Contributions to objective and subjective visual quality assessment of 3d models

Jinjiang Guo

Jinjiang Guo. Contributions to objective and subjective visual quality assessment of 3d models. Other [cs.OH]. Université de Lyon, 2016. English. NNT : 2016LYSEI099 . tel-01715610

HAL Id: tel-01715610
https://theses.hal.science/tel-01715610v1

Submitted on 22 Feb 2018

N° d'ordre NNT : 2016LYSEI099

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'INSA Lyon

École Doctorale EDA 512
INFORMATIQUE ET MATHEMATIQUES DE LYON

Spécialité de doctorat : Informatique
Discipline : Informatique

Soutenue publiquement le 06/10/2016, par :
Jinjiang GUO

Contributions to Objective and Subjective Visual Quality Assessment of 3D Models

---

Devant le jury composé de :

DANIEL Marc, Professeur des Universités, LSIS, Polytech Marseille	Président
DAOUDI Mohamed, Professeur des Universités, LIFL, TELECOM Lille 1	Rapporteur
PUECH William, Professeur des Universités, LIRMM	Rapporteur
DANIEL Marc, Professeur des Universités, LSIS, Polytech Marseille	Examinateur
LARABI Mohamed-Chaker, Maître de Conférences, XLIM-SIC	Examinateur
LAVOUÉ Guillaume, Maître de Conférences HDR, LIRIS, INSA Lyon	Directeur de thèse
VIDAL Vincent, Maître de Conférences, LIRIS, Université Lyon 1	Co-directeur de thèse
BASKURT Atilla, Professeur des Universités, LIRIS, INSA Lyon	Co-directeur de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales – Quinquennal 2016-2020

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec : Renée EL MELHEM Bat Blaise Pascal 3e etage secretariat@edchimie-lyon.fr Insa : R. GOURDON	M. Stéphane DANIELE Institut de Recherches sur la Catalyse et l'Environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 avenue Albert Einstein 69626 Villeurbanne cedex directeur@edchimie-lyon.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Sec : M.C. HAVGOUDOUKIAN Ecole-Doctorale.eea@ec-lyon.fr	M. Gérard SCORLETTI Ecole Centrale de Lyon 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60.97 Fax : 04 78 43 37 17 Gerard.scorletti@ec-lyon.fr
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** **http://e2m2.universite-lyon.fr** Sec : Safia AIT CHALAL Bat Darwin - UCB Lyon 1 04.72.43.28.91 Insa : H. CHARLES Safia.ait-chalal@univ-lyon1.fr	Mme Gudrun BORNETTE CNRS UMR 5023 LEHNA Université Claude Bernard Lyon 1 Bât Forel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 06.07.53.89.13 e2m2@ univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** **http://www.ediss-lyon.fr** Sec : Safia AIT CHALAL Hôpital Louis Pradel - Bron 04 72 68 49 09 Insa : M. LAGARDE Safia.ait-chalal@univ-lyon1.fr	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 avenue Jean Capelle INSA de Lyon 696621 Villeurbanne Tél : 04.72.68.49.09 Fax :04 72 68 49 16 Emmanuelle.canet@univ-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** **http://infomaths.univ-lyon1.fr** Sec :Renée EL MELHEM Bat Blaise Pascal 3e etage infomaths@univ-lyon1.fr	Mme Sylvie CALABRETTO LIRIS – INSA de Lyon Bat Blaise Pascal 7 avenue Jean Capelle 69622 VILLEURBANNE Cedex Tél : 04.72. 43. 80. 46 Fax 04 72 43 16 87 Sylvie.calabretto@insa-lyon.fr
Matériaux	**MATERIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry Ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIERE INSA de Lyon MATEIS Bâtiment Saint Exupéry 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 71.70 Fax 04 72 43 85 28 Ed.materiaux@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** **http://mega.universite-lyon.fr** Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72 .43.71.70 Fax : 04 72 43 72 37 Philippe.boisse@insa-lyon.fr
ScSo	**ScSo*** **http://recherche.univ-lyon2.fr/scso/** Sec : Viviane POLSINELLI Brigitte DUBOIS Insa : J.Y. TOUSSAINT viviane.polsinelli@univ-lyon2.fr	Mme Isabelle VON BUELTZINGLOEWEN Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.86 Fax : 04.37.28.04.48

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

I dedicate this thesis to my beloved grandfathers in heaven and my family.

> 昨夜西风凋碧树　独上高楼　望尽天涯路
> — 晏殊《蝶恋花·槛菊愁烟兰泣露》
>
> "Last night the west wind shriveled the green-clad trees
> Alone I climb the high tower
> To gaze my fill along the road to the horizon" [88]
> — Yan Shu (991-1055), *Love of Butterfly*

Acknowledgements

First and foremost, I would like to express my special appreciation and thanks to my advisor Dr. Guillaume Lavoué, you have been a tremendous mentor for me. Your expertise, patience and cautious attitudes of scientific study shape me to grow as a researcher. I still remember everything vividly: you helped me to debug each tiny error in my messy codes, patiently instructed me with theoretical knowledge, and efficiently corrected my work and papers... Meanwhile, your thoughtful consideration and life supports help me to survive from these four tough years abroad. Your advice on both research as well as on my future life have been invaluable. I would also like to express my great gratitude to my other excellent advisors Dr. Vincent Vidal and Pr. Atilla Baskurt. I am very grateful to Dr. Vincent Vidal for your insightful suggestions and comments. With your talent and skillful programming abilities, you guide me to solve every problem in my work. Also, thank you so much, Pr. Atilla Baskurt, for your general guidelines on my PhD study, instructions and encouragements. All these help me to cultivate an attitude of being a qualified researcher.

I am also thankful to Dr. Irene Cheng and Pr. Anup Basu in University of Alberta, for your thoughtful care and the expertise on my study. I grew very fast, and learned a lot from your informative instructions. I really enjoyed my stay in Canada. Especially, your lovely daughter Vina added the joyful moments.

I would also like to thank my jury members: Pr. Mohamed Daoudi, Pr. William Puech, Pr. Marc Daniel and Dr. Mohamed-Chaker Larabi for serving as my jury members. I am very grateful to Pr. Mohamed Daoudi and Pr. William Puech for taking your precious time to review my manuscript, and I also appreciate Pr. Marc Daniel and Dr. Mohamed-Chaker Larabi for examining my thesis and defense. I also want to thank you all for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

Many thanks to all my lovely colleagues and friends who once helped me and shared the enjoyable moment with me. Thank you very much, Dr. Mingyuan Jiu, Dr. Lilei Zheng, for your generous help and advice on my work. And as such, Ms. Xiaolu Jiang, Ms. Jingping Li, Ms. Wenjun Hao, Mr. Qing Liu, Mr. Alexandre Hilth, Mr. Yue Ma, Mr. Rémi Dufresne, Dr. Min Yao, Dr. Yue Zhang, Mr. Xichen Yuan, Dr. Bin Bao, Dr. Lei Han, Mr. Quanyi Yin, Ms.

Wenchun Jiang, Ms. Wen Chen, Dr. Fei Liu, Ms. Yaoyi Hu, etc., thank you all for sharing those unforgettable moments when we laughed, suffered and strove together for the bright future.

Special thanks to my family. Words cannot express how grateful I am to my mother Mme. Xinhua Wang and father Mr. Zhiyong Guo, for all of your supports and sacrifices that you have made on my behalf.

Finally, I would like to dedicate this thesis to my most respectable and beloved grandfathers Mr. Ping Guo and Mr. Guangde Yu, two great people I lost during my PhD study. I have accomplished your wishes.

Abstract

In computer graphics realm, three-dimensional graphical data, generally represented by triangular meshes, have become commonplace, and are deployed in a variety of application processes (e.g., smoothing, compression, remeshing, simplification, rendering, etc.). However, these processes inevitably introduce artifacts, altering the visual quality of the rendered 3D data. Thus, in order to perceptually drive the processing algorithms, there is an increasing need for efficient and effective subjective and objective visual quality assessments to evaluate and predict the visual artifacts.

In this thesis, we first present a comprehensive survey on different sources of artifacts in digital graphics, and current objective and subjective visual quality assessments of the artifacts. Then, we introduce a newly designed subjective quality study based on evaluations of the local visibility of geometric artifacts, in which observers were asked to mark areas of 3D meshes that contain noticeable distortions. The collected perceived distortion maps are used to illustrate several perceptual functionalities of the human visual system (HVS), and serve as ground-truth to evaluate the performances of well-known geometric attributes and metrics for predicting the local visibility of distortions.

Our second study aims to evaluate the visual quality of texture-mapped 3D model subjectively and objectively. To achieve these goals, we introduced 136 processed models with both geometric and texture distortions, conducted a paired-comparison subjective experiment, and invited 101 subjects to evaluate the visual qualities of the models under two rendering protocols. Driven by the collected subjective opinions, we propose two objective visual quality metrics for textured meshes, relying on the optimal combinations of geometry and texture quality measures. These proposed perceptual metrics outperform their counterparts in term of the correlation with the human judgment.

Keywords : Computer Graphics, 3D mesh models, Visual artifact, Textured mesh, Mesh Visual Quality Assessment, Image Quality Assessment, Perceptual metrics, Subjective study.

Résumé

Dans le domaine de l'informatique graphique, les données tridimensionnelles, généralement représentées par des maillages triangulaires, sont employées dans une grande variété d'applications (par exemple, le lissage, la compression, le remaillage, la simplification, le rendu, etc.). Cependant, ces procédés introduisent inévitablement des artefacts qui altèrent la qualité visuelle des données 3D rendues. Ainsi, afin de guider perceptuellement les algorithmes de traitement, il y a un besoin croissant d'évaluations subjectives et objectives de la qualité visuelle à la fois performantes et adaptées, pour évaluer et prédire les artefacts visuels.
Dans cette thèse, nous présentons d'abord une étude exhaustive sur les différentes sources d'artefacts associés aux données numériques graphiques, ainsi que l'évaluation objective et subjective de la qualité visuelle des artefacts.
Ensuite, nous introduisons une nouvelle étude sur la qualité subjective conçue sur la base de l'évaluations de la visibilité locale des artefacts géométriques, dans laquelle il a été demandé à des observateurs de marquer les zones de maillages 3D qui contiennent des distorsions visibles. Les cartes de distorsion visuelle collectées sont utilisées pour illustrer plusieurs fonctionnalités perceptuelles du système visuel humain (HVS), et servent de vérité-terrain pour évaluer les performances des attributs et des mesures géométriques bien connus pour prédire la visibilité locale des distorsions.
Notre deuxième étude vise à évaluer la qualité visuelle de modèles 3D texturés, subjectivement et objectivement. Pour atteindre ces objectifs, nous avons introduit 136 modèles traités avec à la fois des distorsions géométriques et de texture, mené une expérience subjective de comparaison par paires, et invité 101 sujets pour évaluer les qualités visuelles des modèles à travers deux protocoles de rendu. Motivés par les opinions subjectives collectées, nous proposons deux mesures de qualité visuelle objective pour les maillages texturés, en se fondant sur les combinaisons optimales des mesures de qualité issues de la géométrie et de la texture. Ces mesures de perception proposées surpassent leurs homologues en termes de corrélation avec le jugement humain.

Mots-clés : Informatique graphique, Maillages 3D, Artefact visuel, Maillage texturé, Évaluation de la Qualité Visuelle des Maillages, Évaluation de la Qualité des Images, Métriques perceptuelles, Étude subjective.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Context

In the teachings of Gautama Buddha, the *Five Aggregates*: "Form, Sensation, Perception, Mental Formation and Consciousness", reveal the way that a human observes, experiences and understands the world with his psychological cognitive processes. The cognition based on the visual system is a universal and frequent process, even in today's world, where the growing technology has changed people's lives dramatically. Especially, graphics, from ancient paintings to modern computer graphics, describe the visual perception, and reflect the human cognition. In present days, the graphics have evolved from simple two-dimensional lines to diverse three-dimensional graphical data (see Figure 1.1), which have become commonplace in the vast majority of industrial domains involving medical imaging, mechanical engineering, digital entertainment, e-commerce, scientific visualization, architecture design, virtual reality and so forth. As a result of these proliferating and various uses, the 3D graphical data, generally formed by triangular meshes, are subject to various processing operations, such as compression, simplification, transmission, remeshing, watermarking, etc. Inevitably, these manipulations introduce artifacts, which impact on the visual perception more or less, or in other words, alter the visual quality of the processed data.

Considering the final receivers of the 3D data are human eyes, which appeals to costumers for the satisfying *Quality of Experience*, it is crucial to evaluate the quality of the distorted 3D objects, namely the visual annoyance amount caused by the artifacts. Thus, many researchers proposed numerous quality functions (i.e. metrics), attempting to find a common evaluation standard of the visual perception. For instance, the pioneer researchers only relied on simple geometric attributes (e.g., Hausdorff or RMS distances) to evaluate the distortion quality between a reference model and its processed version. However, these metrics performed poorly in terms of correlation with human vision. Recently, the scientific community has

Figure 1.1: Examples of different 3D graphical data. from left to right: Dragon mesh, vertex number : 50,000; Squirrel mesh mapped with texture, vertex number: 6,185; Bunny mesh, vertex number : 34,835.

devoted to devise different mesh visual quality (MVQ) metrics considering the connections between graphical attributes (including geometric and rendering attributes) and the Human Visual System (HVS), or the functionalities of the HVS. These visual metrics aim to predict the perceived visual quality of distorted 3D data, and have surpassed the previous classical metrics regarding the correlation with visual perception. Nevertheless, there still exist two fundamental problems:

- Among the diverse metrics, which metric (and/or attribute) is able to provide the predictions that are mostly correlated with human perception?
- How to design a visual quality metric that can thoroughly fulfill different evaluation tasks (with a variety of distortions and rendering scenarios), predicting the perceived visual quality efficiently and effectively?

In order to evaluate the performance of the metrics, a metric prediction/value must be compared with the human subjective assessment, usually by computing the correlations between the global quality value produced by the metric and the subjective score from observers. However, besides this global information, in many cases it is also necessary to evaluate their performances using information about the local visibility of the artifacts (i.e. the visually noticeable distortions on small regions of a mesh). Such local information is relevant to drive locally any geometry processing operations, such as mesh simplification, compression and so forth. Moreover, among the existing 3D mesh visual quality metrics, they mainly consider either the geometric distortions or the artifacts generated during the rendering steps (see Figure 1.2). Little work has been done to evaluate the visual impact of the artifacts on textured 3D models in consideration of complex perceptual interactions (i.e. geometric, texture attributes and rendering scenarios).

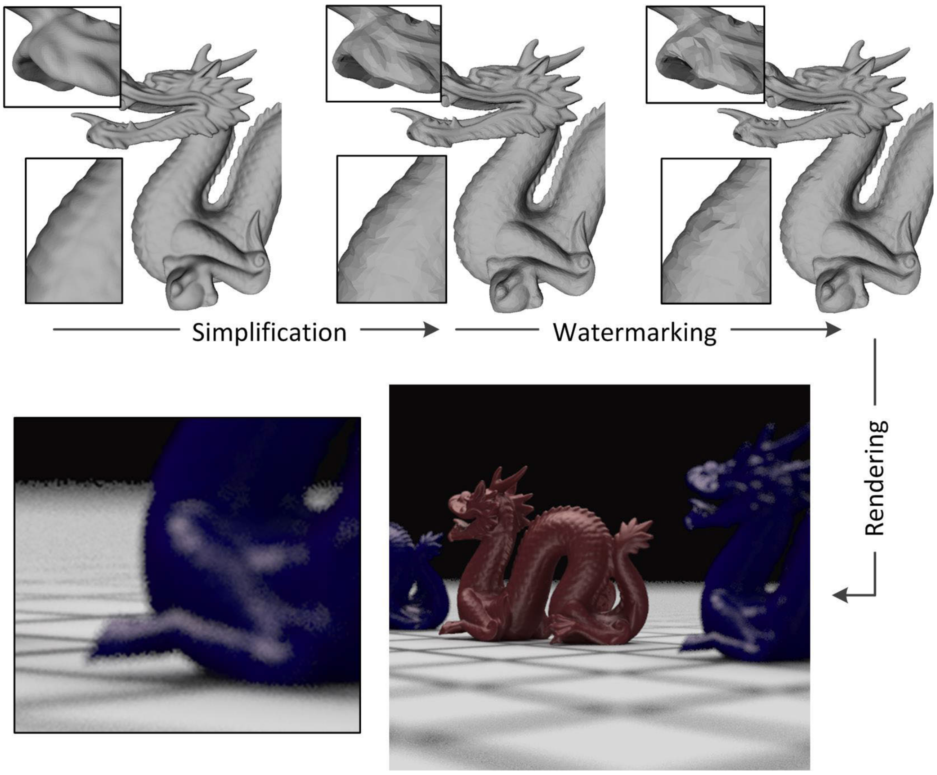


Figure 1.2: Illustration of a typical computer graphics work-flow and its different causes of artifacts. *Top row, from left to right*: An original scanned 3D Dragon model (338K vertices); A simplified version (50K vertices) with uniform high frequency noise; result after watermarking [123] which produces some local bumps on the surface. *Bottom row*: Result after rendering (radiance caching) which creates a non uniform structured noise. This work-flow is presented in [61].

1.2 Objectives and Contributions

The research topic of this thesis work is the design of subjective visual quality assessment protocols and objective visual quality metrics for 3D objects. Our main objectives are: (1) to evaluate the performance of existing 3D mesh quality metrics in the terms of correlation with the human vision on perceiving local distortions; (2) for the texture mapped 3D models, to evaluate the available metrics with regards to correlation with visual perception in different evaluation tasks (with various artifacts and rendering scenarios), and then to propose new objective quality metrics.

In order to accomplish the first objective, we designed a subjective quality study based on evaluations of the local visibility of geometric artifacts, in which observers were asked to mark the areas of 3D meshes that contain noticeable distortions. The collected perceived distortion maps are used to illustrate several perceptual functionalities of HVS, and serve as ground-truth to evaluate the performance of well-known geometric attributes and metrics for predicting the local visibility of distortions.

To achieve the second goal, we introduce 116 processed models with both geometric and texture distortions, conduct a paired-comparison subjective assessment, and invite 98 subjects to evaluate the visual qualities of the models under two different rendering protocols. Driven by the collected subjective opinions, we propose two objective quality metrics for textured meshes based on the optimal combinations of geometry and texture quality measures. Finally, we evaluate these two quality metrics along with other counterparts regarding the correlation with the subjective opinions.

1.3 Outline

The remainder of this manuscript is organized as follows:

Chapter 2 provides background knowledge on digital artifacts (distortions) caused by acquisition, conversion and processing operations. It covers common distortions existing in processed 2D images (involving videos) and 3D meshes.

Chapter 3 presents a comprehensive survey on objective visual quality metrics for 2D images, videos and 3D objects.

Chapter 4 describes the existing subjective visual quality assessments in computer graphics.

Chapter 5 presents our proposed subjective assessment, which quantitatively evaluates several well-known geometric quality metrics and attributes for the task of predicting perceived local distortions.

Chapter 6 details our subjective and objective quality assessments for texture mapped 3D models.

Chapter 7 summarizes the contributions of the manuscript, and proposes several future working directions concerning the research on visual quality assessments for 3D graphical data.

Chapter 2

Background knowledge on digital artifacts by acquisition, conversion and processing

In the computer graphic fields, various visual contents are produced, which can be mainly categorized into two types: 3D data and 2D images (digital videos are essentially sequences of 2D images). According to different sub-realms of applied graphics (e.g., scientific visualization, virtual reality, video gaming, engineering design, medical care, internet industry, etc.), extensive images, videos and 3D data are generated and manipulated. Usually, depending on different manipulation purposes (e.g., remote transmission, animation, scanning, rendering, etc.), the relevant operations, such like compression, sub-sampling, simplification, smoothing and so forth, or the sensor and circuitry of a scanner/camera may cause the noticeable distortions in 2D images or graphical 3D data [43] [32]. These distortions are called 2D image artifacts and 3D geometric artifacts respectively.

2.1 2D image artifacts

2.1.1 Image noise

Usually, image noise is shown as random variance of the color or brightness in digital images, which often occurs during image capture, transmission [32], etc . The range of the noise magnitude can be from almost imperceptible spots on digital images to almost entirely noise due to sophisticated processing, poor acquisition or low precision of a sensor, scanner or digital camera. Depending on different formation causes, the image noise can be generally sorted into following types:

Figure 2.1: Some image noise examples. From left to right: Original Lena image, size: 512×512, White Gaussian noise (standard deviation 25), Multiplicative Speckle Noise. Both noises are generated by *MATLAB*.

Gaussian noise

Gaussian noise usually happens under poor conditions of acquisition, such like poor illumination and/or high temperature, etc. This noise appears as additive, independent at each pixel and independent of the signal intensity [32]. A major noticeable Gaussian noise is amplifier noise, which shows as constant noise level in dark areas or in the blue channel of a color image (See Figure 2.1).

Speckle noise

Speckle noise is of granulate, which exists in active radar and synthetic aperture radar (SAR) images. In traditional radar images, it results from the interference of the received waves from an object which is no larger than a single processing element, and consequently raises the average gray level in a local area. Speckle noise in SAR images is caused by coherent processing of scattered signals from multiple distributed targets [40]. The presence of speckle noise in an image reduces the ability of a human observer to resolve fine details in the image [65]. (See Figure 2.1).

Fat-tail distributed noise

Fat-tail distributed noise is also known as salt-and-pepper noise or impulsive noise, which is caused by malfunctioning pixels in camera sensors, faulty memory locations in hardware, or transmission in a noisy channel [11]. It manifests itself as sparsely occurring dark pixels in bright areas and white pixels in dark areas. (See Figure 2.2).

Figure 2.2: Left: Original Pepper image, size: 256 × 256, Right: Fat-tail distributed noise (added to 10% of the pixels). The noise is added by using *MATLAB*.

2.1.2 Smoothing artifacts

Aiming to handle the image noise or other fine-scale structure/rapid phenomena, image smoothing methods were proposed to capture important patterns in images. During this processing, the individual points of an input signal, which are more intensive than the immediately adjacent points, are modified and reduced, and thus the points that are lower than adjacent points rise. Naturally, a smoother signal is produced after this processing. Most common smoothing algorithms use spatial filters to smooth input data [66], where a set of original adjacent points are multiplied point-by-point by a group of coefficients that define the smooth shape, and then the summation of the products is divided by the sum of the coefficients (average value). The process repeats until all the original points are shifted down to the defined smooth values. Besides taking average values of coefficients, some algorithms take the median values or Gaussian weighted average values as smoothing fliters. The visually blurred artifacts caused by smoothing may be from a high extent of defined smooth shape and excessive smoothing repeations. Figure 2.3 shows a smoothing example using Gaussian filter with different σ values.

2.1.3 Compression artifacts

Transmission and storage of digital images are almost required in every domain related to computer graphics. Thus, in order to decrease the transmission and storage costs, image compression algorithms were proposed to reduce irrelevance and redundancy of images, which can be lossy or lossless. Between the two types of compression methods, the lossy compression may introduce perceptible artifacts, especially, at low bit rates.

Figure 2.3: A smoothing example using Gaussian filter with different σ values. From left to right: Original Lena image, size: 512×512; Image with Gaussian smoothing, σ=3; Image with Gaussian smoothing, σ=5. The original image is manipulated by *MATLAB*.

JPEG

JPEG (Joint Photographic Experts Group) compression [122] is a commonly used lossy compression method based on an 8×8 block-size discrete cosine transform (DCT), which convert the image (or each frame of a video source, which refers to the video compression for reducing spatial redundancy of video sequences in MPEG standards [36]) from spatial domain into frequency domain. More precisely, let $f(x,y)$ represent an unsigned integer image value (e.g., color, hue, intensity, etc.) in $[0, 2^P - 1]$ at the location (x,y) of input image. The output of DCT $F(u,v)$ shifted $f(x,y)$ to a signed integer in the range of $[-2^{P-1} - 1, 2^{P-1} - 1]$, the transform is expressed as follows:

$$F(u,v) = \frac{1}{4}C(u)C(v)[\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}] \tag{2.1}$$

where $0 \leq u < 8$, $0 \leq v < 8$, $C(u), C(v) = 1/\sqrt{2}$ for $u, v = 0$; and $C(u), C(v) = 1$ otherwise. In frequency domain, in an 8×8 pixels block, the quantization algorithm (i.e. DTC), which is related to the discrete fourier transform (DFT), is opted to optimally reduce a large scale of coefficients into a smaller scale by using the equation 2.1 to remove the high-frequency coefficients, since the high-frequency coefficients (such like, sharp transitions in color hue, intensity, etc.) provide fewer contributions to the whole image than other frequency coefficients [122]. Finally, the quantized coefficients are sequenced and losslessly packed into the output bitstream. The quantization step is a many-to-one mapping proccedure, and therefore is fundamentally lossy [122]. Usually, a high level of compression ratio, which means the ratio between size of compressed file and its original file, results in blocking artifacts on the compressed image, because the quantization operation inevitably makes noise around contrasting edges to meet the high compression ratio (see Figure 2.4).

Figure 2.4: A comparison between JPEG and JPEG2000. From top to bottom: Original photo, size: 1024 × 768; JPEG format, size: 576 × 224, compression ratio: 1:33.65; JPEG2000 format, size: 576 × 224, compression ratio: 1:33.65. The photo is captured from *http://www.wikipedia.com*, Author: *Shlomi Tal*.

JPEG 2000

JPEG 2000 compression [103], proposed by Joint Photographic Experts Group in 2000, can be lossy as well. It inherits the JPEG standard, and replaces the previous discrete cosine transform (DCT) with a discrete wavelet-based transform (DWT), which filters the input image signals with low-pass and high-pass filters, and then downsamples the filtered signals by discarding odd indexed samples, whereas JPEG uses an 8×8 block-size DCT. The low-pass filter preserves the low frequencies of a signal while attenuating or eliminating the high frequencies, meanwhile the high-pass filter preserves the high frequencies of an image such as edges, texture and detail, when discarding or attenuating the low frequencies [75]. Since the DWT eliminates blocking artifacts at high compression ratios, the JPEG 2000 performs better-measured fidelity over JPEG, however, the elimination of high frequencies from the low-pass filter results in a blurred version of the original image [75] (see Figure 2.4).

