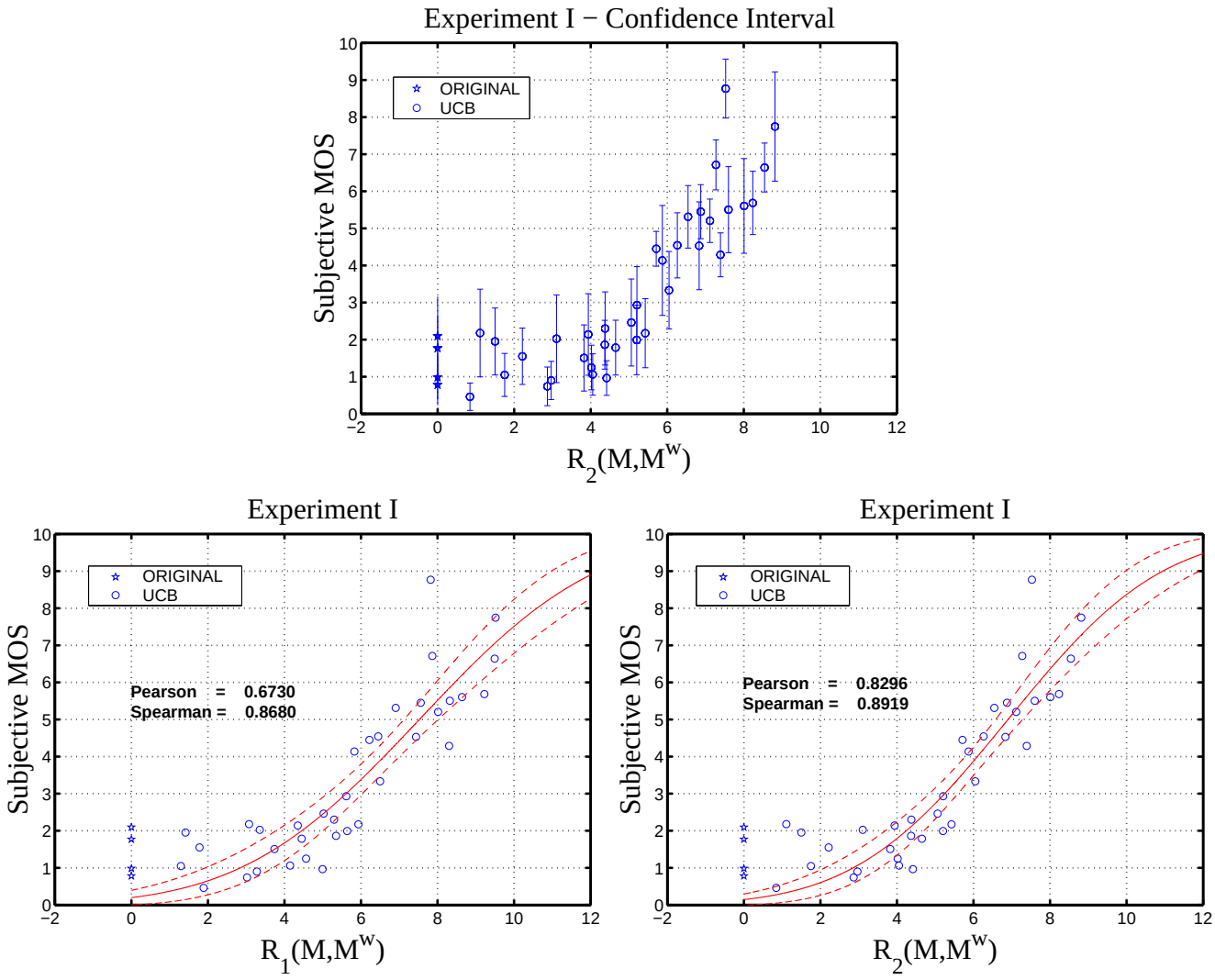


Fig. 13. Experiment I: Subjective MOS versus objective metrics curves fits.