2.1.4 Resampling (Sub-sampling) artifacts

For different image processing purposes, an image may be resized to a finer matrix in order to improve its appearance, or fitted to new position coordinates after a transformation (e.g., rotation, scaling, etc.) [82], etc. Hence, several resampling algorithms were designed to meet the demands. However, the resampling algorithms can result in visible artifacts, such as ringing, aliasing, blocking and blurring [111], especially, when they are used to enlarge subsampled images by interpolating image values (e.g., RGB, hue, grayscale, etc) between pixels in a local neighborhood (see Figure 2.5). Three common resampling methods based on different interpolation algorithms are introduced as below:

Nearest-neighbor interpolation based resampling

Nearest-neighbor interpolation [82] is a simple method for approximating an image value of non-given point at some location in an image while the values of neighboring points are given. It interpolates an image value into the location by convolving the value of the nearest point with a rectangle function. However, this is a basic low pass filter which results in some aliasing or blurring artifacts [82] (see Figure 2.5). This algorithm is commonly used in texture mapping in 3D rendering.

Bilinear interpolation based resampling

Bilinear interpolation performs a linear interpolation in two directions (x and y directions) on a rectilinear 2D grid of an image. In practice, when an image needs scaling up, each image pixel needs to be moved along a certain direction according to the scale value, which inevitably leaves many non-valued pixels (i.e., holes) [87]. In this case, the bilinear interpolation uses the values of 4 nearest pixel in diagonal directions from a non-valued pixel to assign appropriate image value to the hole.

We demonstrate the algorithm using a one-dimensional (1-D) case. Two dimensional (2-D) image data are handled by sequential x direction and y direction 1-D interpolations. Assume $f(x_k)$ is an image value at the location x_k, and $f'(x)$ is the value to be interpolated at the location x. Suppose that the nearest available neighbors of x are x_k (left) and x_{k+1} (right), and the distance between x_k and x_{k+1} equals to 1. Define the distances between x, x_k and x_{k+1} as: $d = x - x_k$ and $1 - d = x_{k+1} - x$. The $f'(x)$ is therefore computed using bilinear interpolation as follows:

$$f'(x) = (1-d)f(x_k) + df(x_{k+1}) \tag{2.2}$$

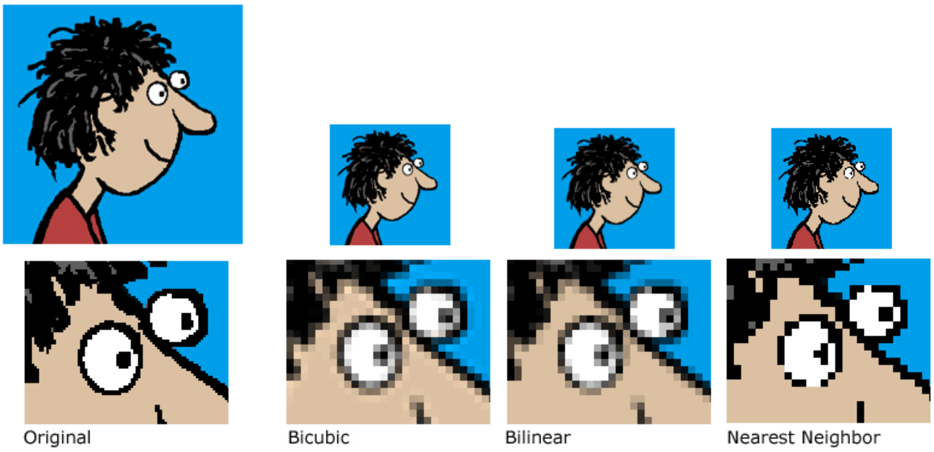


Figure 2.5: Comparison of different resampling algorithms. From left to right: Original image, size: 200 × 200; Resampled image with Bicubic interpolation, size: 100 × 100; Resampled image with Bilinear interpolation, size: 100 × 100; Resampled image with Nearest-neighbor interpolation, size: 100 × 100. The cartoon images are from *http://www.chrismadden.co.uk*.

Like the nearest-neighbor interpolation, this algorithm is commonly used in texture mapping of 3D object as well (see Figure 2.5).

Bicubic interpolation based resampling

Bicubic interpolation [87] is an extension of bilinear interpolation. Unlike the bilinear interpolation, which consider 4 neighboring pixels (2×2), bicubic interpolation takes 16 pixels (4×4) into account, which produces a smoother rescaled image with fewer artifacts. Similar to the demonstration of bilinear interpolation, we interpret the bicubic interpolation in one-dimensional case involving two more values at the coordinates: x_{k-1} and x_{k+2} in the following way:

$$\begin{aligned} f'(x) = & f(x_{k-1})(-d^3+2d^2-d)/2 \\ & +f(x)(3d^3-5d^2-2)/2 \\ & +f(x_{k+1})(-3d^3+4d^2-d)/2 \\ & +f(x_{k+2})(-d^3+d^2)/2 \end{aligned} \tag{2.3}$$

However, this algorithm has a larger time complexity than bilinear interpolation and nearest-neighbor interpolation. Similar to the nearest-neighbor interpolation, both bilinear and bicubic interpolations can cause an amount of blurring in the resampled image [44] (see Figure 2.5).

Figure 2.6: Scanned meshes from stanford 3D model repository [68]. From left to right: Happy Buddha, scan size: 4,586,124 points; Dragon, scan size: 2,748,318 points.

2.2 3D geometric artifacts

In this part, the 3D graphic data that we focus on is 3D surface mesh, which consists of interconnected spatial facets (e.g. triangles, quadrangles. etc). And the 3D geometric artifacts solely refer to the distortions resulted from the geometric modeling and processing (e.g., smoothing, watermarking, etc.) steps, whereas the distortions caused by generating or post-processing 3D content images/videos (e.g., tone-mapping, rendering, etc.) are considered as image artifacts[61].

2.2.1 Mesh acquisition noise

Mesh acquisition noise on 3D mesh is introduced by inaccurate acquisitions from laser scanners. The scanned raw meshes are usually irregular and contain millions of triangles/quadrangles [43]. Figure 2.6 shows two examples of acquisition noise.

2.2.2 Smoothing artifacts

Besides the purpose of removing the high frequency noise caused by acquisition, mesh smoothing methods are motivated by advanced filtering processing, fair surface design, hole-filling and mesh deformation as well. The aim of smoothing is to fair the large polyhedral surface of arbitrary topology [110]. There exist four widely used smoothing algorithms: *Laplacien smoothing* [104], *Gaussian smoothing* [109], *Taubin smoothing* [109] and *Two-step smoothing* [5], of which Laplacien and Gaussian smoothing algorithms may introduce the perceived shrinkage distortion on 3D mesh, while Taubin and Two-step smoothing algorithms prevent the smoothed mesh from shrinkage but may still result in deformation artifacts due to over-iterative smoothing steps (see Figure 2.7).

Laplacian smoothing

Laplacian smoothing moves each vertex v_i on a mesh towards a new position $\bar{v_i}$ by applying an *umbrella operator* [120]. In details, firstly the operator computes a direction vector $U\vec{(v_i)}$ between the vertex v_i and the barycenter of its adjacent vertices v_j:

$$U\vec{(v_i)} = \frac{1}{N}\sum_{j=1}^{N}(v_j - v_i) \tag{2.4}$$

with N the number of adjacent vertices to v_i. Then a new positon $\bar{v_i}$ is set in the direction $U\vec{(v_i)}$, which is computed as: $\bar{v_i} = v_i + \lambda U\vec{(v_i)}$. λ is a positive damping factor smaller than 1 [109]. However, more iteration steps of Laplacian smoothing results in further mesh shrinkage.

Gaussian smoothing

In contrast to the Laplacian smoothing that relies on the umbrella operator, Gaussian smoothing [109] firstly computes a Gaussian weighted direction vector $G\vec{(v_i)}$ between each vertex v_i and the barycenter of its adjacent vertices v_j. $G\vec{(v_i)}$ is computed as follows:

$$G\vec{(v_i)} = \sum_{j=1}^{N}\omega_{ij}(v_j - v_i) \tag{2.5}$$

where N means the number of adjacent vertices to v_i, the Gaussian weights ω_{ij} are positive and add up to 1 [109]. Then the new position v_i' is set by moving v_i in the Gaussian weighted direction: $v_i' = v_i + \lambda G\vec{(v_i)}$, in which the damping factor λ is a positive number smaller than 1 [109]. Still, several repetitions of the smoothing process produce the shrinkage of the mesh.

Taubin smoothing

Taubin smoothing actually consists of two consecutive Gaussian smoothing steps. In a first step, Gaussian smoothing applies a positive scale factor λ to all the vertices of a mesh, whereas in a second step a negative scale factor μ ($0 < \lambda < -\mu$) is applied to all the vertices. The repetitions of these two steps produce a low pass filter, which filters out the perturbations (i.e. zero mean high curvature noise) above the underlying smooth surface [109]. Taubin smoothing effectively prevents a smoothed mesh from shrinkage, and thus contributes to a

Figure 2.7: An example of different smoothing methods. From left to right: Original Vase Lion model with 38,728 points; Laplacian smoothing, iteration steps: 3; Two-step smoothing,iteration steps: 3, feature angle: 30 degree, normal smoothing steps 10, vertex fitting steps 10; Taubin smoothing, iteration steps: 3, $\lambda = 0.5$, $\mu = -0.53$. All the models are processed by the software *Meshlab* [19].

better visual quality than than Laplacien and Gaussian smoothing methods, but it may cause the loss of mesh details after several iterations.

Two-step smoothing

Two-step smoothing is based on the iterative mean filtering [5], where in a first step all the similar normals of all mesh triangles that share a common edge or vertex are averaged together, and in a second step all the vertices are fitted to the new normals. This smoothing performs well in denoising but may lose some salient shape information, which could introduce visually distinguishable artifacts. Some different smoothing methods are shown in Figure 2.7.

2.2.3 Simplification artifacts

Sometime the meshes scanned or processed by acquisition or smoothing procedures are with very high resolution, which are not necessarily required or supported in some applications or devices [37], simplification processing methods are therefore extensively used for reducing or simplifying the size of the mesh. The algorithms incrementally reduce the number of vertices and edges on the mesh surface through clustering or decimation methods. However, oversimplification may lead to the deformation of the mesh, which introduces noticeable artifacts in forms of deformed surface, losses of mesh details, etc. In practice, four simplification algorithms are commonly used, which are presented as follows:

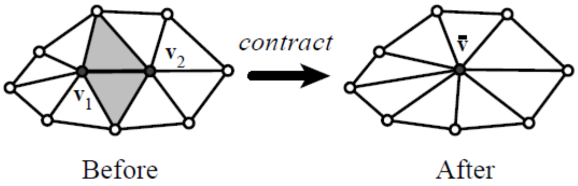


Figure 2.8: The highlighted edge is contracted into a single point. The shaded triangles become degenerate and are removed during the contraction. The figure is adapted from [37].

Vertex decimation

Vertex decimation iteratively removes a vertex and its adjacent faces, and fills the hole with retriangulation [96].

Vertex clustering (Quantization)

Quantization algorithms can be considered as simplification methods by clustering vertices into a center vertex of each 3D grid cell, which will be detailed in the section *Geometry compression artifacts*.

Iterative edge contraction

This method iteratively contracts the edges of a manifold surface to reduce the mesh complexity [91][46]. Figure 2.8 shows the edge contraction procedure, which is called *pair contraction* denoted as: $(v_1, v_2) \rightarrow \bar{v}$. The procedure moves the vertices v_1 and v_2 to the new position $\bar{v}$, connects all their incident edges to v_1, and deletes the vertex v_2.

Quadric edge collapse decimation

Based on the iterative edge contraction, the quadric edge collapse decimation contracts valid vertex pairs iteratively, and preserves surface error approximations relying on quadric matrices [37]. Figure 2.9 shows an instance of different simplified approximations to an original model using this algorithm.

2.2.4 Geometry compression artifacts

As we mentioned before, transmission and storage of digital data involving 2D and 3D graphics are required in every domain related to computer graphics. Thus, geometry compression was proposed as a solution to the issues about the mesh size reduction for the storage and

Figure 2.9: Different simplified Bunny models using quadric edge collapse decimation for different target triangles. From left to right: Original Bunny model with 69,451 triangles; A simplified Bunny model with 1,000 triangles; A simplified Bunny model with 100 triangles. The image is from [37].

transmission. In other cases, geometry compression is adopted for interactive visualization of large meshes on low capability terminals as well [74]. In the first step of the compression, the algorithm usually quantizes all the coordinates of vertices (so-called quantization), which can effectively reduce the amount of data. Even though the quantization enables the encoding without any quality loss, aggressively quantizing the geometry can introduce the loss of information, which can be observed as a blocking or deformed surface [48] (see Figure 2.10). The quantization encoder is usually of two types: scalar quantization and vector quantization.

Scalar quantization

Scalar quantization transforms the vertices of a mesh from floating-point positions into integer number positions. A commonly used compression algorithm, namely Gotsman (TG) compression [74], encodes the mesh vertex coordinates using uniform quantization (a kind of scalar quantization). In details, assume that all the vertices of a mesh are uniformly quantized to n bits meaning each coordinate of any vertex is rounded to the nearest integer value in $[0,2^n-1]$, and accordingly the mesh bounding box is equally divided into a 3D grid, then the coordinates of those vertices v_j that lie inside ith 3D grid cell are removed and replaced by the integer coordinates of the cell center v_i [116]:

$$v_i = round_{j\in k}[v_j] \tag{2.6}$$

where k refers to the number of the vertices that lie inside ith 3D grid cell.

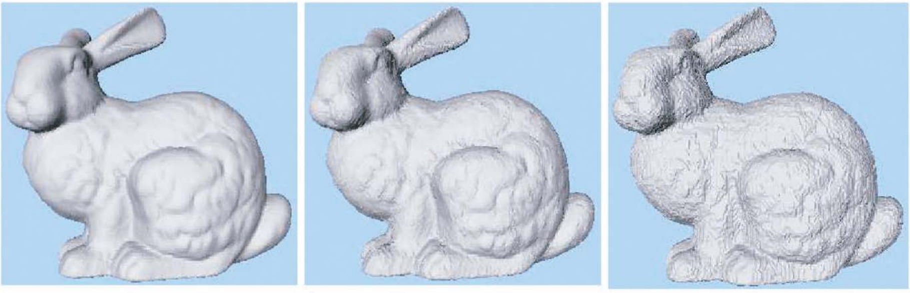


Figure 2.10: Vector quantization vs. scalar quantization. From left to right: Original Bunny model: 96 bits per vertex (bpv); Vector quantization method proposed in [18] (7 bpv); Scalar quantization (7 bpv). The figure is captured from [18].

Vector quantization

Vector quantization partitions vertices of a mesh into arbitrary polygonal cells [92], In contrast with the scalar quantization whose partitioned cells are cuboids, cells of vector quantization can be better fitted to the mesh shape, and thus contribute to a better visual quality than scalar quantization at low quantization bits [64] [18].

2.2.5 Watermarking artifacts

The watermarking technique emerges as a solution for protecting copyrights of 3D digital designs from illegal duplication and redistribution. The algorithms embed a watermark, which is usually a bit string or a pseudo-random sequence, over a 3D digital asset, and thus inevitably introduces distortions after the process. Three recent *blind* watermarking schemes, in which the original cover content (i.e. watermark) is *not* required for watermark extraction, are presented below:

Watermarking based on vertex norms distribution[17]

The method modifies (increases or decreases) the distributions of the mean value or variance value of the vertex norms according to a watermark bit to be inserted. And it usually introduces structured ringing distortions on a mesh surface (see Figure 2.11 and Figure 2.12).

Watermarking based on manifold harmonics [124]

This method opts for a spectrum decomposition tool, namely the manifold harmonics transform [117], which can generalize the conventional Fourier analysis to the functions defined

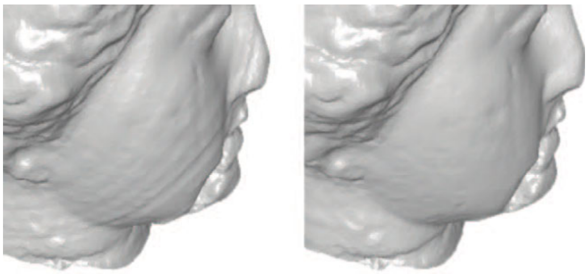


Figure 2.11: A comparison between watermarking methods based on vertex norms distribution [17] and based on manifold harmonics [124]. From left to right: A Venus model embeded a watermark by Cho's method; A Venus model embeded a same watermark by Wang's method using manifold harmonics. This figure is adapted from [124].

on arbitrary 2-manifold surfaces. Based on the manifold harmonics transform, the watermarking algorithm quantizes the amplitudes of low frequency coefficients using scalar Costa scheme [29] to covertly embed a blind and robust watermark in the manifold harmonics spectral domain of a 3D mesh. Figure 2.11 shows a comparison between this algorithm and Cho's method [17]. Apparently, this algorithm introduces fewer visual artifacts.

Watermarking based on volume moments [123]

This scheme introduces the geometric volume moments. The volume moment m_{pqr} of order p, q, r is defined as a volume integral of the function $f(x,y,z) = x^p y^q z^r$ inside the closed surface S, which can be expressed as:

$$m_{pqr} = \iiint x^p y^q z^r \rho(x,y,z)\, dx\, dy\, dz \tag{2.7}$$

where $\rho(x,y,z)$ is the volume indicator function, and it is equal to 1 if (x,y,z) is inside the closed surface; otherwise it is equal to 0. Based on the notion of volume moment, the algorithm firstly normalizes a mesh with its global volume moments, and then discretizes its cylindrical domain to decompose the mesh into patches. In each candidate patch, the algorithm uses a modified scalar Costa [29] quantization of local zero-order moment to insert an one-bit watermark. After the insertion, a smooth deformation mask is used to hide the visible distortion. Figure 2.12 presents a comparison between this method and Cho's algorithm [17]. This method hides a watermark better.

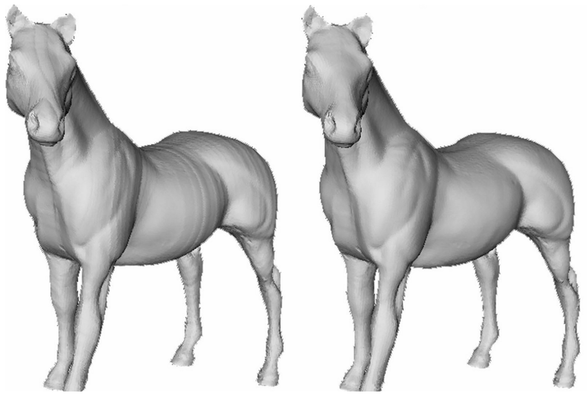


Figure 2.12: A comparison between watermarking methods based on vertex norms distribution [17] and based on volume moments [123]. From left to right: A Horse model embeded a watermark by Cho's method; A Horse model embeded a same watermark by Wang' method using volume moments. This graphic is supplied by [123].

2.3 Conclusion

Through the summary above, we can see that, besides the noises resulted from acquisitions, different types of converting and processing operations on 2D/3D contents alter or modify properties of original digital data on aspects of gray scale, color, hue, intensity, structure, topology, etc., which more or less introduces artifacts (e.g., blocking, blurring, losses of details, shrinkage, deformation, etc.) as well. Some of the artifacts are almost imperceptible for human vision system, whereas others may be very noticeable or even annoying, and bring about unwanted problems and issues in many human centered applications.

Hence, there emerged an increasing demand of evaluating the quality of processed graphics, which aim to evaluate the impacts of the artifacts and thereby improve the acquisition procedure, converting and processing algorithms. Thus, many researchers proposed various objective metrics to measure the global or local quality of processed 2D/3D graphic data. Meanwhile, a great deal of subjective quality assessments, where subjective opinions on evaluating the global or local quality of the processed graphics from a group of human observers are collected and analyzed, was designed to quantitatively measure and compare the performances of the objective metrics in correlation with subjective opinions, and moreover some objective visual quality metrics can be driven by the collected subjective data [126]. Nevertheless, the subjective quality assessments can be very expensive to conduct, since massive subjects and time are involved in the experiments [85]. In the following chapter, several classical and recent objective visual quality assessments of 2D image and 3D mesh will be introduced in details.

Chapter 3

Survey on objective visual quality metrics for 2D images, videos and 3D models

3.1 Introduction

In previous chapter, we described the different types of digital artifacts caused by acquisition, convertion and other processing operations. Since the human visual system (HVS) is the ultimate receiver of the processed graphics, some of the artifacts are imperceptible and of no account, but others may bring about undesired outcomes on the final appearance, and even fatal problems in some applied scientific fields, such like medical imaging, etc. Thus, accompanying the development of computer graphic technologies, diverse objective and subjective visual quality assessments (methodologies) have been proposed to measure artifacts, evaluate processing procedures, and optimize processing methods in return. In computer graphics realm, visual quality (visual fidelity) refers to the visual impacts of the artifacts coming from digital graphic manipulations (e.g., rendering, compression, simplification, etc) [61]. Based on mathematical methods, objective quality assessment evaluates the visual fidelity of processed graphic data using automatic metrics, while subjective quality assessment, where a group of observers provide subjective opinions on estimating processed graphic contents, permits to evaluate the performance of objective assessments in terms of correlation with the subjective opinions. The outputs of an objective metric or a subjective experiment is usually a single global score (e.g., Mean opinion score), which represents a global visual impact of artifacts. However, it can also be a local distortion visibility, which predicts spatial distributions of artifacts. Both subjective and objective quality assessing procedures can be one of the three forms: full-reference (FR), reduced-reference (RR), in which an original graphic (image, video or 3D model) is completely or partially shown for the comparison with its distorted version, and no-reference (NF) where only distorted graphics are available. This chapter is dedicated to present different objective visual quality metrics for 2D image, including video sequence, and 3D mesh with full-reference (FR) setting.

3.2 Image and video visual quality assessments (IQA)

3.2.1 Metrics of first generation

Mean square error (MSE) and Peak signal-to-noise ratio (PSNR) are pioneering objective quality metrics, both of which compute the differences in the statistical distributions of pixel values (y_i and x_i) between degraded image and its reference (original image). In details, assume that X is an original image without any processing, and Y is a distorted approximation to be evaluated. Denote N is the total number of the pixels in each image. The MSE and PSNR can be defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{3.1}$$

And the PSNR is defined as:

$$PSNR = 10 \cdot log_{10}\frac{MAX^2}{MSE} \tag{3.2}$$

MAX refers to the maximum pixel value of an image. When the pixels are allocated with 8 bits/pixel, $MAX = 2^8 - 1 = 255$. Furthermore, a generalization of the MSE can be deducted using Minkowski metric, as below:

$$E_p = (\sum_{i=1}^{N}|x_i - y_i|^p)^{\frac{1}{p}} \tag{3.3}$$

where $p \in [1,\infty)$. When $p = 1$, E_p means mean absolute error measure, $p = 2$ makes E_p square root of MSE (RMSE), $p = 3$ yields cube root of MSE, while $p = \infty$ calls the maximum absolute difference measure:

$$E_\infty = \max_i |x_i - y_i| \tag{3.4}$$

MSE and PSNR metrics are extensively used in image processing fields thanks to their simplicity and convenience for the optimization of processing algorithms. Moreover, the MSE and PSNR have clear physical meanings: the energy error of pixel intensities between a distorted image and its reference. However, the major defect of the MSE and PSNR is their poor correlation with human perception. Figure 3.1 shows a visual example of this defect, where the original Einstein image (a) is processed by several different operations: contrast enhance (b), Gaussian noise (c), Fat-tail distributed noise (d), speckle noise (e), JPEG compression(f), blurring (g). All the measured qualities of these distorted images from MSE are 285.15, and their PSNR measures equal to 23.58, however, obviously, their perceived qualities are dramatically different. For the last three decades, many researchers have launched extensive experiments and studies to criticize and analyze the weakness of the MSE and PSNR metrics [39, 28].

Later on, a great deal of researchers has devoted to design visual quality metrics, which can predict the perceptual annoyance caused by artifacts. The research on the metrics successively evolved in two directions: *bottom-up approaches* and *top-down approaches*.

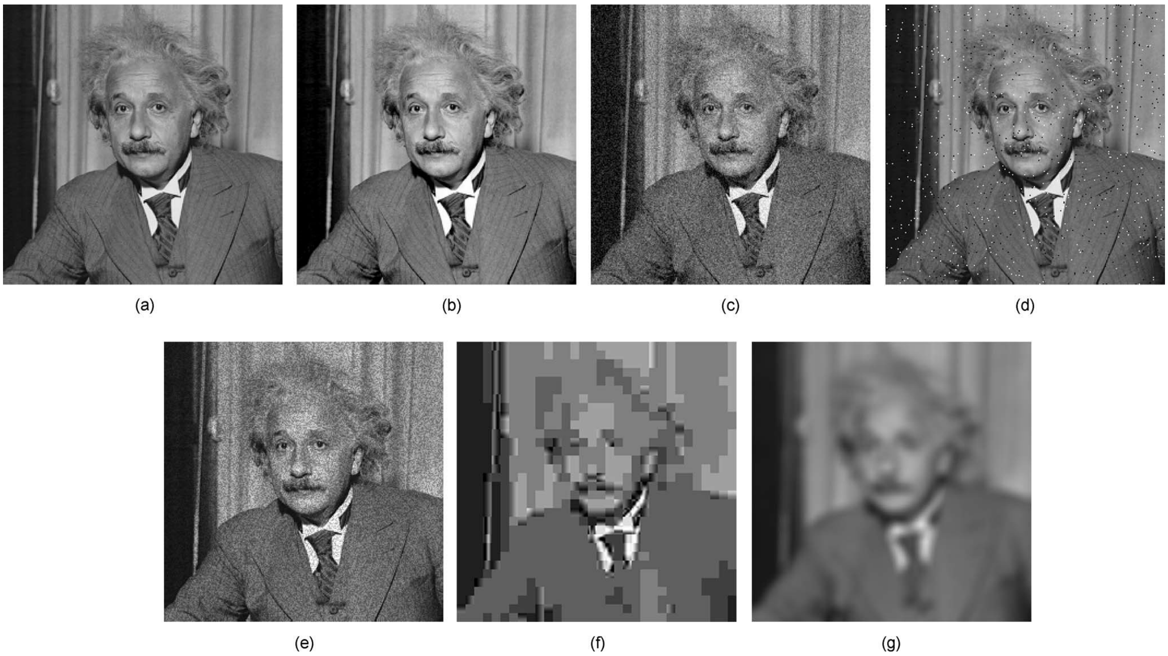


Figure 3.1: A original Einstein image (a) and different processed images: contrast enhance (b), Gaussian noise (c), Fat-tail distributed noise (d), speckle noise (e), JPEG compression (f), blurring (g). ALL the measured MSEs between original image and a distorted versions are 285.15, and PSNR equal to 23.58. The image is captured from [69].

3.2.2 Metrics based on bottom-up approach

The image visual quality metrics based on bottom-up approach try to approximate the low-level functionalities of the *human visual system* (HVS). The most common approach is to build some basic blocks that simulate the relevant psychophysical components and features, and ultimately integrate all the blocks into a mathematical function, which can behave the same way as HVS components.