- The Spearman's coefficients for the NBE and VFA algorithms (second and third rows respectively) demonstrate that both metrics are able to predict impairments introduced by these two algorithms.
- The worst performances of the proposed metrics are obtained for the KDK algorithm. This can be explained by considering that the distortion produced by the KDK algorithm on the surface are *non-uniform*.
- The overall performances of the perceptual metrics for the watermarking algorithms that introduce *uniform* distortions on the surface shown are reported in the 6th row of the table. The values of the correlation coefficients ($r_P = 0.6455$ and $r_S = 0.8416$ for the first metric, $r_P = 0.7383$ and $r_S = 0.8954$ for the second metric) are very high thus resulting in very good prediction of quality evaluation of the developed metrics.
- The overall performances of the perceptual metrics considering all the *uniform* and *non-uniform* watermarking algorithms tested are reported in the last row of the table. Despite the presence of the KDK algorithm, for which the performance are not high, the global prediction of the metrics still remains good. In particular, such performances are excellent comparing with the ones of the two geometry-based metrics.

In order to visualize the results of the table IX the graphs of Figure 14 shown the values of the objective metrics plotted versus the subjective MOS for several watermarking algorithm groups. The curve drawn on this figure does not represent the result of a fit; the same gaussian curve obtained with the data of the Experiment I are drawn for all the picture. In other words these graphs visualize the behavior of the KDK, the NBE and the VFA algorithm with respect to the perceptual metrics developed after Experiment I, that is represented by the red curve (the dashed line is the confidence interval for that curve).

Since the non-linear correlation coefficient of Spearman is based on the rank of the data instead on the data values itself like the Pearson's coefficients, it is interesting to compare the watermarked models ranking by the impairments perceived by the subjects and by the impairments predicted by the metrics. An example of this comparison is reported