To establish the function, the approach usually gets through 5 stages: (1) Pre-processing, which involves image alignment, point-wise nonlinear transform, etc; (2) *Contrast sensitivity function* (CSF) filtering (see Figure 3.3) is a low-pass filter that simulates visual system components (e.g., eye optics, retina, striate cortex, etc.) and transforms color space information; (3) Channel decomposition, which transforms the input signals into different selective channels (subbands) of spatial frequency and orientation; (4) Error normalization, which is used to normalize the signal error in each channel by merging the variation of visual sensitivity in different channels, and the variation of visual error sensitivity resulted from intra- or inter-channel neighboring transform coefficients, (5) Error pooling integrates the signal errors from different channels into a global quality or distortion score [127]. Figure 3.2 illustrates the framework of these image quality metrics. Generally, a *predictor* can provide the local visual detectability of near-threshold distortions (i.e. distortion map), whereas a *metric* generates a single non-negative number, usually ranging from 0 to 1, indicating the global quality of a processed graphic, and furthermore the local measurements of many metrics can provide distortion maps, which be considered as predictors as well. Several traditional image visual quality metrics based on bottom-up perceptual approaches are described below:

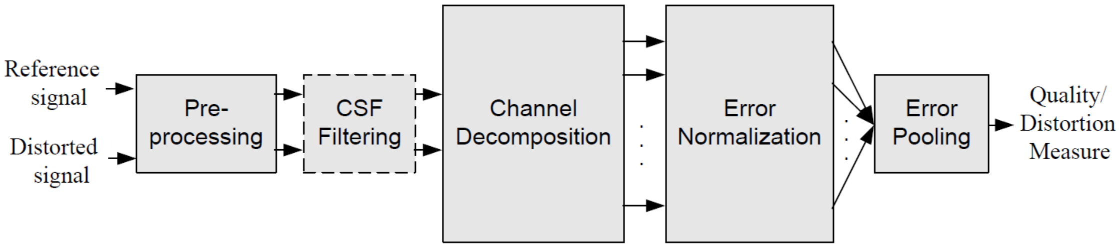


Figure 3.2: Diagram of bottom-up visual quality assessment approaches. Note that the CSF feature can be implemented either as a separate stage (as shown) or within “Error Normalization” . The diagram originates from [127].

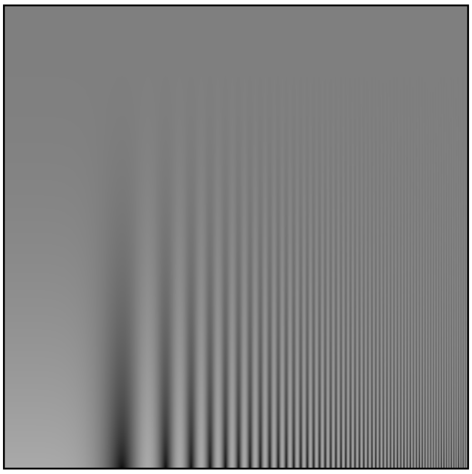


Figure 3.3: An illustration of contrast sensitivity function (CSF) using sinusoid contrast pattern at various spatial frequencies. The contrast decreases linearly from bottom to top and uniformly over all frequencies, which increase from left to right. The perceived contrast, however, peaks near the middle on the horizontal axis, near two cycles per degree. The image originates from Bolin and Meyer [7].

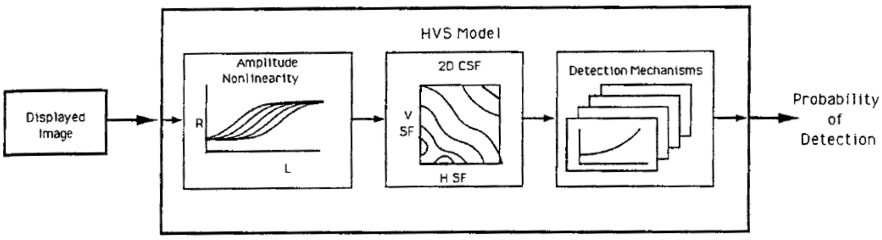


Figure 3.4: A block diagram illustrating the three main components of HVS model designed in the VDP [25].

The visible differences predictor (VDP) [25]

VDP method or Daly model mimics the lower-order processing of the HVS, such as the optics, retina, lateral geniculate nucleus, and striate cortex, by integrating three components (i.e. amplitude nonlinearity, contrast sensitivity function (CSF), and detection mechanism) into a HVS model (see Figure 3.4). The VDP does not provide an absolute image quality but instead focuses on describing the visibility of distortions between a distorted image and its reference. The output of the VDP is a map showing the probability of detecting the differences between two images. Figure 3.5 presents an example of VDP output. The map of the *detection probablity* is able to provide a description of the threshold behavior of vision but do not discriminate among different suprathreshold errors. Hence, the VDP can be summarized as a threshold model for suprathreshold imagery, and capable of quantifying the important interactions among the threshold differences and suprathreshold of image content and structure [25].

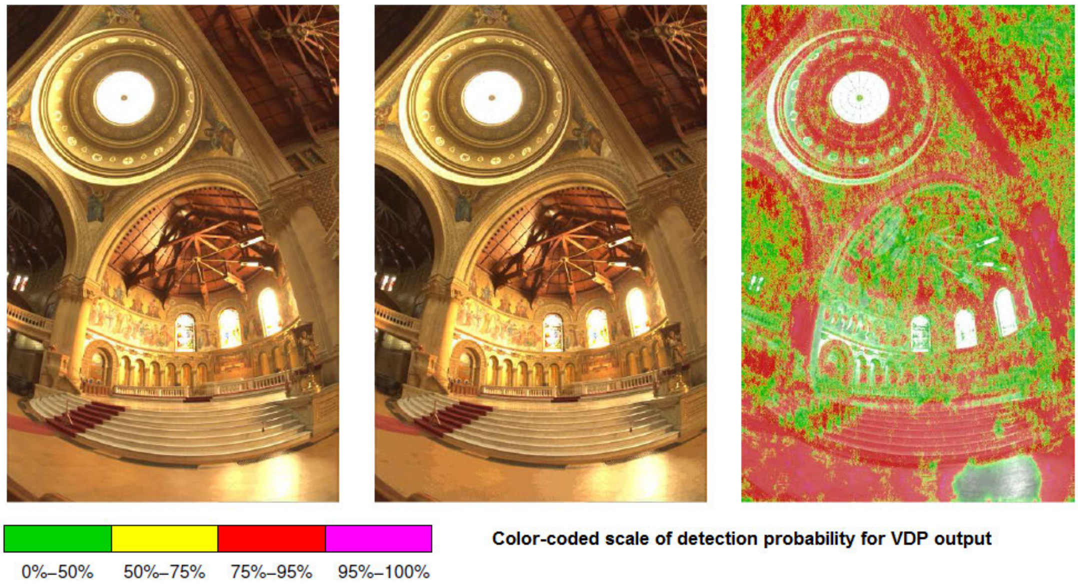


Figure 3.5: An example of VDP output. from left to right: Reference image; Distorted image proccessed by quantization; Detection probability map output by VDP. The figure originates in [76].

Sarnoff JND (just-noticeable difference) model [73]

This model uses point spread function (PSF) as a simulation of eye optics to filter image, and then decomposes the filtered image using Laplacian pyramid into seven resolutions, followed by band-limited contrast calculations. The *visual masking effect* was firstly considered in this work. This effect expresses itself in a decreased ability to distinguish contrast patterns at certain frequencies in the presence of similarly oriented patterns of nearly equal frequencies (see Figure 3.6), which is particularly significant when an image or a mesh is mapped by texture. The model uses a transducer (i.e. sigmoid nonlinearity) to take masking effect into consideration. The transducer makes the discrimination of a contrast increment between one image and the other depend on the contrast response that is common to both images. Similar to the VDP, the Sarnoff JND produces a probability-of-detection map showing the noticeable differences at each spatial location between two input images.

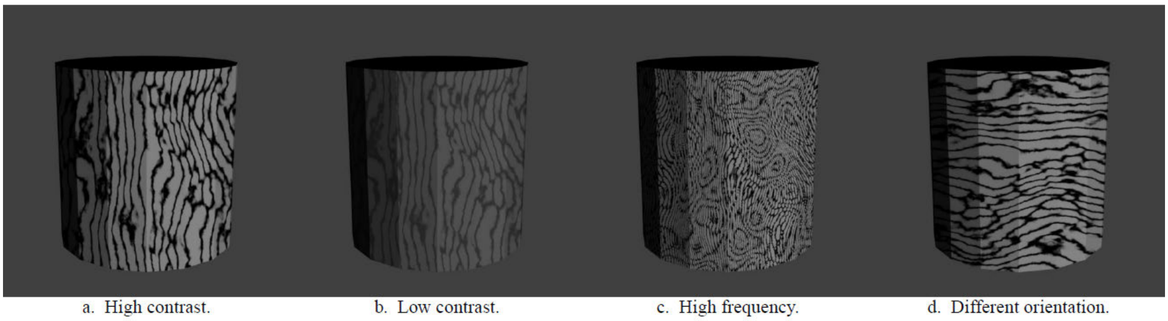

a. High contrast. b. Low contrast. c. High frequency. d. Different orientation.

Figure 3.6: A demonstration of visual masking effect. The edges of the mesh are less perceived when the texture contrast is high and the texture frequency and orientation are nearly same to the pattern of edges. The example is supplied by [34].

Watson's DCT (Discrete Cosine Transform) model [130]

This method was designed for the perceptual quality evaluation of quantization noise from JPEG compressions. It first divides the image into distinct blocks, and a visibility threshold is calculated for each coefficient in each block. Three factors: (1) Baseline contrast sensitivity, (2) Luminance masking, (3) Contrast/texture masking, are involved to determine the visibility threshold. It produces a perceptually measured distortion map and a single distortion value after the frequency error pooling.

Watson's Wavelet model [131]

This measure was proposed to detect the visibility of discrete wavelet transform (DWT) quantization noise. Following inverse wavelet transform, the noise visibility threshold in each spatial subband is measured by subjective experimentation at fixed viewing distances. The visibility threshold for that subband is then determined as the reciprocal of the corresponding visual sensitivity. The combination of DWT noises detection thresholds allows a computation of a perceptually detectable quantization matrix for which all errors are in theory above the visual thresholds.

Several recent metrics based on these bottom-up models were designed to simulate the sensitivity of HVS components. Referring to the Sarnoff JND model, NQM [26] metric utilizes Peli's contrast pyramid [84] to decompose a degraded image and its reference, considering: (1) the variation in contrast sensitivity with distance, image dimensions, spatial frequency; (2) the variation in the local luminance mean; (3) contrast interaction between spatial frequencies; (4) contrast masking effects. Based on the Watson's Wavelet Model, Chandler et al. proposed the visual signal-to-noise ratio (VSNR) [12] metric to quantify the visual quality of images. It takes account for the low-level and mid-level HVS properties of contrast sensitivity and visual masking to determine the threshold of visual detection, and then uses a wavelet-based model to judge whether the distortions are below the threshold. However, the fact that the underlying goal of bottom-up models is to establish systems, which mimic the functionalities of HVS, brings about inevitable defects and potential problems, such as an obscure relationship between visibility error and quality degradation, unreasonable normalizations of visibility thresholds between different channels [126], restricted and simplistic generalizations of HVS simulation [137], and so forth.

3.2.3 Metrics based on top-down approach

A milestone of the development for image visual quality assessments for full-reference (FR) comparison was the top-down model based on Structural SIMilarity (SSIM) [127], which significantly extended the horizons of the studies of HVS mechanisms on evaluating images. While the Bottom-up models attempt to simulate the HVS components, top-down approaches consider the HVS systems as a black box, and try to build a model connecting input-output relationship, where the input(s) are the images to be evaluated, and the output(s) are the subjective elevation opinions given by observers. The following sections detail the structural similarity and some recent and relevant approaches.

Metrics based on Structural similarity (SSIM)

The structural similarity quality measure is based on a top-down assumption that the HVS is highly adapted for extracting structural information from a viewing field, and that a measure of structural information errors should give a good prediction of perceived image distortions [127]. It incorporates three structural attributes in an image: luminance, contrast, and structure. The detailed definitions and computations are as follows:
Let $x = \{x_i | i = 1,2,3,\dots,N\}$ and $y = \{y_i | i = 1,2,3,\dots,N\}$ represent the two image patches, which have been aligned with each other (e.g., spatial patches extracted from each image). Let μ_x, μ_y, σ_x^2, σ_y^2 and σ_{xy} denote the mean values and the variances of x and y, and the covariance of x and y respectively. The comparison functions of luminance $l(x,y)$, contrast $c(x,y)$, and structure $s(x,y)$ are computed as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{3.5}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3.6}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3.7}$$

Where C_1, C_2, C_3 are constant values computed by: $C_1 = (K_1L)^2, C_2 = (K_2L)^2$ and $C_3 = C_2/2$, in which L is the dynamic range of the pixel values ($L = 255$ for 8 bits/pixel gray scale images), and $K_1 << 1$ and $K_2 << 1$ are two scalar constants. The local SSIM index between x and y are defined as:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{3.8}$$

According to [127], $\alpha = \beta = \gamma = 1$, and the resulting SSIM index is deduced as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.9}$$

The SSIM index is applied by sliding window approach, where the index is computed in an 8×8 moving window pixel-by-pixel. In [127], results show that SSIM method outperformed between traditional metrics (e.g., PSNR) and previous bottom-up approaches (e.g., Sarnoff JND (just-noticeable difference) model). However, the SSIM is a single-scale method, which may be incompetent to the image details at different resolutions. Moreover, simply taking the mean of SSIM indexes to obtain a global quality score is lack of qualitative sensible and quantitative manageable principles. Improved metrics: MS-SSIM (Multi-scale structural similarity) [129] and IW-SSIM (Information content weighting structural similarity) [128] were therefore designed for resolving the problems.

Extended from SSIM, MS-SSIM [129] method iteratively applies a low-pass filter and downsamples the filtered image by a factor of 2, indexing the original image pathes as Scale 1 and the highest scale as Scale M. While contrast and structure functions are computed at each scale, the luminance function $l(x,y)$ is computed only on the smallest scale (M). The final local $MS-SSIM(x,y)$ index is obtained after $M-1$ iterations, as described below:

$$MS-SSIM(x,y) = [l(x,y)]^{\alpha_M} \cdot \prod_{j=1}^{M} [c(x,y)]^{\beta_j} \cdot [s(x,y)]^{\gamma_j} \tag{3.10}$$

with j the index of the scale, $\alpha_j = \beta_j = \gamma_j$, $\sum_{j=1}^{M} \beta_j = 1$. The results from [129] show that MS-SSIM could produce better results than its single scale counterpart SSIM regarding the correlation with visual perception. Figure 3.7 shows an example of local measurement maps of MS-SSIM and SSIM.

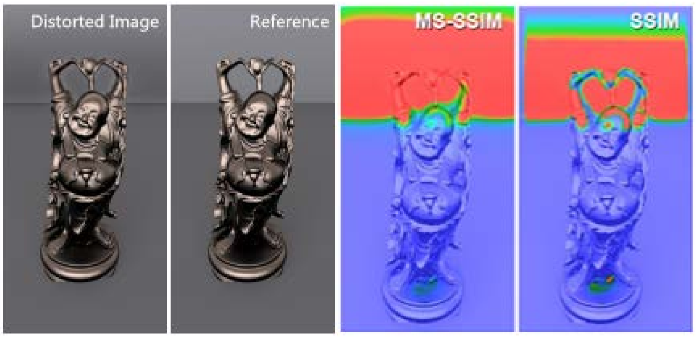


Figure 3.7: An example of prediction maps of MS-SSIM and SSIM local measurements. from left to right: Distorted image with *gradient-based tone mapping artifacts* [33]; the original image; Predicted disortion maps of MS-SSIM and SSIM , in which the distortion probabilities visualized from red (1) to blue (0). The example is adapted from [9].

Soon afterwards, Wang and Li [128] improved the original MS-SSIM to IW-SSIM by considering an information-content weighting (IW) based quality score pooling strategy. Precisely, in the i-th local window at the j-th scale, the information content weight $\omega_{j,i}$ is introduced in the computation, defined by:
for $j = 1,\ldots,M-1$, we have:

$$IW-SSIM_j(x,y) = \frac{\sum_i \omega_{j,i} c(x_{j,i},y_{j,i}) s(x_{j,i},y_{j,i})}{\sum \omega_{j,i}} \tag{3.11}$$

when $j = M$ and,

$$IW-SSIM_j(x,y) = \frac{1}{N_j}\sum_i l(x_{j,i},y_{j,i}) c(x_{j,i},y_{j,i}) s(x_{j,i},y_{j,i}) \tag{3.12}$$

with N_j the number of local windows in the j-th scale, and $\omega_{j,i}$ is derived by modeling the distortion channel and the perceptual channel, and by taking the mutual information between the images into account. Finally, the overall $IW-SSIM$ be expressed as follows:

$$IW-SSIM = \prod_{j=1}^{M} (IW-SSIM_j)^{\beta_j} \tag{3.13}$$

where β_j the values are obtained through psychophysical measurement [129]. Via the extensive tests with six publicly-available independent image databases, IW-SSIM algorithm performs the best compared to SSIM and MS-SSIM for the task of predicting perceived image quality.

Metrics based on information fidelity criterion (IFC)

The philosophical theory of information fidelity criterion (IFC) [101] hypotheses that images and videos of the three dimensional visual environment come from a class of natural scenes, and that most real-world distortion processes disturb the statistics of natural scenes and make the image or video signals unnatural. Based on the Natural Scene Statistics (NSS), the natural scene models in conjunction with distortion models predict the mutual information shared between distorted images and their references.

Based on this hypothesis of IFC, the random field (RF) in a wavelet decomposition subband of an image: RF_d can be described as:

$$RF_d = G \cdot RF_r + V \tag{3.14}$$

where RF_r is the random field of the subband from the reference image, G represents a deterministic scalar attenuation field, and V is a stationary additive zero-mean Gaussian noise random field.

The IFC quantifies the statistical information that is shared between the reference and the distorted images. Regarding the correlation with visual perception, the results from [101, 136] indicate that the IFC is superior to the classical metric (i.e. PSNR), Sarnoff JND metric, and the SSIM, but inferior to the MS-SSIM and the IW-SSIM.

Later on, Sheikh and Bovik [100] extended IFC to the visual information fidelity (VIF) by considering HVS model (see Figure 3.8). The VIF does not only quantify the information shared between a reference and a distorted image, but also provide how much of this reference information can be extracted from the distorted image. The experimental results in [100] show the advantages of VIF measure compared to PSNR, Sarnoff JND metric, SSIM and IFC. However, VIF cannot challenge other advanced visual quality metrics, such like MS-SSIM and IW-SSIM, via the comparisons of their performences correlated to the subjective evaluations in [128, 136].

3.2.4 Data-driven metrics

In recent years, several researchers exploited ways to build reliable full-reference image quality metrics driven by subjective data. One remarkable metric proposed by Čadík et al.

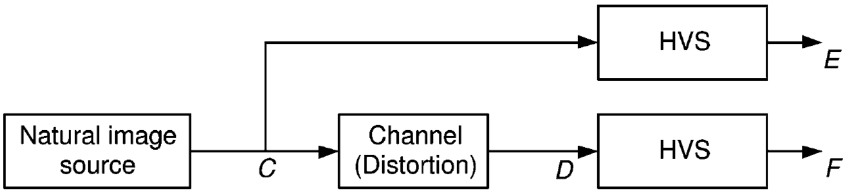


Figure 3.8: The theoretic framework of VIF : Mutual information between C and E quantifies the information that the HVS system could ideally extract from the reference image, whereas the mutual information between C and F quantifies the corresponding information that could be extracted from the test image.

[10] measures 32 defined image difference features between two images, some of which are described by a single value, some by up to 62 dimensions. The features range from a simple absolute difference value to visual attention, and include the predictions of several major image quality metrics (e.g., SSIM, VDP model, etc.). Via *machine learning*, the metric is trained by 37 images with manually labeled distortion maps. The features are optimally selected and incorporated with ensembles of bagged decision trees (decision forest) used for classification. Trough the leave-one-out cross-validation procedure, the classification (trained metric) is verified to outperform between all general purposes metrics. Figure 3.9 shows several predictions (distortion maps) produced by the trained metric and other metrics.

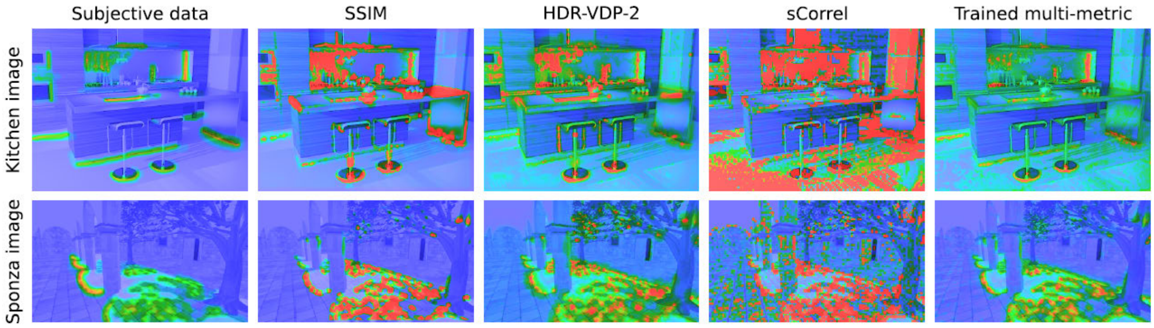


Figure 3.9: Comparison of distortion maps predicted by the trained metric and other metrics. from left to right: manually marked distortions in computer graphics and the predictions of image quality metrics: SSIM, HDR-VDP-2, sCorrel and the trained metric. The trained metric uses the predictions of the existing metrics as features for a decision forest classifier. It is trained to predict the subjective data. The all the distortion maps are captured from [10].

3.2.5 Video visual quality metrics

As we discussed previously, videos are essentially sequences of 2D images. Thus the video visual quality can be measured by image visual quality metrics (e.g., PSNR, VSNR, SSIM, MS-SSIM, IW-SSIM, IFC, VIF, etc) frame by frame on the luminance component of the video, meanwhile, some necessary pre-processes, i.e. spatial/temporal alignment, calculation of processed valid region (abbreviated as PVR and defined as the portion of the processed video image which contains valid picture information), and gain & level offset calibration (i.e. find the reference image that minimizes the standard deviation of each difference image between the reference and processed frames using the sub-sampled Y luminance frames), should be executed to guarantee that measurement of each frame between a processed video and its reference is accurate and efficient. Finally, a visual quality score for the video can be obtained by pooling (e.g., taking the average, the median, the maximum, the minimum, etc) the measured values over all the frames together. The performances of the measures highly depend upon their embedded image visual quality metrics [99].
On the other hand, by considering the specific properties of video, such like attributes of temporal changes in a video sequence, etc, Pinson et al. [85] proposed a more perceptual correlated but complex, video quality metric (VQM), which calculates the perceptual changes in spatial, temporal and chrominance properties from spatial-temporal sub-regions of video streams, and then pools them into a single quality score. The Figure 3.10 shows the diagram of VQM procedure. They firstly extracted the seven quality features (quality parameters) from a spatial-temporal sub-region of a video stream (both original and processed), which are:
(1)Parameter "*si_loss*": it detects a decrease or loss of spatial information;
(2)Parameter "*hv_loss*": it functions as a detection of a shift of edges from horizontal & vertical orientation to diagonal orientation;
(3) Parameter "*hv_gain*": the parameter measures a shift of edges from diagonal to horizontal & vertical;
(4) Parameter "*chroma_spread*": This parameter detects changes in the spread of the distribution of two-dimensional color samples;
(5) Parameter "*si_gain*": it assesses improvements to quality that result from edge sharpening or enhancements;
(6) Parameter "*ct_ati_gain*": this feature is used for measuring the amount of spatial detail, and a temporal information feature, measuring the amount of motion present in the spatial/temporal region;
(7) Parameter "*chroma_extreme*": This measurement is for the detection of several localized

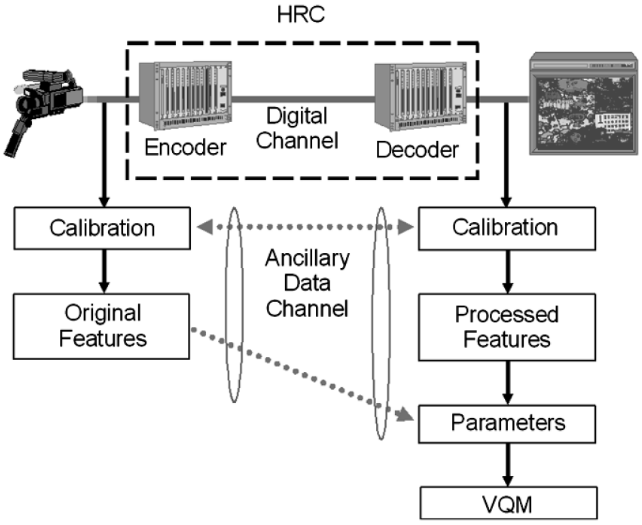


Figure 3.10: The diagram of VQM procedure. The calibration of the original and processed video streams includes spatial alignment, valid region estimation, gain & level offset calculation, and temporal alignment. VQM calculation involves extracting perception-based features, computing video quality parameters, and combining parameters to construct the General Model. This figure is adapted from [85] .

color impairments.

The VQM model is then a linear combination of the seven parameters, whose weights were determined by 1536 subjectively rated video sequences.

$$\begin{aligned} VQM = & -02097 \cdot si\_loss \\ & + 0.5969 \cdot hv\_loss \\ & + 0.2483 \cdot hv\_gain \\ & + 0.0192 \cdot chroma\_spread \\ & - 2.3416 \cdot si\_gain \\ & + 0.0431 \cdot ct\_ati\_gain \\ & + 0.0076 \cdot chroma\_extreme \end{aligned} \tag{3.15}$$

In [85], the Pearson linear correlation between the VQM indexes and Difference Mean Opinion Scores (DMOS) for the testing videos was up to 0.91.

A recent research on video visual quality assessment is MOtion-based Video Integrity Evaluation (MOVIE) index [98], which is an integration of both spatial and temporal measurements of distortions on the reference and the processed videos. For the spatial distortion measurements, the Spatial MOVIE index uses outputs of the spatial-temporal Gabor filters to accomplish the Structural SIMilarity (SSIM) index and the information fidelity criterion (IFC). On the aspect of temporal distortion measurements, the temporal MOVIE index computes motion information from the reference video explicitly, and measures the quality of the processed video along the motion trajectories of the reference video. At last, the MOVIE index is a multiplication of the Spatial MOVIE index and the Temporal MOVIE index. The MOVIE index takes the measures of both spatial and temporal motion distortions into account, and thus outperforms among the counterparts (e.g., PSNR, VSNR, SSIM, MS-SSIM, and VIF based video quality metrics, VQM, etc) in terms of correlation with human judgments [98, 99].

3.3 3D mesh visual quality (MVQ) metrics

As we mentioned in previous chapter, accompanying the developments of 2D image technologies, the advances of computer graphics boost extensive uses of 3D data in many applied domains, like virtual reality (VR), e-commerce, digital entrainments, scientific visualization, etc. The most commonly used 3D data is represented in the form of polygonal meshes.

Similar to 2D image visual quality assessment, diverse 3D visual quality metrics were designed to evaluate/boost all the processing operations on 3D objects (e.g. simplification, compression, watermarking, etc). These 3D visual quality measurements can be mainly classified into two types: (1) model-based metrics, which extract the geometric attributes (e.g., geometric positions, curvatures, dihedral angles, etc.) as basic quality features; (2) image-based metrics, inherited from image/video visual quality metrics, which measure the screen-space distance (or similarity) of the visualized 3D data that are thus represented as images or video sequence (e.g., snapshots of a mesh from different views) [24, 61].

3.3.1 Model-based metrics

Vertex-position based metrics

The first 3D fidelity metrics were based on geometric position distances. In the past decades, the most extensively used metric was the Hausdorff distance, which was firstly designed for evaluating simplification algorithms [96]. The Hausdorff distance is calculated as follows:

Denote $e(p,A)$ represent the minimum distance from a 3D spatial point p to a 3D mesh A:

$$e(p,A) = \min_{v_i^A \in A} d(p, v_i^A) \tag{3.16}$$

in which d is the Euclidian distance, and v_i^A means the i^{th} vertex of A. The asymmetric Hausdorff distance between two meshes A and B is computed as:

$$H_a(A,B) = \max_{v_i^A \in A} e(v_i^A, B) \tag{3.17}$$

Then, the symmetric Hausdorff distance is:

$$H(A,B) = max\{H_a(A,B), H_a(B,A)\} \tag{3.18}$$

Moreover, the ensembles of $e(v_i^A, B)$ can be considered as local predictions of the differences between two meshes A and B. The Figure 3.11 shows an visual example of the local predictions of watermarking distortion measured by $e(v_i^A, B)$.

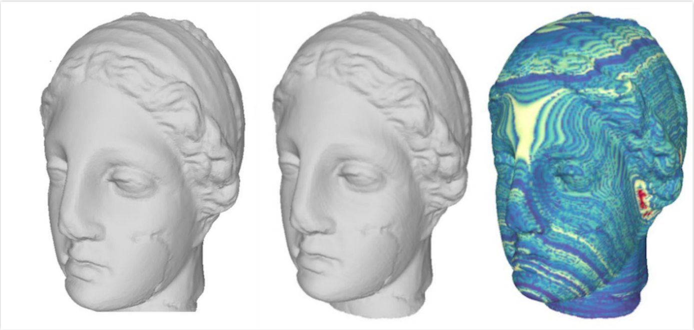


Figure 3.11: From left to right: The Venus model; a distorted version after a watermarking algorithm by [124] ; Distortion map produced by the local measures of Hausdorff distance. Warmer colors represent higher measured distance values.

Referring to the pioneering metrics on 2D image quality, the root mean square (RMS) error is to pool all the distances of corresponding vertices between two meshes:

$$RMS(A,B) = (\sum_{i=1}^{n} ||v_i^A - v_i^B||^2)^{1/2} \tag{3.19}$$

where n means the number of vertices on two meshes respectively, v_i^B of B is the corresponding vertex to the vertex v_i^A on mesh A. This definition limits the comparison between two meshes having the same connectivity. Hence, another mean square error (MSE) was designed based on Hausdorff distance, of which the asymmetric MSE is defined as:

$$MSE_a(A,B) = \sum_{i=1}^{N_A} e(v_i^A, B)^2 \tag{3.20}$$

where N_A is the number of vertices on mesh A. And after the computations of $MSE_a(A,B)$ and $MSE_a(B,A)$, the Maximum Root Mean Square(MRMS) can be derived as:

$$MRMS(A,B) = max\{\sqrt{MSE_a(A,B)}, \sqrt{MSE_a(B,A)}\} \tag{3.21}$$

Another metric based on the geometric Laplacian operator [48] was introduced to perceptually evaluate compression algorithms. The geometric Laplacian can reflect a degree of smoothness of the surface by considering the difference vector $u(\vec{v})$ between a given vertex v and its new position after a Laplacian smoothing step. It can be described as:

$$u(\vec{v}) = v - \frac{\sum_{i \in n(v)} l_i^{-1} v_i}{\sum_{i \in n(v)} l_i^{-1}} \tag{3.22}$$

where $n(v)$ is the number of neighboring vertices' indices of v, l_i refers to the Euclidean distance from v to its neighbor v_i. Based on the geometric Laplacian operator, a mesh visual quality metric GL_1 is derived based on the mean geometric distance between corresponding vertices of two meshes (A and B) and the mean distance of their geometric Laplacian values:

$$GL_1(A,B) = \alpha RMS(A,B) + (1-\alpha)(\sum_{i=n}^{n} ||u(\vec{v_i^A}) - u(\vec{v_i^B})||^2)^{1/2} \tag{3.23}$$

α in [48] is determined as 0.5, while in the work [105], α equals to 0.15, which leads to a different metric GL_2.

Curvature-based metrics

Curvature is an important geometric attribute, which reflects a curving degree of a line or a surface. Many mesh visual quality metrics were developed based on two principal curvatures: minimum curvature K_1 and maximum curvature K_2 (i.e. the eigenvalues of the shape operator at a vertex v on a surface) and their derivatives: Gaussian curvature ($K_1 \times K_2$), Mean curvature ($(K_1 + K_2)/2$), Shape index ($2/\pi * arctan[(K_1 + K_2)/(K_1 - K_2)]$), Curvedness ($\sqrt{(K_1^2 + K_2^2)/2}$) [6, 57], etc. Figure 3.12 and Figure 3.13 present several examples of colorized curvatures of a Dinosaur object from *LIRIS/EPFL General-Purpose database*. A high-level attribute, mesh saliency [63] (see Figure 3.14), was introduced as a measure of visually regional importance for meshes. In details, on a small region with radius σ ($\sigma \in \{2\varepsilon, 3\varepsilon, 4\varepsilon, 5\varepsilon, 6\varepsilon\}$), where ε is 0.3% of the diagonal length of a mesh's bounding box), for the center vertex v, it firstly computes the Gaussian average of the mean curvatures $(G(K_{mean}(v),\sigma))$ over the vertices within the region. Then the local mesh saliency for scale σ: $S(v)$ is calculated as: $S(v) = |G(K_{mean}(v),\sigma) - G(K_{mean}(v),2\sigma)|$.

Based on the principal curvatures and their derivatives, several well-known metrics were designed and will be detailed in the following parts.

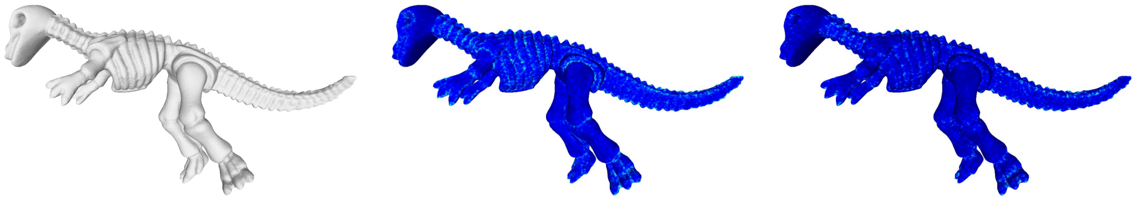


Figure 3.12: Examples of maximum and minimum curvatures. From left to right: Dinosaur object; Colorized maximum curvature measurement; Colorized minimum curvature measurement. The curvatures measured and colorized by Meshlab [19] using APSS algorithm [41]. Warmer color means higher measured value.

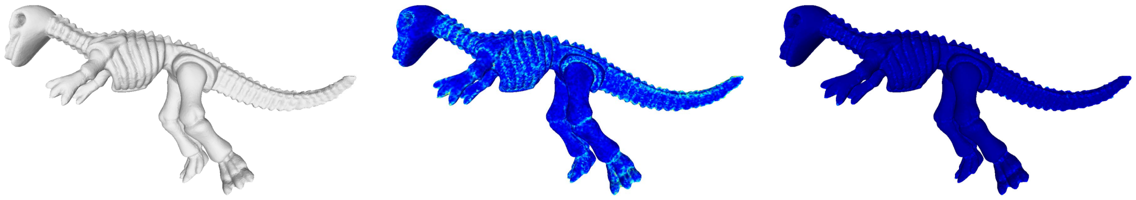


Figure 3.13: Examples of Curvedness and Gaussian curvatures. From left to right: Dinosaur object; Colorized Curvedness measurement; Colorized Gaussian curvature measurement. The curvatures measured and colorized by Meshlab [19] using APSS algorithm [41]. Warmer color means higher measured value.

Inspired by Structural SIMilarity (SSIM) [127] index for 2D image quality evaluation, Lavoué et al. [59] introduced the Mesh Structural Distortion Measure (MSDM) for evaluating the watermarking algorithms. The MSDM performs an analysis of the mean curvature (i.e. average, standard deviation and covariance) on local regions of the meshes. Accordingly, the local measure of the metric *LMSDM* in a region (window) a of A and corresponding b of B is defined as:

$$LMSDM(a,b) = (0.4 \times L(a,b)^3 + 0.4 \times C(a,b)^3 + 0.2 \times S(a,b)^3)^{\frac{1}{3}} \tag{3.24}$$

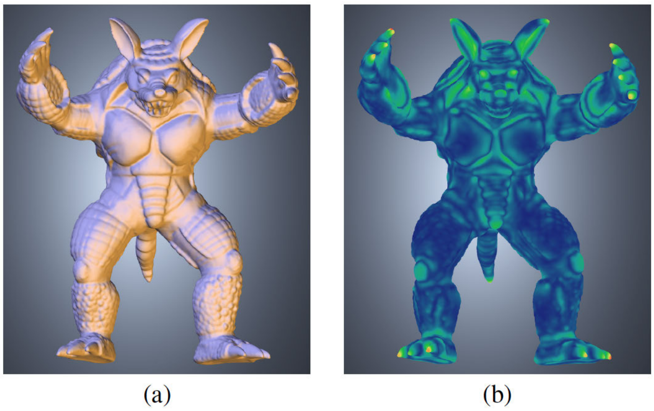

(a) (b)

Figure 3.14: Mesh Saliency: Image (a) shows the Stanford Armadillo model, and image (b) shows its mesh saliency. The image comes from [63].

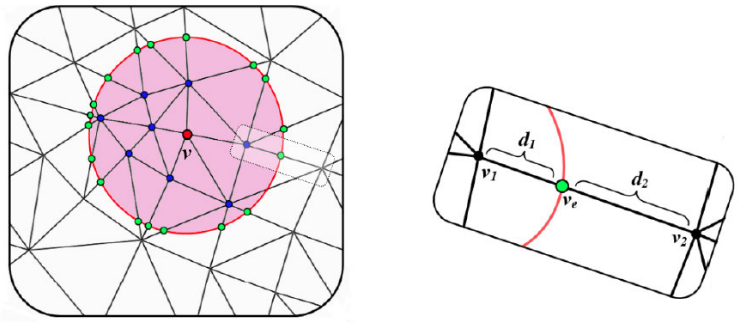
$L(a,b)$, $C(a,b)$ and $S(a,b)$ represent average curvature, contrast and structural comparison functions respectively, which are defined as:

$$\begin{aligned} L(a,b) &= \frac{||\mu_a - \mu_b||}{max(\mu_a, \mu_b)} \\ C(a,b) &= \frac{||\sigma_a - \sigma_b||}{max(\sigma_a, \sigma_b)} \\ S(a,b) &= \frac{||\sigma_a \sigma_b - \sigma_{ab}||}{\sigma_a \sigma_b} \end{aligned} \tag{3.25}$$

with μ_a, μ_b, σ_a, σ_b and σ_{ab} are respectively average , standard deviation and covariance of the mean curvature over the local windows a and b. A local window is denoted as a set of connected vertices and intersected points inside and on a sphere with a given radius: 0.5% of the bounding box length. Figure shows an example of local window computation, and the curvature of a edge point $C(v_e)$ (i.e. the intersection between the sphere and a edge of the meshe) is are computed as: $C(v_e) = \frac{d_2}{d_1+d_2}C(v_1) + \frac{d_1}{d_1+d_2}C(v_2)$. Finally, The global MSDM index between two meshes A and B is computed via *Minkowski* pooling method associated with all local windows:

$$MSDM(A,B) = (\frac{1}{n_w} \sum_{j=1}^{n_w} LMSDM(a_j, b_j)^3)^{\frac{1}{3}} \tag{3.26}$$

where n_w is the number of local windows. The MSDM is an asymmetric visual distortion index ranging from 0, meaning the identical distortion, to 1, implying two completely different meshes.



(a) Local window computation (b) *Edge point* computation

Figure 3.15: Example of local window computation for a vertex v and edge point v_e. The image is obtained from [59].

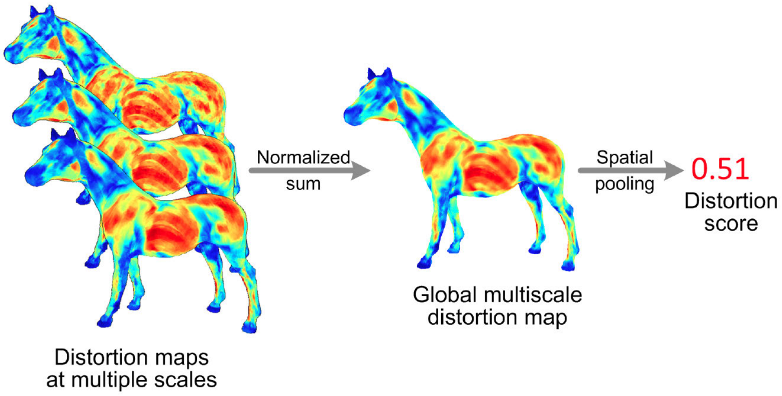


Figure 3.16: Distortion maps at different scales by MSDM2, the different MSDM measures are integrated into one multiscale map and then pooled into a global distortion score. The image comes from [56].

Following the concept of MS-MSSIM [129], Lavoué et al. later on proposed MSDM2 [56], which is a multi-scaled and symmetric version of MSDM (see Figure 3.16). At a given scale h, MSDM2 also upgraded the average curvatures over local windows a and b up to Gaussian average ones, where a Gaussian weighting function $w_v^h()$ is adaptively multiplied to the mean curvature of each neighboring point v_i according to the distance from center vertex v to the v_i. At a give scale h, the $w_v^h(v_i)$ is computed as follows:

$$w_v^h(v_i) = \frac{e^{-2||v_i - v||^2/h}}{\sum_{v_j \in N(v,h)} e^{-2||v_j - v||^2/h}} \tag{3.27}$$

where $N(v,h)$ is the ensemble of vertices and intersections over the local window. The corresponding standard deviations and covariance values are then upgraded to the Gaussian-weighted standard deviations and covariance values following the same scenario as the Gaussian average. The local distortion measure $LD^h(v)$ at the given scale h is then defined as:

$$LD^h(v) = 0.4 \times L^h(a,b) + 0.4 \times C^h(a,b) + 0.2 \times S^h(a,b) \tag{3.28}$$

Thus, for n scales of local windows, the multi-scale local distortion measure $MLD(v)$ is:

$$MLD(v) = \frac{\sum_{i=1}^{n} LD^{h_i}(v)}{n} \tag{3.29}$$

From mesh A to B, the global distortion score $GMD(A,B)$ of the MSDM2 is computed via a *Minkowski* pooling, defined as:

$$GDM(A,B) = (\frac{1}{n_B}\sum_{j=1}^{n_B} MLD(v)^3)^{\frac{1}{3}} \tag{3.30}$$

with n_B the number of vertices over mesh B. Finally, the symmetric MSDM2 index is calculated as follows: $MSDM2 = 1/2(GDM(A,B) + GDM(B,A))$.

Motivated by MSDM2, a recent mesh visual quality metric (Tensor-based Perceptual Distance Measure (TPDM)) [115] considers full information of curvature amplitudes and principal directions, which is called mesh curvature tensor: T.

According to their definition, on each vertex v of a mesh, curvature tensor T is computed as:

$$T(v) = \frac{1}{|B|}\sum_{edges\,e} \beta(e)|e \cap B|\bar{e}\bar{e}^t \tag{3.31}$$

where $|B|$ is the area of the geodesic disk, defined by user, $\beta(e)$ is the signed angle between the normals of the two triangles incident to edge e, $|e \cap B|$ is the length of the part of e inside B, and $\bar{e}$ is a unit vector in the direction of e (see Figure 3.17).

Given a vertex v in the reference mesh A and its corresponding vertex v' on the surface of the distorted mesh B, the curvature tensors on the two vertices are definded respectively by T and T'. Then, the local tensor distance (LTD) measure for each pair of v and v' is computed as:

$$LTD = \theta_{min}\delta_{K_{min}} + \theta_{max}\delta_{K_{max}} \tag{3.32}$$

where θ_{min} (similarly for θ_{max}) is an angle between γ_{min} (i.e. the minimum direction of T) and γ_1' (i.e. the minimum direction of T' that has the smallest angular distance to γ_{min}), while $\delta_{K_{min}}$ (similarly for $\delta_{K_{max}}$) means a Michelson-like contrast of the curvature amplitudes k_{min} (i.e. the minimum curvature of T) and k_1' (i.e. the minimum curvature of T').

In recent years, Wang et al. [125] designed a metric so-called FMPD to measure the perceptual

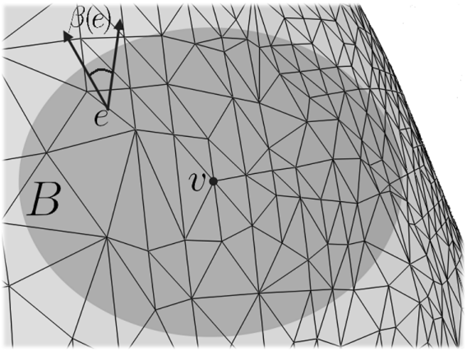


Figure 3.17: Illustration of the curvature tensor computation The image comes from [115].

distance of two meshes. Based on the analysis of Gaussian curvature, the FMPD measures the mesh local roughness. More precisely, for each vertex v_i on a mesh, they defined the local roughness LR_i as a weighted difference between the Gaussian curvatures of v_i and the Gaussian curvatures of the neighbors. Furthermore, the final local roughness measure does not only consider the masking effect by modulating a masking effect function, but also take a surface-weighted average of the mesh's local roughness. The global roughness (GR) of the mesh is then computed as a normalized surface integral of the local roughness. Finally, the perceptual distortion FMPD index between mesh A and B is defined as:

$$FMPD(A,B) = c|GR_A - GR_B| \tag{3.33}$$

where $c = 8.0$ is a scaling factor that scales the FMPD index into the interval $[0,1]$.

Normal-based metrics

Besides geometric position and curvatures, the normal of a vertex (or a facet) can also provide useful information for evaluating geometric properties of a 3D mesh. And an angle between the normals of two adjacent facets on a mesh yields the dihedral angle, which is related to the surface roughness [56]. Two pioneering surface roughness measures based on analysis of normals were the *multi-scale roughness* proposed by Wu et al. [134], and *smoothing-based roughness* by Gelasca et al. [38].

In the *multi-scale roughness* estimation, Wu et al. [134] associated each dihedral angle to a quantity of local roughness by:

$$\rho_d = 1 - \vec{N}_1 \cdot \vec{N}_2 \tag{3.34}$$

where $\vec{N}_1$ and $\vec{N}_2$ are the normals of adjacent triangles. Based on the definition, the roughness of a given triangle T is related the mean $G(v_j)$ and variance $V(v_j)$ values of local roughness

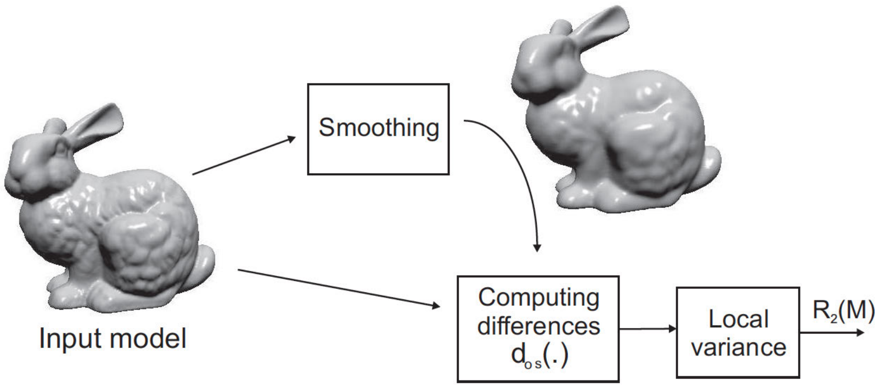


Figure 3.18: Smoothing-based Roughness Estimation. The diagram is from [38].

ρ_d of neighbor triangles, like:

$$\rho_1(T) = \frac{G(v_1)V(v_1) + G(v_2)V(v_2) + G(v_3)V(v_3)}{V(v_1) + V(v_2) + V(v_3)} \tag{3.35}$$

The final local roughness evaluation per-vertex $\rho_1(v)$ takes account for different scales of bumpiness (i.e. find the maximum value among the 1-ring, the 2-ring and the 4-ring computations around a given vertex v). The global roughness $\rho_1(A)$ of a mesh A is the sum of local roughnesses of all vertices.

The *smoothing-based roughness* estimation [38] is based on the consideration whether distortions can be perceived on smoothed surfaces (see Figure 3.18). Accordingly, the approach firstly applies a Taubin smoothing algorithm to smooth the mesh, and then measures the distance between each vertex v of a mesh A and its corresponding vertex v^S of a smoothed version A^S along the normal of v^S (i.e. $\vec{n}_v^S$) in the following way:

$$d_{OS}(v, v^S) = proj_{\vec{n}_v^S}(v - v^S) \tag{3.36}$$

Then the local smooth-based roughness is the division between variance of the 2-ring $d_{OS}(v, v^S)$ and the area of the faces that form the 2-ring of v. The total roughness $\rho_2(A)$ of a mesh A is the sum of local roughnesses of all vertices.

By utilizing these two surface roughness measures respectively, Corsini et al. [23] developed two mesh quality metrics ($3DPM1$ and $3DPM2$) considering the increment of the global roughness $\rho(\cdot)$ (i.e. $\rho_1(\cdot)$ or $\rho_2(\cdot)$) between mesh A and its distorted mesh B:

$$3DWPM(A,B) = \log(\frac{\rho(B) - \rho(A)}{\rho(A)}) - \log k \tag{3.37}$$

with constant k to avoid numerical instability.
Soon afterwards, Váša and Rus [119] defined the *oriented dihedral angle* per-edge to measure the mesh local roughness. In details, on the surface of a mesh A, assume two adjacent triangles $t_1 = \{v_1, v_2, v_3\}$ and $t_1 = \{v_3, v_2, v_4\}$ with normals $\vec{N_1}$ and $\vec{N_2}$, the oriented dihedral angle is expressed as follows:

$$D_{t_1,t_2} = \arccos(\vec{N_1}, \vec{N_2}) * sgn(\vec{N_1}(v_4 - v_3)) \tag{3.38}$$

The corresponding dihedral angle on mesh B is denoted as $\overline{D_{t_1,t_2}}$.
Based on this local roughness, they proposed a mesh visual quality metric DAME (the abbreviation of Dihedral Angle Mesh Error), which computes the local roughness variations between mesh A and mesh B. The global DAME index is the summation of the local roughness variations for all edges of two meshes in the following way:

$$DAME = \frac{1}{n_e} \sum_{\{t_1,t_2\} \in n_e} ||D_{t_1,t_2} - \overline{D_{t_1,t_2}}|| \cdot m_{t_1,t_2} \cdot \omega_{t_1,t_2} \tag{3.39}$$

where n_e is the number of all triangle pairs that share one edge, m_{t_1,t_2} is a weight relative to the masking effect, which increases the values of distortion on smooth surfaces, $\omega_{t_1,t_2}$is a weights relevant to the surface visibility.

So far, we have presented several mesh visual quality metrics based on geometric attributes (e.g., the vertex position, the curvatures, the normal of vertex (or facet), and their derivatives, such like dihedral angles, geometric Laplacian, Laplacian of Gaussian curvature and so forth). A metric [57] based upon the optimal combination of all these attributes was proposed by Lavoué et al. The surveys [58, 61] have well summarized the performance of all these mesh quality metrics in terms of correlation with subjective opinions, and the results indicate that the classical geometric distances, like Hausdorff and RMS, provide a very poor correlation with human judgment, while the most recent model-based metrics (e.g. MSDM2, TPDM, FMPD) provide much better performance.

3.3.2 Image-based metrics

In contrast to the model-based metrics, the image-based ones estimates the visual quality of images or video sequence computed by rendering the 3D objects (e.g., snapshots of an object from different views) by using image quality metrics (e.g., MSE, SSIM, VIF, etc.).

Figure 3.19: The sampled 20 2D snapshots sampled around the 3D mesh as the main criterion for decimation. The image is captured from [72].

The image-based metrics thus simplify the evaluation procedures.

The first use of image-based metrics was proposed by Lindstrom and Turk [72] in the context of evaluating the artifacts caused by simplification. In their work, they applied a simple RMS metric to measure distances between two models using snapshots regularly sampled from 20 different viewpoints on a bounding sphere of each model (see Figure 3.19). Soon afterwards, they[71] replaced RMS with some visual quality metrics (e.g., Sarnoff JND metric, etc), but surprisingly found that RMS performs better than the visual quality metricsin terms of predicting the perceived quality of artifacts.
For the purposes of driving simplification and remeshing of textured meshes, Qu et al.[86] proposed a metric based on the computation the visual masking properties of a texture. By using the Sarnoff JND to compare the original texture map and its Gaussian filtered version, the obtained masking map which can indicate the visual masking potential on the surface, and thus guide the remeshing or simplification algorithms.

In order to design an optimal approach of discrete levels of detail (LOD) for the visualization of complex 3D building facades, Zhu et al. [138] presented a new quantitative analytical method to study the relationship between the viewing distance and visibility of a textured model's façade details using image visual quality metrics (i.e. VDP and SSIM).

Since the image-based metrics evaluates the 3D mesh visual quality in a simple and unified fashion, can they perform better than model-based metrics regarding the correlation with subjective evaluations on 3D distortions? In the study conducted by Lavoue et al. [60], they suggested image-based metrics may perform significantly close to the model-based metrics in evaluating the visual quality of different versions of a same object under a single type of artifact. However, the image-based metrics are less accurate in comparing different artifacts or artifacts applied to different 3D objects.