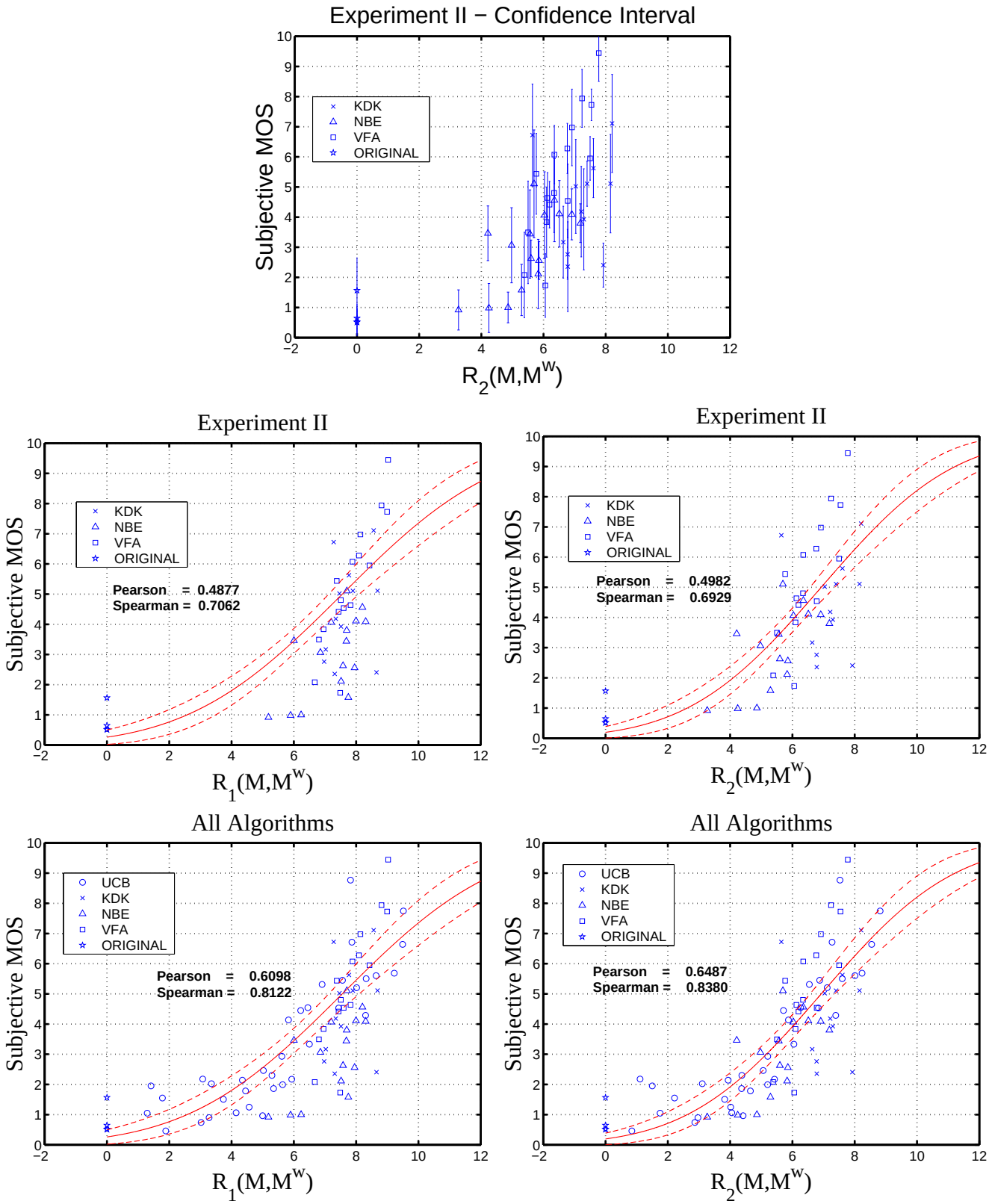


Fig. 14. Experiment II: Subjective MOS vs objective metric curves. The parameters of the fitting curve are the same of the Experiment I.

	Hausdorff Distance		$\mathcal{R}_1^*(M, M^{\mathbf{w}})$		$\mathcal{R}_2^*(M, M^{\mathbf{w}})$	
Algorithms	**Max** (r_S)	**Mean** (r_S)	r_P	r_S	r_P	r_S
UCB	0.6672	0.6595	0.6730	0.8680	0.8296	0.8919
KDK	0.6904	0.3230	0.6154	0.7171	0.5514	0.7111
NBE	0.7087	0.7026	0.5597	0.7917	0.6240	0.8146
VFA	0.4951	0.8815	0.7472	0.9389	0.7763	0.9147
KDK+NBE+VFA	0.3759	0.4853	0.4877	0.7062	0.4982	0.6929
UCB+NBE+VFA	**0.5219**	**0.6183**	**0.6455**	**0.8416**	**0.7383**	**0.8954**
ALL	**0.4993**	**0.5352**	**0.6098**	**0.8122**	**0.6487**	**0.8380**

TABLE IX
PERCEPTUAL METRICS PERFORMANCES.

in Figure 15 where the Bunny and the Feline models are considered. It is possible to see that the smoothing-based perceptual metric, that has values of r_S slightly high with respect to the multi-scale metric, is able to rank the watermarked models in a way very close to the subjective rank.

VIII. Conclusions

In this work first studies about the extension of the idea of perceptual image watermarking to 3D watermarking have been conducted. In particular, a new experimental methodology for subjective quality assessment of watermarked 3D objects has been proposed. The analysis of the data collected by two subjective experiments that use this methodology demonstrate that such methodology is well-designed and provides reliable subjective data. Additionally, two perceptual metrics for 3D watermarking impairment prediction have been developed by combining roughness estimation with subjective data. The very good performance of such metrics have been deeply analyzed. The results of such analysis demonstrate the effectiveness of the proposed perceptual metrics with respect to the state of the art geometric metrics. Such metrics could be used in a feedback mechanism to improve the visual quality of 3D watermarking algorithms.

IX. Acknowledgments

The *"Bunny"*, *"Horse"*, *"Feline"* and *"Venus"* model used in the experiments are made public available by the Multiresolution Modeling Group of Caltech University [50].

Appendix

Experiment Script

Get the subject into position, centered in front of the screen and at the correct distance (about 0.4 cm). Then read the following instructions:

- "This test concerns the evaluation of the **distortions or impariments** introduced by watermarking algorithms on the surfaces of **3D models**.
  What is a 3D model? A 3D model is a collection of data that represent a 3D shape. 3D models are used in particular in entertainment industries, for examples movies and video games use a lot of 3D models.
  What is watermarking? Digital watermarking is a technology used to embed information inside a digital media. Imagine you want to associate some information with a digital media, i.e. the name of the owner of an image to the image itself. A watermarking algorithm, specific for images, can process the image and embed this information inside the image data itself. So, this information can be eliminated only modifying the watermarked image. In order to embed the data watermarking algorithms modify some properties of the digital media producing always some distortions in the watermarked media. The purpose of the test is the evaluation of the distortions introduced by watermarking algorithms on 3D models. Usually these distortions are visible on the model surfaces. So, the test is very simple: you interact with some models and you have to indicate if you see or not a certain kind of distortions. Obviously I will show to you these distortions so you can understand what you have to evaluate.
- During the test you interact with the models and you have to evaluate these impairments. In particular you will indicate whether you detect any distortion or impariment.
- For those 3D models you detect a distortion you will indicate *how much* you perceive such distortion by entering a number that is proportional to its distortion value.
- Additionally you will have to indicate the part of the models *where* the distortions are more evident (this task is present only for Experiment I).
- Here I will show you the models without any distortion. The test includes four models. You have to imagine these models like *statues*. In particular I will show to you a model called "Venus" that represent the head of a statue of Venus, a mythological feline with wings called "Feline", a "Horse" and, finally a "Bunny" .