3.4 Conclusion

In this chapter, we briefly presented several classical and recent objective visual quality metrics for 2D images, videos and 3D meshes. Most of them aim at evaluating the distortions caused by one specific graphic processing algorithm (e.g., Waston's DCT [130] metric for JPEG compression, IFC [101] for image noise, Hausdorff distance [96] for mesh simplification, MSDM [59] for mesh watermarking, etc.), while others focus on evaluating different kinds of distortions (e.g., SSIM [127] , MS-SSIM [129], VQM [85] , FMPD [125], MSDM2 [56], etc.). About the metric outputs, several metrics can only provide one global quality score (e.g., VIF [100], $3DWPM_1$, $3DWPM_2$ [23], image-based mesh metrics [72, 86, 138] , etc.), while some other metrics, besides the global quality score, can also provide local distortion maps (e.g., MSE, PSNR, IFC , SSIM, MS-SSIM, IW-SSIM [128] , Hausdorff distance, MSDM, MSMD2, GL, DAME [119] , etc.).

Ultimately, since the output of a visual quality metric should maximally approximate the human perception, the performance of all the objective metrics are therefore evaluated using the subjective quality experiments, where a group of observers provide subjective opinions on a set of distorted images or 3D models. Thus, this draws forth another important theme that is how to design a subjective experiment, from which we can collect the effective information of human estimations on different artifacts, and quantitatively evaluate the performance of the metrics using the collected subjective opinions. In the following chapter, we turn to the detailed protocols of designing *subjective quality assessments experiments* and the means of evaluating the performance of the objective qualiy metrics.

Chapter 4

Subjective visual quality assessments in computer graphics

4.1 Introduction

The *objective* visual quality metrics introduced in previous chapter focus on objectively and automatically evaluating the visual quality and/or predicting the locally perceived distortions in processed graphics, whereas a *subjective* visual quality assessment consists in asking a bunch of observers to globally or locally evaluate (i.e. rate, select a preference or mark) distorted 3D models or images. Hence, the subjective studies are generally used to evaluate the performance of objective metrics. Furthermore, some objective visual quality metrics can be driven by subjective opinions using machine learning [57]. However, a subjective quality assessment usually requires a sufficient number of observers and experimental trials to draw significant conclusions, which makes the assessment very expensive. Nevertheless, such subjective study still cannot ensure to generate conclusive results [77]. Hence, a great deal of research was prompted to design efficient and rigorous subjective quality assessment experimental protocols. The most prominent subjective study protocols are: *single-stimulus*, *double-stimulus*, *forced-choice pairwise comparison* and *similarity judgments* (see Figure 4.1). These four protocols are able to produce a scaled *global quality* (i.e. a single score) from observers, which aims to reflect the global level of annoyance caused by all the artifacts on a 3D model or an image. The single-stimulus and the double-stimulus protocols may also enable observers to evaluate the *local distortion visibility* of artifacts (i.e. indicating the spatial localization of the distortions). In this chapter, we will cover every subjective quality assessment protocol, including the specific experimental setups, the four subjective assessing protocols, scaling methods of computing collected subjective data, and the means of testing objective quality metrics performances with the scaled subjective data. The performance of all the introduced quality metrics will be summarized at the end of the chapter.

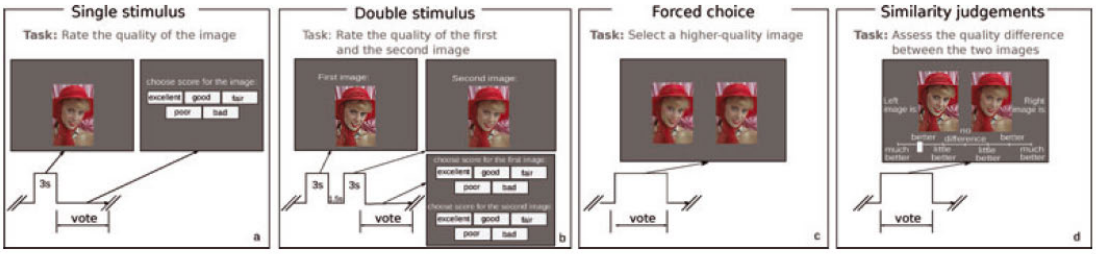


Figure 4.1: Overview of the four subjective quality study protocols. from left to right: (a) *single-stimulus*, (b) *double-stimulus*, (c) *forced-choice pairwise comparison* and (d) *similarity judgments*. The diagram shows the timeline of each protocols and the corresponding screens.The diagram originates from [77].

4.2 Specific setups for subjective experiment

Several important parameters are involved in a subjective experiment, including the selection of graphic data (e.g. images, videos, 3D models, etc.) to be evaluated, the types and the amplitudes of the distortions, and the rating protocols (i.e. single-stimulus, double-stimulus, forced-choice pairwise comparison and similarity judgments). Precisely, the number, types and the amplitudes of stimuli are usually decided by following the criteria [8, 127] below:

- *Stimuli types*: Presenting diverse distortions in order to conclude significant remarks after analysis of objective metrics performances.
- *Distortion levels*: The amplitude of a stimuli should be visible but not too strong; the idea is to span a large range of perceptual impacts.
- *Stimuli number*: Using limited number of stimuli to restrict each experiment time within a generally acceptable range by each observer (30 minutes is usually the accepted limit [47]).

However, for a subjective quality study involving 3D content, additional conditions should be considered as detailed in [8, 61]:

- *Lighting*: Rogowitz and Rushmeier [89] addressed that the spatial location and nature of light source(e.g., spot light, point light, etc.) can strongly influence the perceived quality of artifact.

- *Materials and Shading*: Complex materials and shading methods may enhance the visibility or introduce masking effect [34, 72].
- *Background* The perceived quality of a 3D model may be influenced by its background. Particularly the background may affect the observations of the silhouette [61].
- *Animation & interaction*: The complexity of showing a 3D model to the observers can range from the simplest display (e.g., a static image with fixed view point [132].) to the most complex interactions (e.g., free controls of 3D models, such like rotation, translation, zoom, marking, etc. [23]). For each observer, it is important to get access to different viewpoints of the models, however, allowing this free interaction may result in cognitive overload, which may alter the results [61]. Animation is a promising compromise used in [81], however, a fairly slow animation speed is required, since the speed can impact on the Contrast Sensitivity Function (CSF) [49].

Some public database provide processed 3D models with different types and distortion extents. For instance, Figure 4.2 presents several examples of stimuli from LIRIS/EPFL General-Purpose Database [59], and LIRIS Masking Database [55].

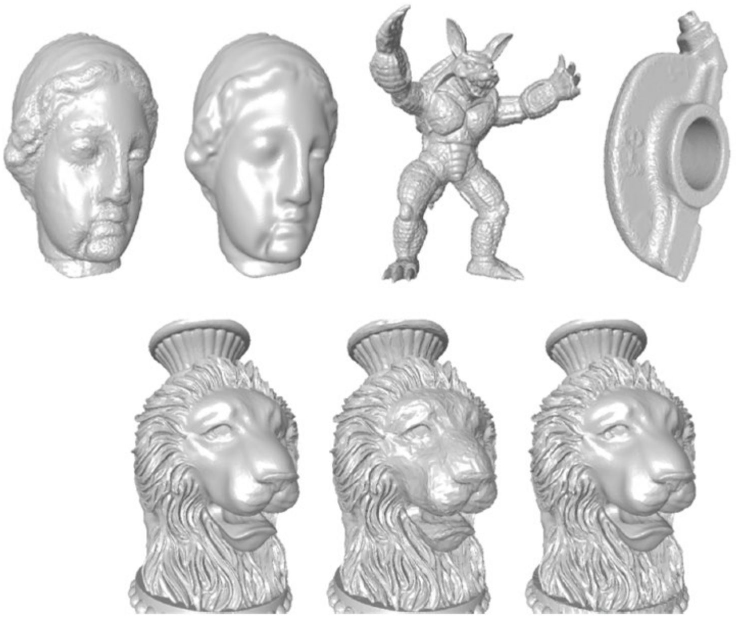


Figure 4.2: Several 3D stimuli, which are processed by smoothing and noise algorithms on different areas of the model surfaces from LIRIS/EPFL General-Purpose Database [59] (Top row) and LIRIS Masking Database [55] (Bottom row). The figure is from [57].

4.3 Four subjective quality evaluation protocols

Four dominant subjective quality evaluation protocols exist: *single-stimulus*, *double-stimulus*, *forced-choice pairwise comparison* and *similarity judgments*. They are adapted in a vast majority of research of image, video and 3D mesh quality assessments. Figure 4.1 shows a general view of these four protocols.

4.3.1 Single stimulus

Global quality rating experiment

The single stimulus scenario (see Figure 4.1 (a)) for global fidelity of processed graphics often involves presenting an image, a video clip or a 3D model within a short and fixed duration, and then demanding an observer to rate it using one of describing categories (e.g excellent, good, fair, poor or bad) or a quality scale (e.g., from 1 (the worst quality) to 10 (the best quality)) [80, 15]. The discrete categorical scales are widely used in major subjective experiments, however some experiments prefer continuous to categorical scales to avoid quantization artifacts [77]. A branch method of the single stimulus scenario is so-called *absolute category rating with hidden reference* [80], where the reference is displayed among other stimuli. For instance, Corsini et al. [23] utilized this protocol to collect the subjective opinion on the different artifacts caused by watermarking. Similarly, Lavoué et al. [59] adapted the same scenario to study the HVS sensitivity on general artifacts (e.g. noise addition, watermarking and smoothing, etc.) with varying distortion strengths. In [77], Mantiuk et al. addressed that presentation time is a variable that affects the overall length of the experiment and thus the efficiency of the experimental method, therefore 3-10s presentation time for each stimulus (including the reference) is recommended. The single stimulus scenario is very efficient, only $n+1$ trials (one additional trial for the reference) are sufficient for assessing n stimuli.

Local distortion visibility experiment

Instead of global rating, the single stimulus scenario for local distortion visibility assessment requires the observers to directly mark regions with annoying distortions. To the best of our knowledge, Čadík et al. [9] implemented this protocol (see Figure 4.3), instructing observers to mark all the areas containing unpleasant distortions in an image, where only distorted images were exhibited to the observers. Each observer took on average 30 minutes to finish marking 11 processed images (e.g., noise, alias, brightness, etc.) in one experiment. In their work, the collected data are used to reveal weaknesses of both simple (PSNR, MSE, etc.) and advanced (SSIM, MS-SSIM, etc.) metrics.

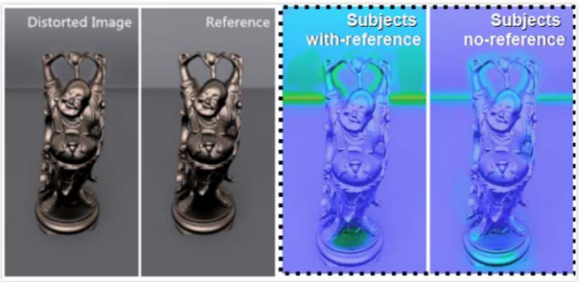


Figure 4.3: Subjective detection probability maps of local visible distortion in [9] (two figures on the right), warmer color means higher probability. The middle right figure shows the detection probability map from double-stimulus protocol, in which observers were asked to mark all the areas in the image containing noticeable distortions compared with the reference, and the right figure presents the map from single-stimulus scenario, where observers marked all the areas in the image containing objectionable distortions without the reference.

4.3.2 Double stimulus

Global quality rating experiment

Double stimulus categorical rating for the global fidelity (see Figure 4.1(b)) is analogous to the single-stimulus method, but a reference and its distorted version are simultaneously shown to an observer for the fidelity rating. This method requires n trials to assess n stimuli. Subjective quality studies conducted by Watson [132] and Rogowitz and Rushmeier [89] followed this scenario to collect user opinions on different types of simplified 3D models at different simplification ratios. The purpose of Watson [132] was to evaluate the performance of image-based and geometric metrics for the task of predicting distortion quality of simplified 3D models, whereas Rogowitz and Rushmeier [89] attempted to study whether 2D snapshots of a 3D model can be suited to predict the visual quality. Other works devised by Rushmeier et al. [94] and Pan et al. [81] also relied on the double stimulus method to study the interactions between texture and mesh resolutions influencing the visual appearance of a simplified textured 3D model. In order to study the visual impact of masking effects on 3D mesh surface, Lavoué et al. [55] conducted a double-stimulus experiment by asking observers to rate the fidelity of meshes with varying strength noises applied on either smooth or rough areas.

Local distortion visibility experiment

The double stimulus protocol (see Figure 4.3) can be used for evaluating the local distortion visibility as well, where researchers instruct observers to mark the visually different areas on processed graphic by comparing its reference. In [9], Čadík et al. also conducted this double-stimulus experiment, verifying the high correlation between the locally perceived distortions with full-reference (i.e. double stimulus protocol) and with no-reference (i.e. single stimulus protocol).

4.3.3 Forced-choice paired comparison

Comparing to assigning a discrete or continuous quality scale to a each stimulus in the single- or double-stimulus scenarios, the forced-choice paired comparison protocol (Figure 4.1(c)) simply requires each observer to provide a preference (e.g., {'better', 'worse'}) between each pair of stimuli over a number of comparison trials [67]. For instance, in the experiment designed by Wills et al. [133], each observer was asked to click on the appropriate button to indicate which side of the image (left or right) appears more similar to the center image (see Figure 4.4). This protocol benefits the advantages of simplicity and reliability enhancement, especially when the quality difference between the processed graphics is not easily distinguished or the multiple modalities of quality variations (e.g., spatial resolution, temporal resolution and so on) are involved [14, 121, 67]. Moreover, in contrast to the single- or double-stimulus methods that produce a score to describe the subjective visual quality of the artifacts, paired comparison method usually provides a ranking of all stimuli from the best perceived quality to the worst quality or vice versa.

However, the forced-choice paired comparison can be very tedious, since it requires $\binom{N}{2}$ comparison trials to get all the possible evaluations on N stimuli. Especially, when N is large, the trials number may be too large to be feasible [67]. In such cases, it is necessary to use some sorting algorithms to reduce the trials number [102]. Efficient sorting algorithms, such as quicksort, can reduce the number of comparisons necessary to order a set of conditions to approximately $NlogN$ [77, 97].

4.3.4 Pairwise similarity judgment

Further than forced-choice paired comparison, pairwise similarity judgment method (Figure 4.1(d)) does not only ask observers to choose their preferences but also indicate a quality

Figure 4.4: An example of experimental platform for forced-choice paired comparison protocol. An observor is asked to click on the appropriate button to indicate the center image appears more similar to the image on the left or to the image on the right. This platform is adapted from [133].

number on the how large the difference in quality is between two stimuli, in which number '0' means no difference is observed between the pair [77]. Pairwise similarity judgments can also use the sorting algorithms to reduce the comparisons number.

4.4 Analysis of subjective data

4.4.1 Inter-observer agreement

During the experiments, the observers may provide implausible quality scores, preferences or local distortion markings due to their misunderstandings or derelictions of the experiments or their divergent opinions on evaluating one stimulus [77]. Hence, before exploiting the results from the subjective experiments, it is critical to analyze the inter-observer agreement between all observers, which indicates the coherency of observers' opinions.

To assess the inter-observer agreement, a well-known mathematical method *Kendall's coefficient of agreement U* introduced in [95] can be used for analyzing the consistency of raters. Generally, u ranges from $u = -1/(m-1)$, meaning no agreement between m observers, to 1, indicating the consensus of all the observers [95]. The *Kendall's coefficient of concordance W* [51] can aslo reflect the inter-observer agreement of the rankings from paired comparison and similarly judgment. For the agreement of local distortion markings, Čadík et al. [9] calculated the coefficient u per pixel of a distorted image and made an average coefficient $\bar{u}$ across all the pixels to indicate the overall inter-observer agreement of the image.

Usually, a high coefficient u (or W) value means a high consensus of the observers or, in other words, no statistical evidence that they differ in quality [77]. However, a low u (or W) value (e.g., less than 0.5) may indicate a strong disagreements between observers or some unreliable rating qualities. In such case, it is necessary to remove those unreliable results. Rohaly, A. M., Corriveau, P. J., Libert, J. M., Webster, A. A., Baroncini, V., Beerends, J., ... & Winkler [90] proposed a removal method. Precisely, for each stimulus, the method firstly counts the number of rating results given by each observer that lie out the range $[-2\sigma, 2\sigma]$ (σ: the standard deviation of the results over all the observer for one stimulus), and then rejects the observer whose results meet one of the following conditions: either (1) more 5% of his results are outside the range, or (2) the results outside that range are evenly distributed so that the absolute difference between the result numbers that exceed the lower and the upper bound of that range is not more than 30%. If u (or W) remains a low value after removal measures, it possibly implies a highly difficult task for observers to perceive the artifact.

4.4.2 Scaling methods

Once all the collected experimental data is validated for high reliability by the analysis of inter-observer agreement, it needs scaling into a mean quality index for each stimulus over all the observers.

For the global results from single- and double-stimulus protocols, addition to the mean opinion score (MOS) over the observers, we can also compute the the differential mean opinion score (DMOS) as the difference of MOS values between the reference and its distorted version:

$$d_{i,j,k,r} = q_{i,ref(k),k,r} - q_{i,j,k,r} \tag{4.1}$$

The indices correspond to i-th observer , j-th processing algorithm, k-th processed graphic (image, video, 3D model, etc.), and r-th repetition. $ref(k)$ is the reference condition for the k-th processed graphic.

However, each observer tends to provide a separate quality scale for each particular stimulus, resulting in different scales associated with each observer. Thus it is necessary to unify all the results over observers in order to make their scores comparable [118]. To unify the results, the easiest way is to apply a linear transformation that makes the mean and the standard deviation equal for all observers. The uniformed result after the transform is called $z-$score, and expressed as:

$$z_{i,j,k,r} = \frac{d_{i,j,k,r} - \bar{d}_i}{\sigma_i} \tag{4.2}$$

with $\bar{d}_i$ the mean DMOS and σ_i the standard deviation over all the stimuli
For the scaling methods on local distortion markings, Herzog et al. [45] computed a probability of detection on local distortion by averaging marking labels (0: unmarked; 1: marked) per pixel across all the observers. And finally, subjective distortion map is produced based on the probabilities of all the pixels (see Figure 4.3).
After obtaining the rankings from paired comparison and similarity judgments, a scaling method is required to convert the rankings of stimuli into quality scales. An easy method [62] uses a so-called *preference matrix* (see Table 4.1) to transform the rankings of the stimuli into *vote score*, which is a sum of preferences number of one stimulus over others across all the observers.

	d_1	d_2	d_3	d_4	d_5	d_6	score
d_1	-	1	0	0	1	1	3
d_2	0	-	0	1	1	0	2
d_3	1	1	-	1	1	1	5
d_4	1	0	0	-	0	0	1
d_5	0	0	0	1	-	1	2
d_6	0	1	0	1	0	-	2

Table 4.1: Example preference matrix for one observer when shown six different stimuli from d_1 to d_6 . Each stimulus in a row, d_i is compared with another d_j in each column. If the observer prefers d_i to d_j, then 1 is recoded in the cell (d_i, d_j), otherwise 0 is filled in the cell. Suggested by [62], the vote score of each stimulus d_i is the summation of ones in each row.

4.5 Measures of evaluating objective quality metric

Suggested by video quality experts group (VQEG) Phase I FR-TV [27], the performance of an objective metric is evaluated with respect to three attributes of their ability to predict the results of subjective quality assessments.

- *Prediction accuracy*: The ability to predict the subjective quality results with low error;
- *Prediction monotonicity*: The degree to which the model's predictions agree with the relative magnitudes of subjective quality results;
- *Prediction consistency*: The degree to which the model maintains prediction accuracy over corpus.

Following these criteria, after obtaining the subjective quality score or the perceived distortion map, researchers usually rely on several statistical measures to do some regression analyses and compute correlation coefficients between these collected scores and the predictions of the objective quality metrics. The coefficients can indicate the performance of the metrics, and thus evaluate and compare them quantitatively.

4.5.1 Performance of global quality prediction

In [27], four measures (i.e. RMS error, Pearson Linear Correlation Coefficient (LCC) [83], Spearman Rank Order Correlation Coefficient (SROC) [106] and Non-linear regression analysis) are advised and broadly used to evaluate the performance of the global quality predictions of the objective metrics.

- *RMS error*: It measures the root-mean-square of differences between DMOS and predictions of an objective quality metric over all stimuli. RMS error is simple but not competent to indicate the prediction consistency [27] (see Figure 4.5).

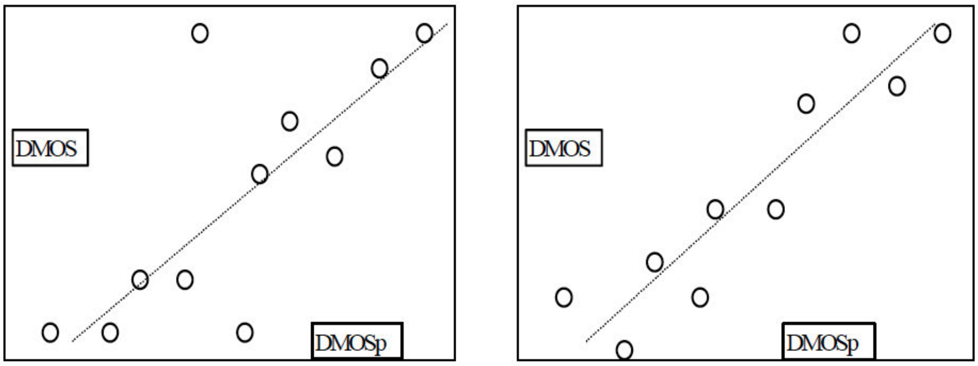


Figure 4.5: Figures left and right show two objective metrics with approximately equal RMS errors between metrics predictions (shown as $DMOS_p$ in figures) and collected $DMOS$. Left figure is an example of a metric that has quite accurate predictions for the majority of stimuli but has large prediction error for the two points in the middle of the figure. Right figure is an example of a metric that has a balanced set of prediction errors – it is not as accurate as the metric of left figure for most of the sequences but it performs "consistently" by providing reasonable predictions for all the stimuli. However, RMS errors cannot reflect this difference in terms of their measured values. The figures come from [27].

- *Pearson Linear Correlation Coefficient (LCC)*: It computes the covariance of DMOS and the objective predictions divided by the product of their standard deviations, and higher coefficient value means better correlation. This coefficient can well reflect the prediction accuracy.

- *Spearman Rank Order Correlation Coefficient (SROC)*: This coefficient is defined as the Pearson Linear Correlation Coefficient between the ranked values of DMOS and objective predictions. It is a sensitive measure of prediction monotonicty.
- *Non-linear regression analysis*: Prior to computations of correlation coefficients , it is suggested to use a non-linear function, which is a four-parameter logistic curve, to fit each objective prediction, when recognizing the potential non-linear mapping of the objective metric outputs to the subjective quality scores. The non-linear function is used to transform the set of objective predictions to a set of predicted DMOS values, and correlation coefficients are computed between the predicted DMOS and the subjective DMOS. Moreover, the outlier ratio (percentage of the number of predictions outside the range of ±2 times of the standard deviations) of the predicted DMOS can function as a measure of prediction consistency [127] (see Figure 4.6).

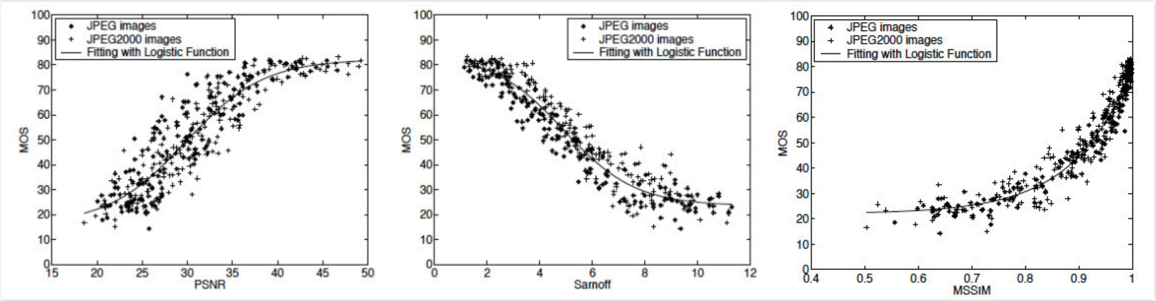


Figure 4.6: Scatter plots of subjective mean opinion score (MOS) versus metric prediction, presented in [127]. Each sample point represents one stimulus. All the points are fitted with four-parameter logic curves. from left to right, the predictions are produced by : PSNR, Sarnoff model and SSIM. The scatter plot of SSIM is the most convergent to the non-linear curve, meaning that SSIM has the highest prediction consistency among three metrics.

4.5.2 Performance of local distortion prediction

To the best of our knowledge, few measures evaluate the performance of the objective quality metrics for the task of predicting the local perceived distortion of the artifacts, while many objective metrics, such like SSIM, MS-SSIM, MSDM2, FMPD and so on, are able to provide such local distortion predictions. Researchers usually evaluate this performance by utilizing the subjective detection probability map of local visible distortion, namely the *subjective distortion map* (see. Figure 4.3). Particularly, the subjective distortion map can indicate the probabilities of visually noticeable distortion over all pixels of an image (or vertices of a 3D mesh surface), but cannot provide any information about the perceived magnitude of a distortion. Hence, such map is propitious to benchmark a metric ability to spot visibly distorted regions in terms of binary classification. To analyze such ability of an objective metric, Čădík et al. [9] used the receiver-operator-characteristic (ROC) [4], which reveals the relation between the size of areas that contain visible distortions and are correctly predicted by an objective metric (true positives), and the areas that do not contain visible distortions but are still predicted (false positives). ROC shows the relation of these two quantities for a varying classification threshold. The objective metric that gives a larger area under the ROC curve (AUC) is assumed to perform better (see Figure 4.7).