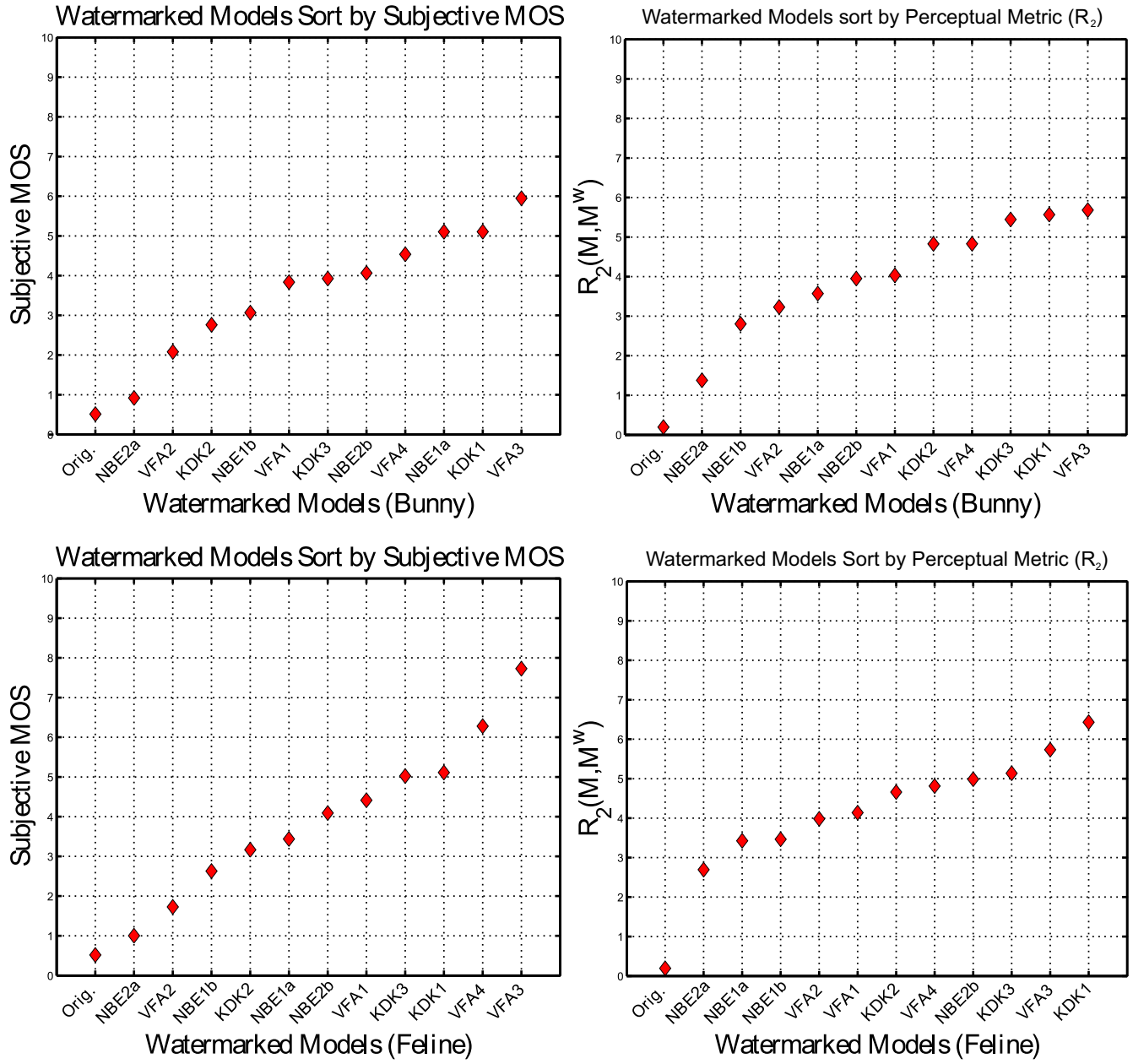


Fig. 15. Experiment II: Comparison between models' impairment ranking. On the left the models are ranking by subjective MOS. On the right the models are ranking by smoothing-based roughness.

- [Show originals]
- Are you able to remember these models? Do you want to see it again?
- Before we start the experiment, you will see how the typical distortions introduced by watermarking process look like. In few words the roughness of the surface of the model is increased in some way. We have to recognize this roughness, so it is important that you remember, for each model, the roughness of its parts. Another important aspect that you have to consider during your evaluation is that the distortions introduced by the algorithm is uniform on the surface.
- [Show the watermarked models]
- Have you understood how these distortions look like? Do you have any questions?
- In this phase, you will learn how to interact with the 3D model. You interact with the models by using the mouse. To rotate the model push the left button and moves the mouse. When you want to stop to rotate the model release the left button. To zoom the model push the right button and move the mouse ahead or back to zoom in or to zoom out. When you want to stop to zoom the model release the mouse button.
- [Interaction trial - Try to rotate the model. Try to zoom it. You can move the model left/right/up and down with

the arrow keys. Try.].