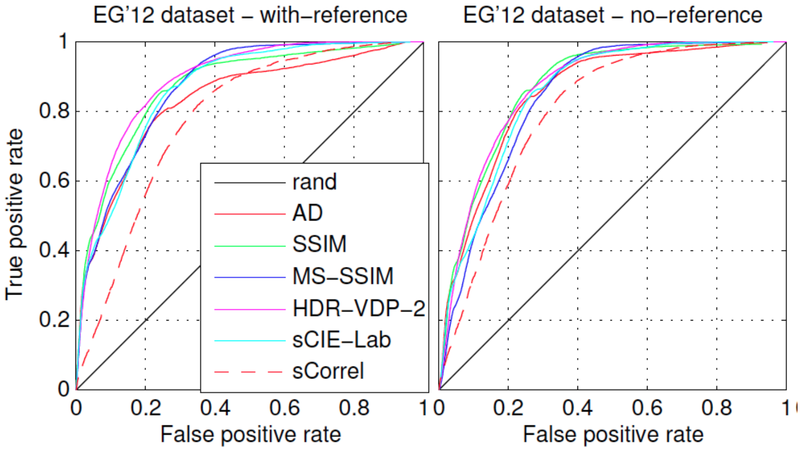


Figure 4.7: The performance of the local distortion prediction of the objective quality metrics is analyzed with ROC curve (the higher area-under-curve (AUC), the better the classification into distorted and undistorted regions). The results are from NoRM dataset [45], and subjective distortion maps of left figure are from double stimulus protocol, and subjective distortion maps of right figure are from single stimulus protocol. The threshold percentage of observers for determining the distortion visibility is 50%. The figure is captured from [9].

4.6 Performance comparison

Thus far, we have presented each procedure for implementing a subjective quality experiment, involving the design of the protocol, the analysis of the collected subjective data and the measures for quantitatively evaluating the performance of the objective quality metric in correlation with the subjective data. The performance of all the classical and recent objective quality metrics in digital graphics (images, videos, 3D models) are evaluated and summarized in the published works [70, 99, 58, 61, 9, 60]. Thus, we briefly introduce their conclusions in the following paragraphs:

- **Performance of image quality metrics**
  For the performance of global quality prediction, as in [70], comparing to the classical image metrics (e.g., PSNR), perceptual visual quality metrics (SSIM, VIF, VSNR) perform much better in terms of correlation with subjective opinions (DMOS). In particular, SSIM performs fairly consistently over the different distortion types, as contrasted with VIF and VSNR. Regarding the ability of predicting locally perceived distortion, Čadík et al. [9] suggested that a VDP based metric: HDR-VDP-2 performed better than its counterparts (e.g. SSIM, MS-SSIM, etc.) concerning the correlation with subjective distortion maps.

- **Performance of video quality metrics**
  Presented by Seshadrinathan et al. [99], the video quality metrics MOVIE and VQM are superior to other video metrics embedded with image quality metrics, such like PSNR, SSIM, VIF and so on, as regards to correlation with subjective scores (DMOS).

- **Performance of mesh quality metrics**
  Summarized in [61], the comparison results show that the classical geometric distances, like Hausdorff and RMS, provide a very poor correlation with human judgment, while the most recent model-based metrics (e.g. MSDM2, TPDM, FMPD) provide much better performance. In order to quantitatively compare the performance between model-based and image based metrics, the subjective study conducted by Lavoué et al. [60] , and indicates that the image-based metrics perform significantly close to the performance of model-based metrics for evaluating the quality of different versions of a same object under a single type of artifact. However, they also pointed out that the image-based metrics are less accurate in comparing different artifacts or artifacts applied to different 3D objects.

4.7 Conlusion

By now, we have presented an overview of objective and subjective quality assessments of digital graphics. The research on quality assessment has made a significant progress, with many applications. However, due to the complex modern graphic processes, this research is still in its early stage, and there is an increasing demand for more efficient and effective quality metrics, considering the interactions between geometry and rendering scenarios, functionalities of human visual system (HVS) in complex tasks, and so forth. For instance, to the best of our knowledge, Čadík et al. [9] explored a way to capture the information about HVS ability on perceiving the local distortions of 2D images. Based on this information, they evaluated the existing image quality metrics for the task of predicting locally perceived distortions. Unfortunately, hardly any research studies this HVS ability on perceiving 3D objects, while the information about the local visibility of 3D artifacts is also important to locally drive any geometry processing operations, such as mesh simplification, compression and distortion localization, etc. Meanwhile, as mentioned in [61], the current mesh quality metrics measure the distortions either on 3D meshes or on the final rendered images. We barely see an quality metric that considers the interactions between geometric and texture image information, while these interactions (e.g., a simplified 3D mesh mapped with a compressed texture) may cause complex visual artifacts, which could be stronger than either of two distortion impacts due to the additive effects, or weaker, thanks to the masking effects. Thus, the sudy on these interactions rises as an obvious trend in current research.

In next two chapters, we will detail our two major contributions to the subjective and objective visual quality assessments to meet the latest demand. Firstly, in Chapter 5, we present a newly designed subjective quality study based on evaluating the local visibility of geometric artifacts. The collected perceived distortion maps are used to represent several perceptual functionalities of the human visual system (HVS), and serve as ground-truth to evaluate the performance of well-known geometric attributes and metrics for predicting the locally perceived distortions. Then, in Chapter 6, we describe our second contribution: the subjective and objective quality evaluation of texture-mapped 3D models.

Chapter 5

Evaluating the local visibility of geometric artifacts

5.1 Introduction

As narrated in previous chapters, 3D meshes are subject to a wide range of processes which include compression, simplification, filtering, watermarking and so forth. These processes inevitably introduce distortions which alter the geometry of these 3D data and thus their final rendered appearance. Classical metrics such as Hausdorff distance and root mean square error have proven to be very poor predictors of the human perception and the visual annoyance of geometric artifacts [58]. As a consequence, mesh visual quality (MVQ) metrics have been recently introduced by the scientific community [119, 125, 56], mostly inherited from image quality assessment metrics [126]. Their objective is to evaluate the visual impact of the geometric artifacts (e.g. geometric quantization noise, smooth deformations due to watermarking, simplification artifacts and so on). A comprehensive review has been recently published about 3D mesh quality assessment [24].

The MVQ metrics provide good results in term of correlation with the subjective opinion, however they are mostly specialized in outputting one global quality score which predicts a global level of annoyance caused by all artifacts present in the 3D model. Such global quality index is relevant for many computer graphics applications, for instance, to evaluate a progressive compression method or to reduce/increase the quantization of either vertex position or mesh attributes in mesh compression algorithm. However, beside this global information, it is also important in many cases to obtain an information about the local *visibility* of the artifacts (i.e. a per-vertex measure of the perceived distortion). Such local information may allow to drive locally any geometry processing algorithms (e.g. simplification, remeshing, filtering and so forth). While some metrics like MSDM2 [56] do produce such local distortion maps, they have never been evaluated quantitatively with regards to this task.

In that context we propose a new full reference dataset of localized distortion maps for several types of geometrical artefact and meshes, obtained through a subjective experiment. For this purpose we gather a set of 3D models associated with several types of distortions (watermarking, quantization, non-uniform noise, smoothing and simplification). In the experiment, we ask observers to paint the vertices of distorted models where they perceived noticeable differences as compared with the reference ones (using a brush painting interface). After an analysis of the inter-observer agreement, we show that the results illustrate some perceptual mechanisms of the Human Visual System (Masking effect and frequency sensitivity). We then use this dataset to perform quantitative analysis and comparison of the performance of existing metrics and geometric attributes for the task of local visual difference prediction.

The rest of this chapter is organized as follows: section 5.2 presents our subjective experiment. Section 5.3 details the tested attributes and metrics. Finally, section 5.4 presents the performance measure and the results of our study.

5.2 Localized distortion experiment

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

5.2.1 Experimental design

As mentioned in previous chapter, the design of subjective quality evaluation experiment involving 3D content requires the choice of many rendering parameters (e.g. lighting, material, background, level of interaction). Several authors have made such 3D subjective quality experiments [132, 89, 23, 119]; whereas none of them tackled the task of local artifact marking, these works inspired us to use the following experimental setting:

- We considered a *full-reference* experiment, i.e. the observers see the distorted 3D model to mark, together with the reference not-distorted one. A no-reference task would have been too difficult since an observer, even expert, cannot have a sufficient *a priori* on a 3D shape to be able to notice that a region presents an artifact without knowing the reference.

- We considered a *full interaction* scenario (the use of static viewpoints is not recommended [89]).

- To minimize the influence of the light and the camera positions on the perception of artifacts, we simply consider a front directional lighting without specular reflections.

- As in [23], we chose a non uniform background in order to minimize the influence of the silhouette.

- We provide to the user a mesh painting tools, with a customable brush size.

Our interface is built upon MeshLab [19].

5.2.2 Stimuli creation

The choice of an appropriate corpus of stimuli for a subjective quality experiment is critical and is a complex task, since we have to chose a set of reference 3D models and their distorted versions with appropriate distortion types and appropriate distortion strengths, fulfilling the criteria detailed in section 4.2 Chapter 4. Particularly in our case, all the reference 3D models should represent meshes used either in digital entertainment, or virtual museum with a sufficient population diversity. Indeed, imagine that we chose a too simple reference object (e.g. a sphere) and too simple distortions associated with high amplitudes (e.g. adding sharp geometric creases) then our experiment will be useless since even a simple metric can predict the visibility of such obvious artifacts. On the contrary, the objective of our study is to challenge the ability of existing metrics to detect *complex* artifacts that occurs in practical operations (e.g. smooth structured patterns) in *complex* scenarios (e.g. near the visibility threshold). Indeed, even a basic Euclidean geometric distance is able to predict the visibility of a too obvious artifact like a sharp geometric crease added on a sphere. Besides this need of *complexity*, our dataset must also present a sufficient degree of diversity to allow us to draw significant conclusions regarding the performance of the metrics. Finally, given the *limited* but *necessary* time for the painting task, the dataset has to remain of *limited* size. Indeed, it is not realistic to ask people to paint 80 models, which is the typical size of datasets used for global quality assessment [119, 56].

For these goals, we selected a small set of surface models exhibiting very different shapes, containing both smooth and rough parts, and associated with different sampling densities. All the models are formed by closed 2-manifold triangular meshes. For the distortion we have considered realistic processing operations (#1 to #3 and #9 to #11 in Table 5.1):

- *Watermarking* - Three different algorithms are used, they respectively modify the volume moments [123], the low frequency spectral coefficients [124] and the vertex norms [17] (i.e. their distance to the centroid of the model). The introduced distortions are respectively smooth deformations [123, 124] and structured ring-like patterns [17].
- *Compression* - We consider uniform geometric quantization, the most common lossy process of compression algorithms.
- *Simplification* - We consider the QEM algorithm [37].

On top of these realistic distortions we also consider *noise addition* and *smoothing* [110], in order to emulate further complex processings. These distortions are applied in a local manner on selected regions of the object, either rough or semi-rough. The concerned stimuli objects (#4 to #8 in Table 5.1) have been picked from the LIRIS General Purpose subjective database. All the distortions are prestented in *Appendix A*. As stated above, the idea is to create a compact set of challenging and diversified stimuli (11 in total). We still have introduced a certain amount of consistency: we conducted the same attack (noise addition) on different models and different attacks on the same model (Egea), in order to be able to draw conclusions about these cases. This set of challenging stimuli constitute a good base to analyze complex perceptual mechanisms (e.g. frequency sensitivity, masking effect) and to create a robust metric.

5.2.3 Participants and procedure

The participants were shown each distorted object next to its reference version, side by side on the same screen. They were able to fully interact with the models (rotation, zoom and translation) and to change the size of the painting brush.They were instructed to mark the regions (on the distorted model) were they could see any visible difference with respect to the reference one, and were given no time limit for the deliberation. The 11 pairs of models were presented in random order to prevent any order bias. Before starting the test session, each participant was subject to a training phase to make him familiar with 3D object manipulation

ID	Original Model (vertex number)	Distortion type	Method & ref.	Settings	MRMS (10^{-3})
#1	Egea (100K)	Watermarking	Volume moment [123]	75 bits inserted, α=0.7	0.85
#2	Egea (100K)	Watermarking	Spectral [124]	16 bits inserted	0.59
#3	Egea (100K)	Watermarking	Vertex norms [17]	64 bits inserted, method I	0.08
#4	Egea (50K)	Smoothing	Taubin [110]	30 iterations, rough areas	0.33
#5	Egea (50K)	Noise	Uniform	$\alpha = 0.9\%$, rough areas	0.16
#6	Egea (50K)	Noise	Uniform	$\alpha = 0.9\%$, intermediate areas.	0.16
#7	Egea (50K)	Smoothing	Taubin [110]	30 iterations, intermediate areas	0.26
#8	Dinosaur (42K)	Noise	Uniform	$\alpha = 0.75\%$, intermediate areas	0.17
#9	Vase Lion (39K)	Quantization	Uniform	9 bits	0.33
#10	Bunny (25K)	Simplification	Garland & Eckbert [37]	50% removed	0.15
#11	Bunny (25K)	Simplification	Garland & Eckbert [37]	80% removed	0.29

Table 5.1: Details of our dataset. MRMS stands for Maximum Root Mean Square error, calculated using the Metro tool [20].

(rotation, scaling, translation) and with the mesh painting tool.

Three different sizes of screen were used: 14, 19 and 22 inches due to the limit of our experimental equipment. Given the fact that each pair of stimuli (distorted and reference) spanned the full screen, each stimuli thus subtended at least 15 degree of visual angle horizontally. 20 participants took part to our experiment (3 females and 17 males), they were aged from 16 to 27 years. They were mostly students and staffs from University of Lyon and from the Northwestern Polytechnical University of Xi'an, China. The experiment was conducted, on purpose, in uncontrolled viewing conditions (to reflect real-world conditions). This pool of observers contained both computer graphics experts and naive subjects.

The time to mark all 11 objects was 90 minutes on average (with a standard deviation of 30 minutes). This may seem long (e.g. 30 minutes is usually recommended for a rating experiment [47]), however observing and painting a small number of models is far less tiring that rating hundreds of stimuli, because the full interactions between observers and objects alleviate boredom compared to hundreds times of repetitive ratings. It is interesting to observe that these painting times highly depend on the objects: 5 minutes on average were necessary to mark each simplified Bunny object (#10 and #11) while an average of 12 minutes was taken for the Vase Lion (#9). These times seem to depend both on the object shape and attack complexity. An example of distortion map obtained from the human subjects is shown in figure 5.1 (object #8). It results from averaging the 20 observers' binary maps and reflects the local probabilities of artifact detection. In this example, more than

50% of the observers have detected the geometric noise added on the neck and the ribs of the dinosaur.

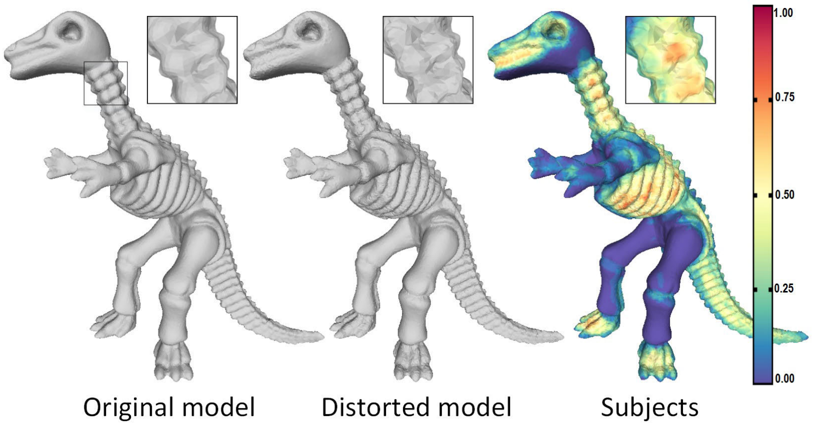


Figure 5.1: Example reference and distorted models from our dataset (object #8), along with human-perceived distortions (color coded probabilities of artifact detection)

5.2.4 Analysis of subjective dataset

Observer agreement

Before exploiting the results of our experiment, it is critical to analyze the agreement between participants. Indeed, a high disagreement would mean that observers have randomly made their choices and thus would imply that the task was too difficult or ill-defined. To assess this inter-observer agreement we first consider the Kendall's coefficient of agreement [52]. This coefficient, noted as u, is usually used for assessing the consistency of votes in paired comparison studies [62, 93]. It ranges from 1 (perfect agreement) to $-1/(n-1)$, if n is even, and $-1/n$, if n is odd, where n is the number of observers. In their experiment related to image local distortion assessment, Čadík et al. [9] computed this coefficient per pixel; following this idea we compute it per vertex by considering the number of participants that have chosen the *distorted* choice over the *not-distorted* one and vice versa. We can then derive an coefficient $\overline{u}$ by averaging all vertex coefficients over each object. Still as in [9], and considering the fact that the high number of unmarked vertices will push the coefficient toward high values, we also compute $\overline{u_{mask}}$ for which we average only vertices marked as distorted by at least 5% of the observers. Results are detailed in table 5.2. The values of $\overline{u}$ and $\overline{u_{mask}}$ averaged over all the objects are respectively 0.60 and 0.43. That constitute good agreement values very close to those from Čadík et al. [9] (0.78 and 0.41).

We could consider that these values sufficiently assess the agreement of the observers and thus the reliability of our experimental results. However, computing the agreement for each vertex separately may lose some important information about the randomness of the observer's choices. As an illustration, imagine that the observers each vote for a different subset of the mesh vertices, then $\overline{u}$ (as well as $\overline{u_{mask}}$) will be high since all the vertices will be chosen as unmarked by n or $n-1$ observers. This high $\overline{u}$ value may be considered as correct because observers mostly agree on the fact that no vertex seems distorted. However, in that case we could also consider that their votes are random and thus that their agreement should be low. To face this issue we also computed a more global statistical agreement indicator: the Krippendorff's α coefficient [54]. This indicator is able to measure the agreement among an arbitrary number of observers assigning values to unstructured phenomena. It ranges from 0 (observers agree as if chance had produced the results) to 1 (observers agree perfectly). It is actually a generalization of several statistical indicators (e.g. for two observers and ordinal data, it is identical to the Spearman's rank correlation coefficient). For each object, we do not compute a single α value but the distribution of α by bootstrapping (as recommended by the author [54]) and we thus obtain a mean value and the associated confidence interval. Table 5.2 details these values as well as the 5^{th} and 95^{th} percentiles. The α averaged over all the object is 0.102. This is a relatively low value suggesting that the task was indeed difficult. However, the percentiles show that these values are statistically significant so there really exists an agreement among the observers, suggesting that their task has been reliably performed. Local distortion marking is still a very unusual task in subjective quality assessment, hence it is hard to judge what can be considered as a typical α value for this kind of data. What we can assert is that our observer agreement is very similar to the one observed in the image localized distortion experiment from [9] (this is assessed by the similar $\overline{u}$ and $\overline{u_{mask}}$ values and illustrated, in the appendix A, by the similarity in the obtained maps). We argue that the Krippendorff's α should become a gold standard to evaluate the observer agreement for localized distortion experiments.

When looking at the per-object agreement in table 5.2, we can observe that it varies from 0.011 to 0.218 (Krippendorff's α). These values directly reflect the difficulty for the observers to perceive the artifacts. For instance, the noise added on the Egea model (#5 and #6) is quite easy to detect ($\alpha > 0.20$). On the contrary, a smooth deformation due to watermarking (#1) is far more difficult to perceive ($\alpha = 0.029$). These agreement values allow us to detect two problematic cases: object #2 owns a very high $\overline{u}$ and a very low α, and object #11 owns both low $\overline{u}$ and α values. The artifacts of object #2 are of very low frequency (see figure 5.3) and almost invisible, thus observers have marked very few and almost random vertices. On the contrary object #11 is too much simplified, hence observers have painted almost all vertices. These two special cases will thus be removed in some of our evaluation experiments.

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

Table 5.2: Inter-observer agreement statistics, in terms of Kendall's u and Krippendorff's α coefficients

Observations

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

Visual masking defines the reduction in the visibility of one stimulus due to the simultaneous presence of another. For 3D surfaces, this phenomenon points out the maximized visibility of geometric distortions on smooth regions rather than on rough ones. This effect is perfectly illustrated by object #9: whereas the same uniform 9-bits quantization is applied for all vertices, we can observe that the subjective artifact probability is much higher on the face of the Lion which is rather smooth, than on the mane which is rough (see figure 5.2).

The human visual system is also sensitive to the frequencies of a visual stimulus. Studies of this phenomenon have led to define the *contrast sensitivity function (CSF)*, characterized by a band-pass filter with a peak frequency between 4 and 6 cpd (cycles by degree) and a

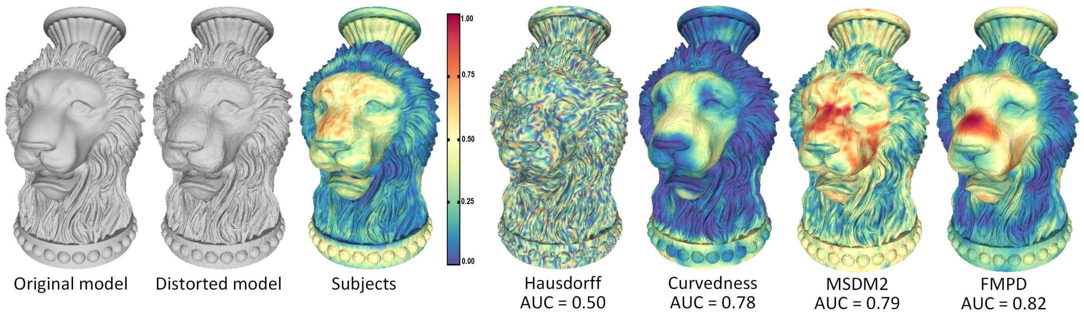


Figure 5.2: Reference and distorted models from our dataset (stimulus #9) along with human-perceived distortions (color-coded probabilities of artifact detection) and distortion maps from different metrics (colors mapped in the min-max range). The distortion prediction performance of each metric is given in term of Area Under ROC Curve (AUC).

quick drop on each side of this peak. For 3D shapes, this generally leads to observe that high-frequency distortions have a much higher probability to be visually noticeable than low-frequency ones. This *frequency sensitivity* is particularly well illustrated when looking at objects #2 and #3. Their distortions are respectively of very low and very high frequency (see figure 5.3). Whereas the *power* of the distortion is much higher in object #2 (as reflected by the MRMS values - 0.59 against 0.08), the number of vertices marked as distorted is much lower: on average observers have marked 1.2% of the vertices against 5.4% for object #3.This is also illustrated by the observer maps in figure 5.3.

5.3 Set of geometric attributes and metrics

Our dataset provides the opportunity to analyze the performance of existing metrics and well-known vertex attributes for the task of visual difference prediction. We also select a set of low-level geometric attributes commonly used in perceptually-based geometry processing algorithms. We implemented 12 features detailed below.

- *Curvatures* (Minimum ($a_1 = K_{min}$), Maximum ($a_2 = K_{max}$), Mean($a_3 = (a_1 + a_2)/2$) and Gaussian ($a_4 = a_1 * a_2$) per vertex respectively) - Following the solution adopted in [1], we evaluate the curvature tensor using the method from Cohen-Steiner and Morvan [22] on a fixed-size neighborhood around each vertex, *i.e.* a geodesic disk approximated by the intersection of the surface with a sphere centered at the vertex.

- *Shape Index* [53] - $a_5 = 2/\pi * arctan[(a_2 + a_1)/(a_2 - a_1)]$.

- *Curvedness* [53] - $a_6 = \sqrt{(a_1^2 + a_2^2)/2}$.

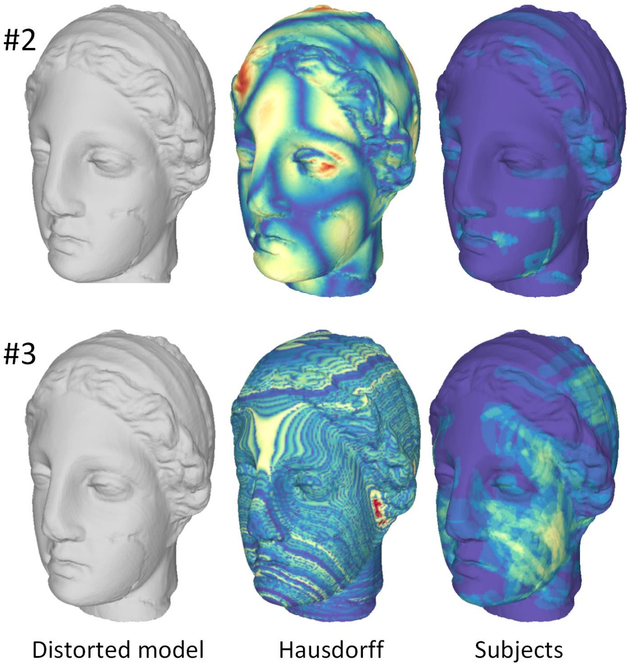


Figure 5.3: Distorted models #2 (low frequency watermark) et #3 (high frequency watermark) and their Hausdorff and observer maps. Low frequency distortions are much less visible.

- The *3D Geometric position* a_7 - $a_7(v_i) = P_i$ is the position of each vertex v_i .

- *Mean dihedral angle* a_8 - Since all the objects are 2-manifold meshes, the mean dihedral angle is then computed for each edge and corresponds to the angle $D < t_j, t_{j+1} >$ between the normals of its adjacent faces t_j and t_{j+1}. The computed values are then averaged for each vertex v_i. This attribute was used in the perceptual metric from Váša and Rus [119]. At each vertex v_i:
  $a_8(v_i) = \frac{\sum_j D<t_j,t_{j+1}>}{N_i^e} \quad for \quad j \in N_i^e$
  N_i^e is the number of adjacent edges around the center vertex v_i.

- *Normal* a_9 - This attribute was considered in several simplification algorithms such as QEM [37] and VSA [21]. At each vertex v_i, it is computed as:
  $a_9(v_i) = \frac{\sum_j \vec{n_j}}{||\sum_j \vec{n_j}||} \quad for \quad j \in N_i^T.$
  N_i^T is the number of adjacent facets around the center vertex v_i.

- *Mesh Saliency* a_{10} - We have implemented one scale saliency (not the multi-scale one) from Lee et al. [63], based on a difference of Gaussian operator. At each vertex v_i, it is computed as:
  $a_{10}(v_i) = |G(a_3(v_i), \sigma_i) - G(a_3(v_i), 2\sigma_i)|$

$G(a_3(v_i), \sigma)$ is defined as the Gaussian-weighted average of the mean curvature per vertex, σ is a distance to the center vertex v_i.

- *Laplacian* a_{11} - The relevance of the difference of geometric Laplacian regarding the perceived artifact visibility was raised by Karni and Gotsman[48]. It is computed at each vertx v_i: $a_{11}(v_i) = \frac{\sum_j (P_i - P_j)}{N_i^T}$. P_i is the position of center vertex v_i, P_j is the position of its adjacent vertex v_j. N_i^T is the number of adjacent facets around the center vertex v_i.

- *Laplacian of Gaussian curvature* a_{12} - It corresponds to the Laplacian operator applied on the Gaussian curvature field. This attribute was used in the perceptual metric from Wang et al. [125]. Note that the authors use the *cotangent weights* discretization of the Laplace operator while we consider the simpler graph Laplacian, defined as *LR* (Local Roughness) in their work.

Geometric positions, normals and geometric Laplacian are 3D vectors, while the others are scalar values. Besides most simple geometric features (position and normals), we selected attributes either commonly used in mesh quality metrics [56, 119, 125] or pointed as *relevant* by previous learning-based works related to perception [6, 16].

Our objective is to predict the local visual distortion in a *full reference* scenario. Hence, for this task, we consider not the attributes themselves but *differences* of attributes. To compute these differences, we first establish a correspondence between the distorted mesh $\mathscr{D}$ and the reference one $\mathscr{R}$. For that purpose, we simply perform a fast projection of the vertices from $\mathscr{D}$ onto the surface of $\mathscr{R}$. For each vertex v of the distorted mesh $\mathscr{D}$ (associated with given attribute $a(v)$), we compute its nearest 3D point $\hat{v}$ on the surface of the reference model $\mathscr{R}$ using the efficient AABB tree structure [2] from the CGAL library. For each 3D point $\hat{v}$, its attribute $a(\hat{v})$ is interpolated from the triangle it belongs to using barycentric coordinates. The features that we consider are thus the local per-vertex differences of each attributes and are computed as follows:

$$f_v = ||a(v) - a(\hat{v})|| \tag{5.1}$$

where $||\cdot||$ represents Euclidean norm.
Additionally to this simple difference, we also consider a *normalized* difference, defined as follows:

$$f_v^N = \frac{||a(v) - a(\hat{v})||}{max(||a(v)||, ||a(\hat{v})||) + K}. \tag{5.2}$$

where $K = 1$ is a constant to avoid instability when denominators are close to zero. The denominator has goals, (1) it normalizes the features f_v (between 0 and 1), which is necessary for further machine learning and (2) it acts as a visual masking filter (e.g. it gives more weight to a same curvature variation in a flat region than in rough or highly curved ones). Of course, for vector attributes (geometric position and normal) this denominator is discarded. The geometric position difference is normalized by the bounding box size.

To strengthen the robustness of our metric, we consider each feature at several scales; indeed on top of each *per vertex* attribute values $a(v)$, we consider also Gaussian-weighted averages $a^h(v)$ computed as follow:

$$a^h(v) = \sum_{v_i \in N(v,h)} w_v^h(v)\, a(v) \tag{5.3}$$

$\mathcal{N}(v,h)$ is the connected set of vertices belonging to the sphere with center v and radius h (including the intersections of edges with the sphere) and $w_v^h()$ is a Gaussian weighting (an isotropic weighting) function centered on v with standard deviation of $h/2$ (i.e. the size h of the neighborhood is twice the standard deviation of the Gaussian filter).

Besides these attribute differences, we also consider three well known metrics, commonly used to drive or evaluate geometry processing operations, and which produce distortion maps:

- The *Hausdorff distance* - Each vertex of the distorted mesh is associated with its Hausdorff distance to the reference surface during the first step the computation. This measure is commonly used to drive or evaluate geometry processing operations.
- *MSDM2* [56] - This metric computes multi-scale differences of curvature statistics between vertices of the distorted mesh and their projections on the reference one, at multiple scales. These local differences are then pooled into a single global score. Hence, this metric natively produces a distortion map. However, while it has shown to provide excellent results for global quality prediction, this metric has never been evaluated for the task of localized distortion visibility prediction.
- *FMPD* [125]- Like MSDM2, this perceptual metric has proven to be an excellent predictor of the global quality of 3D meshes but has never been evaluated for the task of visible difference prediction. The original FMPD computes a single difference of global roughness value between the meshes to compare. Here, we consider local differences of their local roughness estimator, based on the Laplacian of Gaussian

curvature modulated by masking effect functions. The algorithms are well detailed in [125]. This local roughness difference is also subject to normalization and averaging (see eq. 5.1, 5.2 and 5.3).

For each attribute (including the local roughness difference of FMPD), we consider two scales $h = \{0.012R, 0.036R\}$ (R is the maximum length of the bounding box) in addition to the *per vertex* version (the saliency does not have per vertex version), and then get $13 \times 3 - 1 = 38$ features. Besides these normalized features, we also take their non-normalized forms as the other 38 features. These scales correspond to those used in [63]. Finally, including the MSDM2 and Hausdorff computations, we therefore have 78 features in total.

5.4 Evaluation of geometric attributes and metrics

5.4.1 Performance measure

As in [9], our purpose is to evaluate the metric/attribute ability of predicting the problematic regions in terms of binary classification: marking the vertices that contain noticeable distortions. Therefore, we use the receiver-operator-characteristic (ROC) that represents the relation between probability of false positives (x axis, meaning the vertex classified as *distorted* while it is not) and probability of true positive (y axis, meaning the vertex correctly classified as *distorted*) by varying a decision threshold on the metric (resp. attribute difference) output. To determine the binary ground-truth data, it is convenient to assume that a certain percentage of observers need to mark the distortion to consider it *noticeable*. In our case, the binary ground-truth data are obtained by considering as *distorted* vertices marked by more than 25% of the observers. The area under the ROC curve (AUC) can then be used as a direct indicator of the performance (1 corresponds to a perfect classification while 0.5 corresponds to a random one).

For each feature/metric we compute the ROC performances per-object which are then averaged on the whole database. The AUC values are then averaged over the 11 models. To test how the metrics can adapt to several objects and distortion types together, we also compute the ROC curve for all objects together. For this latter experiment, we duplicate vertices of smallest models in order to balance the respective importance of each object in this global ROC calculation.

Attributes	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
Minimum	0.88 (1,N)	0.92 (2,S)	0.71 (2,N)	0.83 (1,S)	0.87 (1,S)	0.84 (0,N)	0.79 (0,S)	0.77 (0,N)	0.58 (1,S)	0.60 (0,S)	0.54 (1,S)
Maximum	**0.95** (2,S)	**0.97** (2,S)	**0.86** (2,N)	**0.90** (2,S)	**0.91** (2,S)	**0.89** (2,N)	**0.86** (2,S)	0.74 (0,S)	0.77 (2,N)	**0.73** (2,N)	0.57 (2,N)
Mean	0.93 (1,N)	0.96 (2,S)	0.85 (2,N)	**0.89** (2,S)	0.90 (2,S)	0.87 (2,N)	**0.86** (2,S)	0.78 (1,S)	0.77 (2,N)	0.70 (2,N)	0.56 (1,N)
Gauss	0.86 (1,S)	0.91 (2,S)	0.68 (1,N)	0.84 (2,S)	0.90 (2,S)	0.87 (1,N)	0.81 (2,S)	**0.79** (1,S)	0.54 (1,S)	0.62 (2,N)	**0.58** (0,S)
Shape index	0.87 (1,N)	0.88 (2,S)	0.74 (2,S)	*0.79* (0,S)	0.86 (1,S)	0.82 (1,S)	0.78 (1,S)	0.75 (1,S)	0.71 (1,S)	0.63 (1,S)	*0.53* (1,S)
Curvednesss	**0.95** (2,S)	**0.97** (2,S)	**0.87** (2,N)	**0.89** (2,S)	**0.91** (2,S)	**0.90** (2,S)	**0.85** (2,S)	**0.80** (1,S)	**0.78** (2,N)	**0.73** (2,N)	**0.58** (2,N)
Position	0.90 (2,N)	*0.83* (2,S)	0.68 (2,N)	0.83 (1,N)	0.88 (2,N)	0.86 (2,N)	0.81 (1,N)	0.79 (1,N)	0.59 (1,N)	*0.50* (0,N)	0.56 (0,N)
Dihedral	0.89 (1,N)	*0.59* (2,N)	0.68 (0,N)	0.82 (1,N)	0.86 (1,N)	0.84 (0,N)	0.79 (1,N)	0.76 (0,N)	0.58 (2,N)	0.65 (2,N)	0.57 (0,N)
Normal	0.91 (2,N)	0.93 (2,N)	0.61 (0,N)	0.87 (2,N)	0.88 (1,N)	0.85 (1,N)	0.84 (1,N)	0.77 (1,N)	0.61 (1,N)	0.58 (1,N)	0.60 (2,N)
Saliency	0.85 (2,N)	0.91 (2,N)	0.64 (1,N)	*0.79* (1,S)	0.83 (1,S)	0.79 (1,S)	*0.76* (1,S)	0.67 (2,N)	0.56 (1,S)	0.63 (1,N)	0.57 (2,S)
Laplassian	0.90 (2,S)	0.65 (0,N)	*0.60* (0,N)	0.82 (1,N)	0.87 (1,N)	0.85 (0,N)	0.78 (1,N)	0.77 (0,N)	0.54 (1,N)	0.69 (2,N)	0.63 (2,N)
LaplaceGauss	0.88 (2,S)	0.89 (1,S)	0.71 (2,S)	0.86 (2,S)	**0.92** (2,S)	**0.91** (2,N)	0.83 (1,N)	0.78 (1,N)	0.72 (2,N)	0.56 (0,S)	0.66 (0,S)
FMPD	**0.96** (2,S)	**0.95** (2,S)	**0.87** (2,S)	0.83 (2,S)	0.83 (1,S)	**0.82** (0,S)	0.79 (0,S)	0.74 (0,S)	0.82 (2,N)	0.63 (0,N)	0.66 (0,N)
MSDM2	**0.97**	**0.97**	0.85	0.84	**0.90**	0.88	0.81	0.77	**0.79**	**0.71**	0.54
Hausdorff	*0.76*	0.83	*0.60*	*0.79*	*0.83*	0.80	*0.76*	*0.73*	*0.50*	*0.50*	0.56

Table 5.3: AUC values from best performed attribute-based features including LRF from FMPD, and AUC values from MSDM2 and Hausdorff distance.

5.4.2 Evaluation procedures

In the evaluation section, we firstly used Naive Bayes classier to train a dataset from each procedure detailed below, which is trained using ground-thruth data (supervised learning) for a single feature/metric, then, and get class information (distorted or not for each vertex). Then based on such information, we tested the same dataset to obtain the correlation between the feature/metric and class. To evaluate all the features/metrics robustly and comprehensively, we design three different train and test procedures based on cross validation:

1. For each object separately, we perform a train and test procedure based on the leave-one-out cross validation, evaluating each feature/metric, where the training dataset is

Rank	9Objects indv	Mean AUC	11Objects indv	Mean AUC	Whole 9Objects	AUC	Whole 11Objects	AUC
1	MSDM2	0.84	MSDM2	0.87	FMPD (2,S)	0.88	FMPD (2,S)	0.87
2	Curvedness (2,S)	0.83	Curvedness (2,S)	0.87	Laplace Gauss (2,N)	0.86	Curvedness (1,S)	0.85
3	Maximum (2,S)	0.83	Maximum (2,S)	0.87	MSDM2	0.86	Maximum (1,S)	0.85
4	Mean (2,S)	0.81	LRF(FMPD) (2,N)	0.85	Curvedness (1,S)	0.86	Mean (1,S)	0.84
5	LRF(FMPD) (2,N)	0.80	Mean (2,N)	0.84	Maximum (1,S)	0.86	Gauss (1,S)	0.84
6	Laplace Gauss (2,N)	0.79	Laplace Gauss (2,N)	0.79	Mean (1,S)	0.85	Laplace Gauss (2,S)	0.83
7	Shape indx (1,S)	0.76	Position (2,N)	0.79	Gauss (1,S)	0.83	MSDM2	0.83
8	Normal (1,N)	0.76	Normal (2,N)	0.78	Minimum (1,S)	0.80	Minimum (1,S)	0.82
9	Laplace (1,S)	0.75	Gauss (2,S)	0.76	Dihedral (0,N)	0.79	Saliency (1,S)	0.77
10	Gauss (1,S)	0.75	Shape indx (1,S)	0.76	Saliency (1,S)	0.77	Dihedral (1,N)	0.77
11	Position (2,N)	0.75	Minimum (1,S)	0.74	Normal (0,S)	0.77	Normal (0,S)	0.74
12	Minimum (1,S)	0.74	Laplace (1,S)	0.73	Shape indx (1,N)	0.76	Shape indx (1,S)	0.73
13	Dihedral (0,N)	0.74	Hausdorff	0.73	Laplace (0,S)	0.73	Laplace (0,N)	0.72
14	Saliency (1,N)	0.71	Saliency (1,N)	0.72	Position (2,S)	0.66	Position (1,S)	0.67
15	Hausdorff	0.70	Dihedral (1,N)	0.71	Hausdorff	0.60	Hausdorff	0.59

Table 5.4: From left to right: Ranks of average AUC values for best performing features and metrics among 9/11 models (second and third column); Ranks of AUC values for best performing features and metrics on a model taking vertices from 9/11 as a whole dataset (fourth and fifth column).

feature/metric values on the vertices from 10 models and testing dataset is the same type of feature/metric values from the rest model (similar to the other 10 models). The average ROC curve is recorded as the final performance estimate per model.

2. We perform a train and test procedure based on 10-fold cross validation, evaluating each attribute/metric on the whole set of vertices from the 11 models together.

3. According to the Krippendorff's α values in Table 5.2, we excluded two models #2 and #11, whose α values are smallest and second smallest respectively. And then we perform a train and test procedure evaluating each attribute/metric on the whole set of vertices from the rest 9 models. This procedure, to some extent, guarantees the robustness and credibility of the evaluation by eliminating those too difficult or ill-defined tasks.

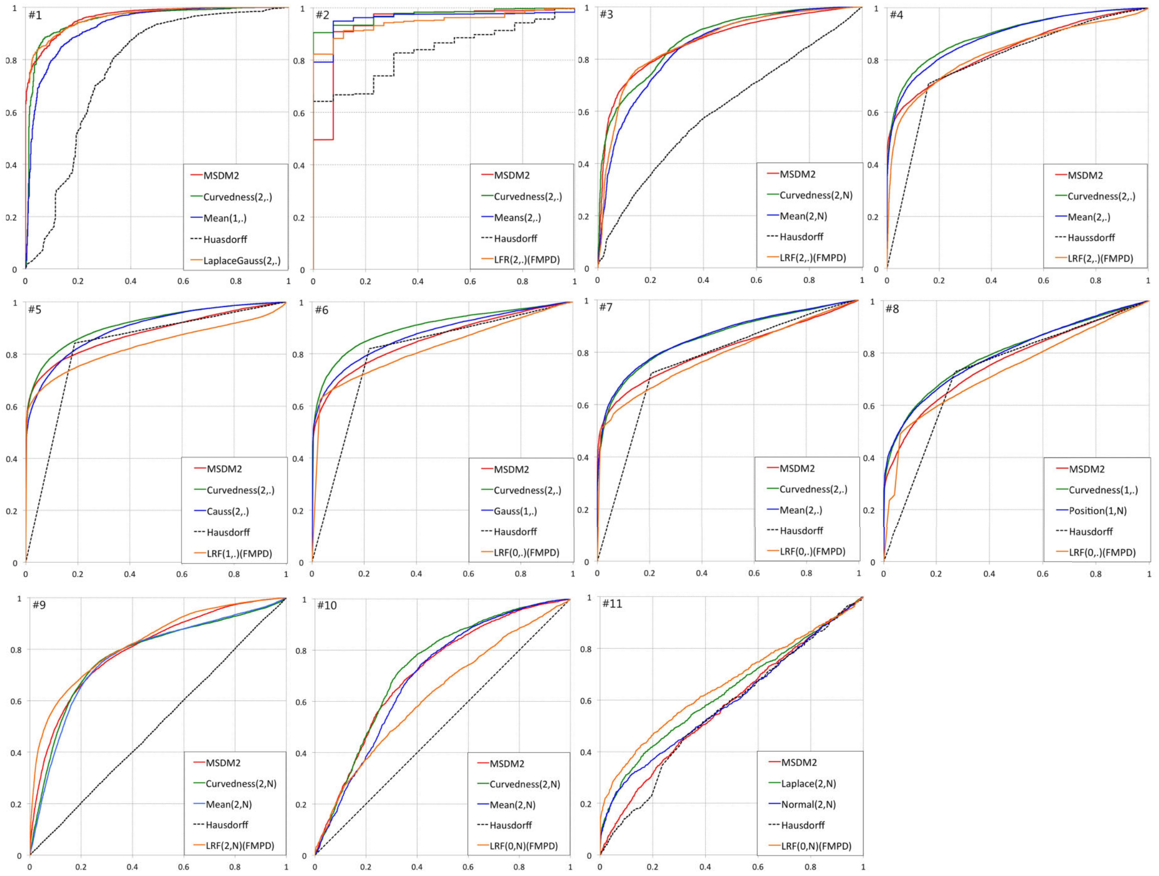


Figure 5.4: the ROC curves from MSDM2, FMPD, Hausdorff distance and 2 other features among best performing ones for each model.

5.4.3 Results and comparisons

Performance of the attributes and metrics

We have trained and tested all the features and metrics with the 3 procedures mentioned in the subsection 5.4.2. Figures 5.2 and 5.5 illustrate some visual results, while table 5.3 and details the quantitative performance in terms of AUC values for each feature and for each model from the dataset. Average values are presented in table 5.4 and figure 5.6 (according to agreement statistics, models #2 and #11 have been excluded from this averaging). Additional results, including #2 and #11, are available in the appendix A. As all the 76 features are derived from 13 attributes, thus, we present the results obtained by the best filter: (i,S) and (i,N) respectively refer to simple (eq. 5.1) and normalized (eq. 5.2) attribute difference, while i refers to the averaging scale (eq. 5.3): 0 means per-vertex attribute and 1, 2 respectively refer to small and large averaging scale.

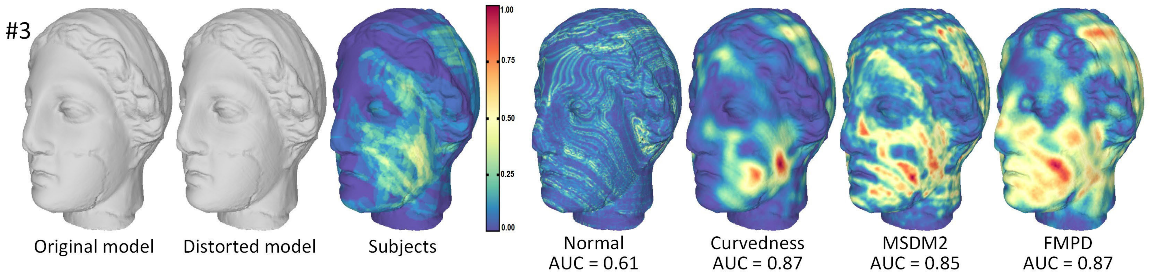


Figure 5.5: Reference and distorted models from our dataset (object #3), observer data and distortion maps from different attributes and metrics (colors are mapped in the min-max range).

For the second and third procedure, Table 5.4 and Figure 5.6-*bottom row* present the performance comparisons.We also computed other statistics of all the results from first procedure. In details, we firstly took an average AUC value of every feature/metric among 11 models and 9 models (excluding #2 and #11) respectively. Then, similar to the work before, we selected the best performing features among same attribute-based features according to the average AUC values. Similarly, for the results from 9/11 models as a whole dataset, we show the results from best performing features among same attribute-based features and FMPD, MSDM2, Hausdorff distance.

In Table 5.4, the second and third column show the ranks according to the average AUC values, the comparisons confirm our previous deduction: MSDM2 performs outstandingly well among all the features/metrics when a certain feature/metric measures 9/11 models separately. And yet, the performance of curvedness-based features is still quite good. This suggests that on *a model whose distortion is unknown*, MSDM2 metric could predict visually correlated distortion better than other feature/metric. In contrast, when we take 9/11 models as whole dataset, the performance of FMPD suddenly rises to the first place. This means that in scenario where one single threshold has to be determined for detecting artifacts for several objects and/or distortion types, then the FMPD metric is the most suited. This may be due to the fact that this metric explicitly incorporates global and local normalization pre-processes (linked to the masking effect). In general, Hausdorff and some geometric attributes-based features (e.g. geometric metric, mean dihedral angle) could not predict the perceived local distortion.

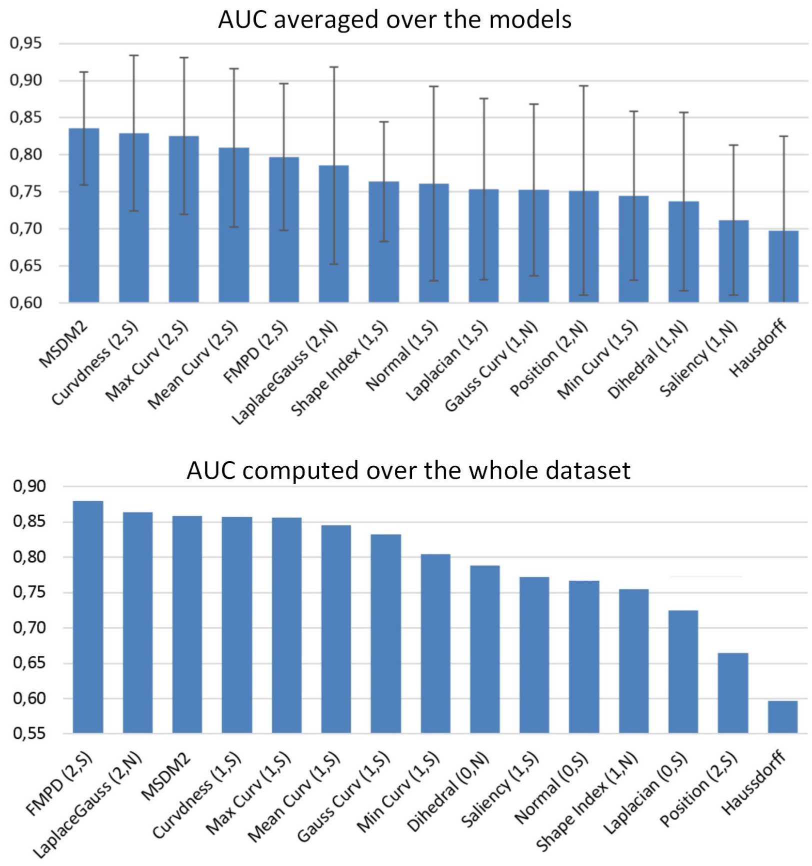


Figure 5.6: *Top*: Mean of AUC values for all the tested features. Error bars denote standard deviations. *Bottom*: AUC values when computing ROC curves over the whole dataset.

To observe directly, for each model, Figure 5.4 shows the ROC curves of the performances from MSDM2, FMPD, Hausdorff distance and 2 other features among best performing ones. With the same scenario, Figure 5.7 shows the ROC curves for 9/11 models as a whole.

Influence of the filters

Table 5.5 illustrates the AUC values averaged over all features and objects, for each filter separately. We run paired t-test with confidence level 0.95 to evaluate the effect of each filter. The normalization does not have a significant impact ($p\text{-}value = 0.11$). However Scale 1 is significantly better than Scale 0 ($p\text{-}value = 6 \times 10^{-8}$) and Scale 2 is significantly better than Scale 1 ($p\text{-}value = 0.02$). The conclusion to draw here is that averaging the features has a significant impact on their ability to predict the visible distortion. The main reason is that

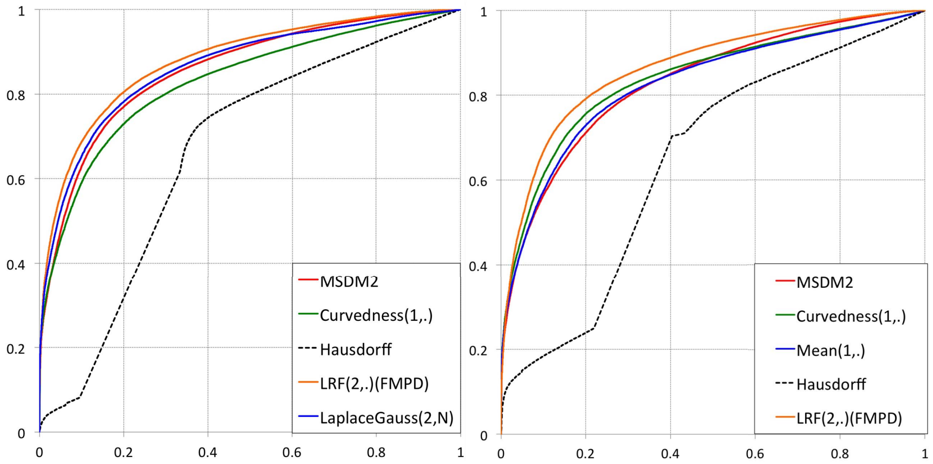


Figure 5.7: the ROC curves from MSDM2, FMPD, Hausdorff distance and 2 other features among best performing ones on a model taking vertices from 9 as a whole dataset(left) and on a model taking vertices from 11 as a whole dataset (right).

such averaging (on geodesic neighborhoods) makes the computation more independent of the sampling density.

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

Table 5.5: AUC values averaged over all features, for each filter separately.

5.4.4 Discussion

Several conclusions can be drawn from the results presented above:

- Previous studies [58, 24] showed that geometric distances (e.g. Hausdorff and root mean square error) are not good predictors of the perceived global visual quality. Our study completes this analysis by demonstrating that the geometric distance (Hausdorff as well as averaged position difference) is not a good predictor of the local visible difference.

- Recent metrics like MSDM2 [56] and FMPD [125] provide the best results for this task of visible difference prediction; these good performances are mainly due to two key features: (1) being multi-scale improves the robustness, and (2) smart masking filters improve the stability across different models and distortions.

- Simple curvature measures (e.g. the curvedness difference) may produce excellent results. However, we noted unstable performance with a high and unpredictable effect of the normalization. For instance, AUC values for model #9 are respectively 0.78 and 0.67 for curvedness $(2,N)$ and curvedness $(2,S)$ (see the appendix A).
- Predicting the perceived local distortions on the surface of a 3D mesh appears to be a very challenging task. Figures 5.2 and 5.5 illustrate the fact that even best metrics are still far from perfectly predicting the ground-truth distortion maps. Hence, much works remain to be done in this area and our dataset is an excellent basis for that.

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

5.5 Conclusion and perspective

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

These publicly available dataset and results[1] constitute the very first steps toward the efficient prediction of the perceivability of geometric artifacts, and more broadly toward the understanding of local perceptual mechanisms involving 3D geometry. In particular, this dataset will be very useful to evaluate and/or to train future metrics. Machine-learning is appearing as a powerful tool to design full reference image quality metrics [78, 10, 13] and

---

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

could indeed provide good results for 3D mesh artifact localization. However, we believe that such top-down approach should also integrate low-level human vision processes (such as a model of contrast sensitivity function) to be really efficient.

Our study focused on the geometry and thus considered a simple rendering style. We now plan to investigate how complex textures and shaders, used in modern rendering pipelines, influence the visibility of artifacts. Recent studies about the influence of material [35] and lighting [31] on the shape perception, should bring highly relevant cues for this goal.