- (Only for Experiment I) Remember that you have to decide even *where* the distortions are more evident. This is very important, so take in account, while you are interacting with the model, that you have to decide how much you perceive the distortion and where these are more evident. To indicate the part of the model that presents the most evident distortions we have to use this viewfinder, this sight [activate selection mode]. You can move the models as usual. The rectangle can be resized by using the keys '**A**', '**D**', '**X**' and '**W**'. Pay attention. You have to move the model to select the part, not the selection rectangle. Press <**ENTER**>, on the keyboard, to confirm your selection. Now, try to indicate some parts on this model. [Interaction trial - selection]
- Have you understood? Do you have any questions?
- Now I will show to you some cases that help you to evaluate numerically *how much* you perceive the distortions. You have to choose the worst case and mentally assign a value of 10 to it. This will give you an idea of the distortions that you will be seeing. The distortions could be present or not on the model. So, you are assigning a score of 10 to the most evident distortions. If the perception of the distortions during the test is half of the worst examples you chose, give it 5; if it is 1/10th as bad give it 1, if it is 1.5 times as bad, give it 15. This is important, you can give score higher than 10. Remember that the question is *how much you perceive, how much you notice such kind of distortions*. I will show these examples now.
- [Show worst cases]
- Before we start the experiment you will have six practice trials to be sure that you understand the task. You will respond in these trials just like you will in the main experiment. The questions appear on the screen. So, you have to provide three answers (two for the Experiment II) after the interaction with the model. The first question is: did you notice any distortion? You can answer <**YES**> or <**NO**> at this question. Then, in case of positive answer you have to give a score to indicate *how much* the distortion is evident. You use the numeric keypad to enter the perception value. And finally, the third question, (only for Experiment I) *where* you noticed the distortions. To answer to this question you have to select the part of the model with the most evident distortions.
- Do you have any question? Do you want to repeat it?
- [Practice Trials]
- 40 models will be shown to you during the test (48 in Experiment II). This takes about 20 minutes (24 minutes for Experiment II) plus the time you need to indicate where the distortions are more evident (only for Experiment I). So, the test will takes about 30 minutes.
- Before to start the test I would like to give you the following practical recommendations:
  1) In case of input error, please tell me what you want to do and I correct your answer at the end of the test.
  2) You can take a break at any time by entering your answers (score and part) for the most recent models, but waiting to hit <**ENTER**> until you are ready to go on.
  3) Finally, at the end of the test I will ask to you few questions.
- Do you have any question before to start?
- [Start the experiment]
- Interview
  1) What is your feeling with the models? I mean, have you experienced any problem to identify the distortions on a specific model and why?
  2) How would you describe the distortions that you saw?
  3) Have you comments or remarks about the tests?

References

[1] A. H. Tewfik and M. Swanson, "Data hiding for multimedia personalization, interaction and protection," *IEEE Signal Processing Magazine*, vol. 14, no. 4, pp. 41–44, 1997.

[2] I. Cox and M. L. Miller, "A review of watermarking and the importance of perceptual modeling," in *Proceedings of SPIE: Vol. 3016. Human Vision and Electronic Imaging II*, Feb. 1997, pp. 92–99.

[3] R. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1108–1126, 1999.

[4] P. G. Barten, "Evaluation of subjective image quality with the square-root integral method." *Journal of Optical Society of America*, vol. 7, no. 10, pp. 2024–2031, 1990.

[5] A. B. Watson, "Efficiency of an image code based on human vision," *Journal of Optical Society of America*, vol. 4, no. 12, pp. 2401–2417, 1987.

[6] ——, "DCT quantization matrices visually optimized for individual images," in *Proceedings of SPIE: Vol. 1913, Human Vision, Visual Processing and Digital Display IV*, 1993, pp. 202–216.

[7] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *Journal of Optical Society of America*, vol. 70, no. 12, pp. 1458–1471, 1980.

[8] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 525–539, Apr. 1998.

[9] D. Kundur and D. Hatzinakos, "A robust digital watermarking method using wavelet-based fusion," in *Proceedings of IEEE International Conference on Image Processing '97*, vol. 1, Oct. 1997, pp. 544–547.

[10] R. B. Wolfgang, C. I. Podilchuk, and E. J. D. III, "Perceptual watermarks for digital images and video," in *Security and Watermarking of Multimedia Contents*, P. W. Wong and E. J. D. III, Eds., vol. 3657, no. 1. San Jose, CA, USA: SPIE, 1999, pp. 40–51.
[11] R. G. V. Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Proceedings of IEEE International Conference on Image Processing '94*, vol. 2, Nov. 1994, pp. 86–90.
[12] F. Bartolini, M. Barni, V. Cappellini, and A. Piva, "Mask building for perceptually hiding frequency embedded watermarks," in *Proceedings of the 5th IEEE International Conference on Image Processing, ICIPS̈98*, vol. I, Chicago, IL, USA, Oct. 1998, pp. 450Ű–454.
[13] J. F. Delaigle, C. D. Vleeschouwer, and B. Macq, "Watermarking algorithm based on a human visual model," *Signal Processing*, vol. 66, no. 3, pp. 319–336, 1998.
[14] R. S. P. Cignoni, C. Rocchini, "Metro: measuring error on simplified surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167–174, 1998.
[15] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "Mesh: Measuring error between surfaces using the hausdorff distance," in *Proceedings of the IEEE International Conference on Multimedia and Expo 2002 (ICME)*, vol. I, 2002, pp. 705–708.
[16] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proceedings of the first IEEE International Conference on Image Processing*, vol. 2, Nov. 1994, pp. 982–986.
[17] S. Daly, "The visible difference predictor; an algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993.
[18] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed. New Jersey: World Scientific, 1995, pp. 245–283.
[19] P. Lindstrom and G. Turk, "Image-driven simplification," *ACM Transaction on Graphics*, vol. 19, no. 3, pp. 204–241, 2000.
[20] N. Williams, D. Luebke, J. D. Cohen, M. Kelley, and B. Schubert, "Perceptually guided simplification of lit, textured meshes," in *Proceedings of the 2003 symposium on Interactive 3D graphics*. ACM Press, 2003, pp. 113–121.
[21] B. Rogowitz and H. Rushmeier, "Are image quality metrics adequate to evaluate the quality of geometric objects?" in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 4299. SPIE Proc., 2001, pp. 340–348.
[22] Y. Pan, I. Cheng, and A. Basu, "Perceptual quality metric for qualitative 3D scene evaluation," in *Proceedings of International Conference on Image Processing 2003*, vol. 3, Sept. 2003, pp. 169–172.
[23] M. Reddy, "Perceptually modulated level of detail for virtual environments," Ph.D. dissertation, University of Edinburgh, 1997.
[24] M. R. Bolin and G. W. Meyer, "A perceptually based adaptive sampling algorithm," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM Press, 1998, pp. 299–309.
[25] K. Myszkowski, P. Rokita, and T. Tawara, "Perceptually-informed accelerated rendering of high quality walkthrough sequences," in *Proceedings of the Tenth Eurographics Workshop on Rendering*, 1999, pp. 5–18.
[26] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 73–82.
[27] J. A. Ferwerda, P. Shirley, S. N. Pattanaik, and D. P. Greenberg, "A model of visual masking for computer graphics," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 143–152.
[28] M. Corsini, "Towards blind and robust watermarking of 3d objects," Ph.D. dissertation, University of Florence, Italy, 2004, in preparation.
[29] R. Ohbuchi, H. Masuda, and M. Aono, "Watermaking three-dimensional polygonal models," in *Proceedings of the fifth ACM international conference on Multimedia*. Seattle, Washington, United States: ACM Press, 1997, pp. 261–272.
[30] F. Cayre and B.Macq, "Data hiding on 3-d triangle meshes," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 939–949, Apr. 2003.
[31] O. Benedens, "Two high capacity methods for embedding public watermarks into 3D polygonal models," in *Proceedings of the Multimedia and Security-Workshop at ACM Multimedia 99*, Orlando, Florida, 1999, pp. 95–99.
[32] T. Harte and A. Bors, "Watermarking 3D models," in *Proceedings of IEEE International Conference on Image Processing 2002*, vol. III, Rochester, NY, USA, Sept. 2002, pp. 661–664.
[33] O. Benedens, "Watermarking of 3D polygon based models with robustness against mesh simplification," in *Proceedings of SPIE: Security and Watermarking of Multimedia Contents*, vol. 3657, 1999, pp. 329–340.
[34] M. G. Wagner, "Robust watermarking of polygonal meshes," in *Proceedings of Geometric Modeling and Processing 2000. Theory and Applications.*, Hong Kong, China, Apr. 2000, pp. 201–208.
[35] E. Praun, H. Hoppe, and A. Finkelstein, "Robust mesh watermarking," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 49–56.
[36] F. Uccheddu, M. Corsini, and M. Barni, "Wavelet-based blind watermarking of 3d models," in *Proceedings of the 2004 multimedia and security workshop on Multimedia and security*. ACM Press, 2004, pp. 143–154.
[37] S. Kanai, H. Date, and T. Kishinami, "Digital watermarking for 3D polygons using multiresolution wavelet decomposition," in *Proceedings of the Sixth IFIP WG 5.2 International Workshop on Geometric Modeling: Fundamentals and Applications (GEO-6)*, Tokyo, Japan, Dec. 1998, pp. 296–307.
[38] R. Ohbuchi, S. Takahashi, T. Miyazawa, and A. Mukaiyama, "Watermarking 3D polygonal meshes in the mesh spectral domain," in *Proceedings of Graphics interface 2001*. Ottawa, Ontario, Canada: Canadian Information Processing Society, 2001, pp. 9–17.
[39] R. Ohbuchi, A. Mukaiyama, and S. Takahashi, "A frequency-domain approach to watermarking 3D shapes," in *Proceedings of EUROGRAPHICS 2002*, Saarbrucken, Germany, Sept. 2002.
[40] F. Cayre, P. R. Alface, F. Schmitt, B. Macq, and H. Maître, "Application of spectral decomposition to compression and watermarking of 3D triangle mesh geometry," *Image Communications - Special issue on Image Security*, vol. 18, pp. 309–319, Apr. 2003.
[41] *Methodology for Subjective Assesment of the Quality of Television Pictures Recommendation BT.500-11*. International Telecommunication Union, Geneva, Switzerland, 2002.
[42] *Subjective Video Quality Assesment Methods for Multimedia Applications Recommendation P.910*. International Telecommunication Union, Geneva, Switzerland, 1996.
[43] T. Akenine-Möller and E. Haines, *Real-Time Rendering (second edition)*. AK Peters, 2002.
[44] R. Shacked and D. Lischinki, "Automatic lighting design using a perceptual quality metric," *Computer Graphics Forum*, vol. 20, no. 3, 2001.
[45] A. Watt, *Fundamentals of three-dimensional computer graphics*. Addison-Wesley Publishing Co., 1990.
[46] E. B. Goldstein, *Sensation and Perception*. 4th Edition: Internazional Thomson Publishing Company, 1996.
[47] K. Shoemake, "Arcball rotation control," in *Graphics Gems IV*, P. Heckbert, Ed. Academic Press, 1994, pp. 175–192.
[48] B. Watson, A. Friedman, and A. McGaffey, "Using naming time to evaluate quality predictors for model simplification," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 2000, pp. 113–120.

[49] M. S. Moore, "Psychophysical measurement and prediction of digital video quality," Ph.D. dissertation, University of California, Santa Barbara, USA, 2002.
[50] [Online]. Available: http://www.multires.caltech.edu/software/pgc
[51] G. W. Snedecor and W. G. Cochran, *Statistical Methods*. Iowa State University, Press, Ames, 1989.
[52] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, Massachusetts: MIT Press, 1993.
[53] F. Uccheddu, M. Corsini, M. Barni, and V. Cappellini, "A roughness-based algorithm for perceptual watermarking of 3d meshes," in *Proceedings of the Tenth International Conference on Virtual System and Multimedia*, Nov. 2004.
[54] J.-H. Wu, S.-M. Hu, J.-G. Sun, and C.-L. Tai, "An effective feature-preserving mesh simplification scheme based on face constriction," in *Proceedings of the 9th Pacific Conference on Computer Graphics and Applications*. IEEE Computer Society, 2001, p. 12.
[55] G. Taubin, "A signal processing approach to fair surface design," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM Press, 1995, pp. 351–358.
[56] L. Kobbelt, "Discrete fairing," in *Proceedings of the Seventh IMA Conference on the Mathematics of Surfaces '97*, 1997, pp. 101–131.
[57] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr, "Implicit fairing of irregular meshes using diffusion and curvature flow," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 317–324.
[58] T. R. Jones, F. Durand, and M. Desbrun, "Non-iterative, feature-preserving mesh smoothing," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 943–949, 2003.