Chapter 6

Subjective and objective visual quality assessment of textured 3D meshes

6.1 Introduction

In the previous chapter, the objects we study are primary 3D surface meshes without texture. However in practice, texture mapped 3D graphics are now commonplace in many applications, including digital entertainment, cultural heritage and architecture. They consist of geometric surfaces, on which are mapped several texture images that serve to make the rendering more realistic. Common texture maps include diffuse map, normal map and specular map. After their creation (by a designer or using a scanning/reconstruction process), these textured 3D assets may be subject to diverse processing operations including simplification, compression, filtering, watermarking and so on. For instance, with the goal of accelerating the transmission for remote Web 3D visualization (e.g., for a virtual museum application), the geometry may be simplified and quantized, and the texture maps may be subject to JPEG compression. Similar geometry and texture degradations may also occur when these assets have to be adapted for lightweight mobile devices; in that case, textures may have to be sub-sampled or compressed using some GPU-friendly random-access methods (e.g., [107]). These geometry and texture content corruptions may impact the visual quality of the 3D asset. Therefore, there is a need for efficient perceptual metrics to evaluate the visual impact of these textured model artifacts on the visual quality of the rendered image.
Many visual quality metrics have been introduced in the field of computer graphics. However, most of them can only be applied on images created during the rendering step. They mostly focus on detecting artifacts caused by global illumination approximation or tone mapping [3, 45, 135, 10]. On the contrary, another class of method focuses on evaluating the artifacts

introduced on the 3D assets themselves. However most consider only geometric distortions [56, 119, 125]. Little work has been done to evaluate the visual impact of both geometry and texture distortions on the appearance of the rendered image. Studying the complex perceptual interactions requires a ground-truth of subjective opinions on a variety of models with such degradations. To the best of our knowledge, only Pan et al. [81] conducted such a subjective study. However, they considered only geometry and texture sub-sampling distortions. In this chapter, we present a large-scale subjective experiment for this purpose, based on a paired comparison protocol. As in [81], we restrict the texture information to the diffuse maps. TOur dataset contains more than 272 videos of animated 3D models created from 5 reference objects, 5 types of distortions, and 2 rendering parameters. The experiments involved more than 100 people. After an analysis of the influence of lighting on the perception of geometry and texture artifacts, we then use this subjective ground-truth to evaluate the performance of a large set of state-of-the-art metrics (dedicated to image, video and 3D models). We finally propose a new metric based on an optimal combination of geometric and image measures.
The rest of this chapter is organized as follows: 6.2 describes our subjective experiments and their results. Section 6.3 presents a comprehensive evaluation of state-of-the-art image and mesh metrics with respect to our subjective ground-truth, and details our proposed perceptual measure and their validation. Finally, concluding remarks and perspective works are given in Section 6.4.

6.2 Subjective experiment

We conducted a large-scale subjective experiment to evaluate the visual impact of texture and geometry distortions on the appearance of textured 3D models. We chose a paired comparison technique, where observers are shown two stimuli side by side and are asked to choose the one that is most similar to the reference (forced-choice methodology). This protocol was shown to be more accurate than others (e.g. single stimulus rating) due to the simplicity of the subjects' task [77]. This section provides details on the subjective study and its results

6.2.1 Stimuli generation

We selected 5 textured triangle meshes created using different methodologies and targeting different application domains (see Table 6.1 and Figure 6.1). The *Hulk* and *Sport Car* are artificial models created using a modeling software. They have been selected from a community model repository (ShareGC.com). They both have small numbers of polygons, structured texture content and smooth texture seams (i.e., vertices associated with multiple texture coordinate pairs). The *squirrel* and the *Easter Island statue* come from a reconstruction process using multiple photographs, they are courtesy of the EPFL Computer Graphics and Geometry Laboratory. Finally, the *Dwarf* is a scanned model, courtesy of the Visual Computing Laboratory of ISTI-CNR, Pisa (http://vcg.isti.cnr.it). These three latter models, created respectively from reconstruction and scanning, exhibit noisier texture seams and content. The numbers of vertices from the five models goes from 6,000 to 250,000 and the texture size goes from 256×256 to 4096×4096. Note that the *Hulk* and *Sport Car* are associated with several texture images (resp. 2 and 15). To summarize, these objects span a wide variety of geometry and texture data.

	#Vertex	#Texture	Texture Size	Average Curv.	Text. Charac.	Map. Complex.
Squirrel	6,185	1	2048×2048	High	High freq. & Noisy	Simple
Hulk	10,236	2	1024×1024 & 512×512	High	Structured	Simple
Statue	104,019	4	355×226 to 4096×4096	Low	Noisy	Complex
Sport Car	122,873	15	256×256 to 1024×768	Low & sharp edges	Structured	Simple
Dwarf	250,004	1	4096×4096	Intermediate	Intermediate	Complex

Table 6.1: Details about our 3D models.

Figure 6.1: 5 Models used in the subjective study (from left to right: Sport Car, Easter Island statue, squirrel, Hulk and Dwarf).

These reference models have been corrupted by five types of distortions (three applied on the geometry and two applied on the texture), applied each with four different strengths:
On the geometry:

- **Compression** - We consider uniform geometric quantization, the most common lossy process of compression algorithms.

- **Simplification** - We consider the *Quadric Error Metric* algorithm from Garland and Heckbert [37].
- **Smoothing** - We consider Laplacian smoothing [110].

On the texture map:

- **JPEG** - The most commonly used algorithm for the lossy compression for 2D images.
- **Sub-sampling** - We reduce the texture size by resampling through bilinear interpolation.

The strength of these distortions was adjusted manually in order to span the whole range of visual quality from imperceptible levels to high levels of impairment. For this task, a large set of distortions were generated and viewed by the authors and a subset of them that spanned the desired visual quality (i.e., “Excellent”, “Good”, “Fair” and “Poor”) were chosen to be included in the database. This perceptual adjustment of the distortion strength was also done for the LIVE Video Quality Database [99]. As stated by the authors, it allows us to test the ability of objective metrics to predict visual quality consistently across varying content and distortion types. We thus generated 20 distorted models (5 distortion types × 4 strength) per reference object.

To challenge the generalization ability of the objective metrics, we also included mixed distortions. For this task, we manually selected 36 distorted versions among the $12 \times 8 = 96$ possible combinations of the 12 geometry and 8 texture distortions detailed above. In practice, we applied the 20 single-type distortions on the *Hulk*, *Sport Car*, *squirrel* and *Easter Island statue* model, and mixed 36 mixed distortions on the *Dwarf* model, resulting in a database of 116 models (+5 references), including the reference. Table B.1 provides details on the distortion parameters while Figure 6.2 and 6.3 illustrate some visual examples.

6.2.2 Rendering parameters

User interaction: In existing subjective studies involving 3D content, different ways have been used to display the 3D models to the observers, from the most simple (as static images, as in [132]) to the most complex (by allowing free rotation, zoom and translation, as in [23]). While it is important for the observer to have access to different viewpoints of the 3D object, the problem of allowing free interaction is a cognitive overload which may alter the results. A good compromise is to use animations, as in [81]. For each object of our database, we generate a low-speed rotation animation around the vertical axis.

Figure 6.2: Examples (*Sport Car* and *squirrel*) of distorted models from our dataset. The rendering is the same as in our videos.

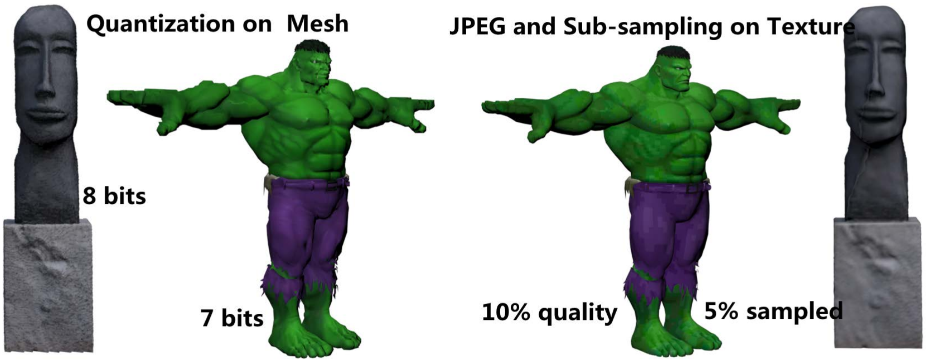


Figure 6.3: Examples (*Hulk* and *Easter Island statue*) of distorted models from our dataset. The rendering is the same as in our videos.

Lighting and shading: As noticed by Rogowitz and Rushmeier [89] the position and type of light sources have a strong influence on the perception of the artefacts. Lighting from the front have a masking effect, hence we chose an indirect illumination. Sun and Perona [108] showed that people tend to assume light is above and slightly to the left of the object when they interpret a shaded picture as a 3D scene. Their observations have been confirmed by O'Shea et al. [79] who demonstrated that the viewer's perception of a 3D shape is more accurate when the angle between the light direction and viewing direction is 20-30 degree above the viewpoint and to the left of 12 degree from vertical. We follow this lighting condition by putting a spot light at this position. For the material, we kept the original parameters from the source reference objects, which are mainly diffuse. The video size is 1920 × 1080. The duration of the video is 15 seconds for the single-type distortions and 10 seconds for the mixed-distortions (for this latter case, videos have been made shorter in order

ID	**Distortion type**	**Squirrel**	**Hulk**	**Statue**	**Sport Car**	**Dwarf**
L1	Smoothing	1 iteration	1 iteration	10 iterations	1 iteration	15 iterations
L2	Smoothing	3 iterations	2 iterations	20 iterations	2 iterations	25 iterations
L3	Smoothing	5 iterations	3 iterations	30 iterations	3 iterations	40 iterations
L4	Smoothing	7 iterations	4 iterations	50 iterations	4 iterations	50 iterations
Si1	Simplification	50% removed	30% removed	50% removed	50% removed	80% removed
Si2	Simplification	70% removed	40% removed	70% removed	60% removed	92% removed
Si3	Simplification	75% removed	50% removed	87.5% removed	75% removed	97.5% removed
Si4	Simplification	87.5% removed	70% removed	95% removed	87.5% removed	98.7% removed
Q1	Quantization	10 bits	10 bits	10 bits	10 bits	11 bits
Q2	Quantization	9 bits	9 bits	9 bits	9 bits	10 bits
Q3	Quantization	8 bits	8 bits	8 bits	8 bits	9 bits
Q4	Quantization	7 bits	7 bits	7 bits	7 bits	8 bits
J1	JPEG	18% quality	18% quality	80% quality	10% quality	12% quality
J2	JPEG	14% quality	14% quality	16% quality	5% quality	10% quality
J3	JPEG	10% quality	10% quality	12% quality	3% quality	8% quality
J4	JPEG	6% quality	8% quality	8% quality	1% quality	6% quality
Su1	Sub-sampling	40% sampled	40% sampled	25% sampled	50% sampled	10% sampled
Su2	Sub-sampling	30% sampled	30% sampled	20% sampled	20% sampled	8% sampled
Su3	Sub-sampling	20% sampled	20% sampled	10% sampled	10% sampled	5% sampled
Su4	Sub-sampling	10% sampled	10% sampled	5% sampled	5% sampled	3% sampled

Table 6.2: Details about the distortions applied on each reference model. For the *Dwarf* model, details about distortions are available in *Appendix B*.

to keep a reasonable duration for the comparison experiment).

In order to study the influence of shading on the results, we also re-generated the same set of videos by keeping only the reflectance (i.e. without shading). Examples are shown in Figure 6.4. We thus obtain 10 video sets to rate (5 models $\times$ 2 rendering settings), for a total of 232 videos.

6.2.3 Experimental procedure

As stated above, we opted for a paired comparison methodology since it has been demonstrated to be more reliable than rating methods [77]. The participants are shown two videos of distorted models at a time, side by side, and are asked to choose the one that is most similar to the reference. The observer can replay the videos as much as (s)he wants. For sake of readability, the reference video is not displayed on the same screen but is presented just before the beginning of the comparisons, and can then be viewed at any time on a pop-up window by clicking on a button. The interface was developed in JavaScript, as web platform, and is illustrated in Figure 6.5.

Figure 6.4: Examples of the appearances of distorted models with different lighting and shading settings from our dataset . The rendering is the same as in our videos. Note that for the Dwarf, we present here compound distortions (geometry+texture) from our validation set (see Section 6.3.4).

The main issue with the paired comparison protocol is the large number of possible comparisons: $\binom{2}{20}$=190 per model for the single distortion setting, and $\binom{2}{36}$ = 630 for the mixed distortion one. It is therefore unrealistic to ask a participant to perform a complete test even on a single model. Fortunately, this high number of trials per model can be reduced by using sorting algorithms as recommended in [102, 77]. The idea is to embed a sorting algorithm into the experiment platform; this algorithm then decides in an on-line fashion which pairs of videos to compare based on the previous comparisons.

We introduced a simple yet efficient sorting algorithm for the single-distortion setting. The idea is obtain a global ranking of the 20 stimuli by interleaving them progressively, one distortion type at a time (i.e., compression, simplification, smoothing, JPEG and Subsampling). For a given distortion D, we assume that the distortion strength goes from D_1 (weak) to D_4 (strong). The assumption behind our algorithm is that for a given distortion D, the quality of D_i is always better than D_j for $j > i$ (noted as $Q_{D_i} > Q_{D_j}, \forall j > i$). First, two distortions types Q and J are randomly chosen (in this example, Quantization and JPEG), Q_4 and J_4 are then compared. The index of the *not-selected* video (Q_4) is pushed into a list

Figure 6.5: Illustration of our browser-based interface for the paired comparison task. Pairs of 3D models are presented as videos ; the user selects the one he prefers and then clicks on the *Next* button. The reference model can be displayed at any time by clicking on the *See Original Model* button below the *Next* button.

(*List 1*), as the poorest quality version. In the next trial, the *selected* model from the previous round (J_4) and a distorted model with a *decreased* level from the other type (Q_3) are shuffled and displayed to the user. This process continues until all 8 models are sorted from the worst to best quality. This sorting process is repeated with two other distortions types (Smoothing and Simplification in the example) to form a second list (*List 2*) which is then interleaved with the remaining distortion type (Sub-sampling) and then with *List 1* to obtain the final ranking of the 20 distortions. In our study, the average comparison number was 36 (instead of 190 for the full design). This sorting algorithm is illustrated in *Appendix B*.
For the mixed-distortion setting, the hypothesis $Q_{D_i} > Q_{D_j}, \forall j > i$ does not hold anymore. Hence, we implemented a more classical self-balancing binary tree, as in [77]. The average number of comparison for sorting the 36 distorted models was 140 (instead of 630).

6.2.4 Participants

A total of 101 subjects took part in the experiment, aged between 20 and 55, all with normal or corrected to normal vision. Participants were students and staff from the University of Lyon in France and the University of Alberta in Canada. On average, it took 12 minutes for one observer to finish the experiment for one video set. 89 subjects rated 1 set, 8 rated 2 sets and 4 rated 3 sets. None of these repetitions took place on the same day in order to prevent any learning effect. The experiments were all conducted on the same 15-inch MacBook Pro screen, in a dark room. In total, each of the 10 video sets (5 models × 2 rendering settings) was judged by between 11 and 15 observers (see Table 6.3).

6.2.5 Computing scores

For each video set and each subject, we obtain a global ranking of the n_m distorted models (n_m equals 20 or 36 in our experiments). From this ranking, it is easy to retrieve the full preference matrix ($n_m \times n_m$), by applying the transitive relation: if object A is better than object B and B is better than C, then we can deduce that A is better than C. These per-subject preference matrices can then be summed into a single one (per video set). In this matrix P, each element $P_{i,j}$ represents the number of times the stimulus i was judged to have higher quality than stimulus j (the matrices are available in *Appendix B*). As in [62, 77], we then consider the number of votes received by each stimuli as its quality score, which may then be divided by the number of human subjects n_s for normalization among video sets:

$$s_i = \frac{\sum_{j=1}^{n_m} P_{i,j}}{n_s} \tag{6.1}$$

We then obtained a subjective score for each distorted model with belong to [0,19], for the single distortion type setting and [0,35] for the mixed distortion one. Note that the value 19 (resp. 35) occurs only for models which have been ranked first by every subjects. Note that more sophisticated statistical methods exist for inferring scale values from a preference matrix. We computed scores using Thurstone's Law of Comparative Judgments, Case V [112], which assumes that the observers' choices can be thought of as sampled from a normal distribution of underlying quality scores. Since the values obtained were very close to the simple vote counts described below (more than 0.99 Pearson correlation), we decided to keep the latter.

6.2.6 Analysis and discussion

Observer agreement

It is essential to analyze the agreements between the subjects before studying the results from the experiment. Since each observer outputs a global ranking of the stimuli, the best way to evaluate their agreement is to compute the Kendall's coefficient of concordance W [50] which assesses the agreement among raters. Table 6.3 gives details on the results. W ranges from 1, meaning complete agreement, to 0, meaning no agreement, while the p-value associated with W provides the likelihood of null hypothesis, which means no agreement between all the subjects. Table 6.3 shows that the overall Kendall's W coefficients are at least as large as 0.69, implying a strong agreement among the subjects, confirmed by the very low *p-values*.

	Rendering with shading			**Rendering without shading**		
	Observer Number	Kendall's W	p-value	Observer Number	Kendall's W	p-value
Squirrel	11	0.69	<0.0001	11	0.71	<0.0001
Hulk	11	0.77	<0.0001	13	0.70	<0.0001
Statue	11	0.83	<0.0001	15	0.76	<0.0001
Sport Car	11	0.74	<0.0001	11	0.76	<0.0001
Dwarf	11	0.72	<0.0001	12	0.77	<0.0001

Table 6.3: Agreement between observers (Kendall's W between their ranks) for each reference model and each rendering condition.

Observation

The visual impact of the distortions highly depends on the textured 3D model, as illustrated in Figure 6.6. The geometric quantization, for instance, is less visible on 3D shapes with fewer vertices (like the Squirrel). The reason is that the sampling density of the surface directly influences the frequency of the visual distortions created by the quantization and thus their perceptual impact (related to the contrast sensitivity function). The resolution and the content of the texture also strongly influence the perception of artifacts.

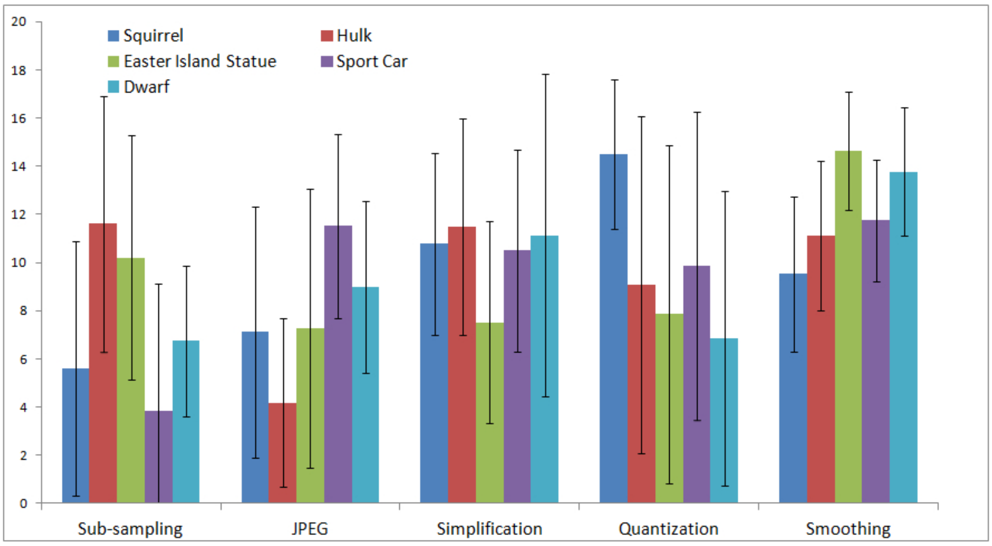


Figure 6.6: Mean scores (averaged among the 4 strengths) for the 5 types of distortions and the 5 models for the rendering with shading. Higher scores mean better visual quality.

Effect of the rendering

In this section, we investigate the influence of the rendering on the visual impact of the distortions. Figure 6.7 presents the quality scores, averaged over the models for the two rendering conditions: with and without shading. As expected, when only the reflectance of the surface is taken into account, the quality scores of the geometric distortions are consistently better since the impact of the geometry on the rendering is basically limited to silhouette. By conducting one-tailed paired t-tests, we found a significant increase of the quality of geometric distortions (p-value=0.012) and a significant decrease of the quality of texture distortions (p-value=0.001). These intuitive results have to be kept in mind, when assessing the quality of textured 3D models. A calibration according to the rendering may be necessary.

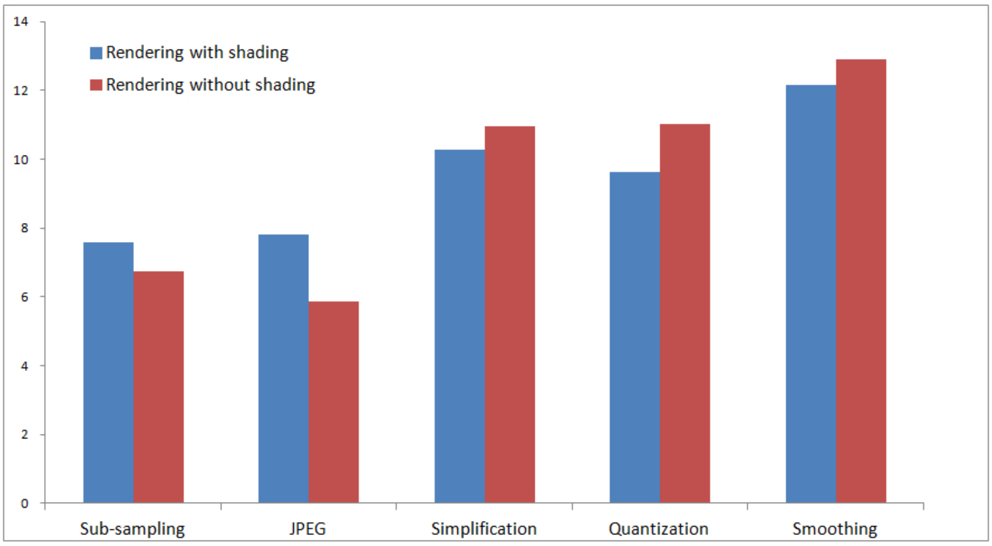


Figure 6.7: Mean scores (averaged among the 5 types of distortions and the 4 models) for the two types of rendering: with shading (top-left light source, diffuse material), and without shading (diffuse albedo only).

6.3 Toward an optimal metric for textured mesh quality assessment

In this section, we propose an objective metric for textured mesh quality assessment as a simple linear combination of mesh quality and texture quality. We use the subjective dataset presented below to evaluate the performance of different mesh and texture metrics for this task. We also compare their performance to video metrics computed on the rendered videos. In the next section, we also propose a new metric for geometry quality assessment.

6.3.1 A new metric for geometry quality assessment

By looking at the subjective scores obtained for geometric distortions (smoothing, quantization and simplification), we observed that the ranks of the distortions with the highest strengths from these three types show a pattern common to all models: the distorted model ranked as the worst visual quality always comes from either the strongest quantization method or from the strongest simplification. Distorted models from the strongest smoothing never appear at the end of the ranks. For instance, among the 20 distorted Hulk models, the distortion with the worst visual quality is the 7 bits quantization (subjective score: 0.91), while the distortion of the worst smoothing has a fairly high subjective score (7.45) (see Figure 6.8). Quality scores from L3,L4 are actually significantly higher than Q3,Q4 (p-values = 0.011 and 0.0063 for shaded/non-shaded data) and significantly higher than Si3,Si4 (p-values = 0.011 and 0.03). This subjective pattern is related to the perceptual mechanisms of the human visual system which is more sensitive to high frequency variations on local areas (e.g. distortions caused by simplification or quantization) rather than to more global low-frequency variations (e.g. caused by smoothing).

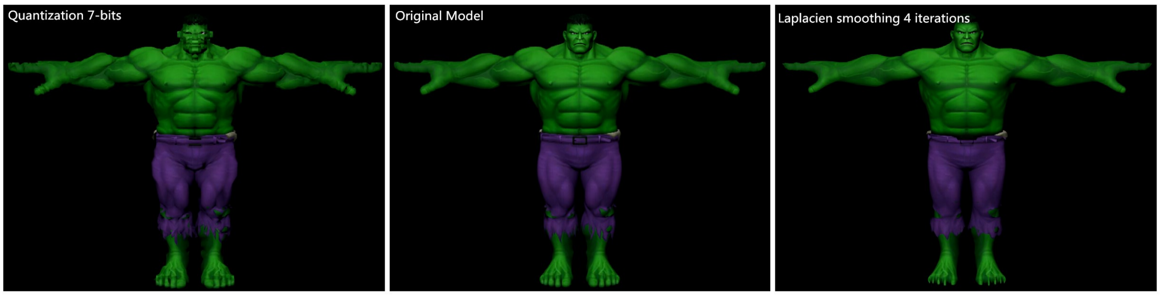


Figure 6.8: From left to right: Hulk model with 7-bits quantization (subjective score: 0.91), Original Hulk model and Hulk model with Laplacian smoothing of 4 iterations (subjective score: 7.45).

Based on these observations as well as previous studies which emphasize the good behavior of curvature for predicting visual distortions [56, 125, 115, 42], we propose a novel local distortion measurement by computing the amount of variance of curvature in local corresponding neighborhoods between two meshes (a distorted mesh M_d and a reference mesh M_r). We first establish a correspondence between M_d and M_r (as in [56]), we then compute the mean curvature C on each vertex of M_d and its corresponding curvature value $\hat{C}$ on corresponding point of M_r. For each vertex v from M_d, we then compute the standard deviation δ_v^h of the local curvature differences in a geodesic neighborhood of size h around v. δ_v^h is computed as follows:

$$\delta_v^h = \sqrt{\frac{1}{k}\sum_{j=1}^{k}[\frac{\hat{C}_j - C_j}{max(\hat{C}_j, C_j)+a} - E(\frac{\hat{C}_j - C_j}{max(\hat{C}_j, C_j)+a})]^2} \tag{6.2}$$

where k is the number of vertices of the neighborhood, a a constant to avoid instability when denominators are close to zero, and $E(\frac{\hat{C}_j - C_j}{max(\hat{C}_j, C_j)+a})$ the mean value of curvature differences in the neighborhood. δ_v^h is theoretically upper bounded by 2 but can be clamped in [0,1]. As in [56], we make this computation for three different neighborhood sizes h_i to capture the perceptually meaningful scales, and improve the efficiency and robustness of the metric. We took three scales $h_i \in \{2\varepsilon, 3\varepsilon, 4\varepsilon\}$, where $\varepsilon = 2.5\%$ of the max length of the bounding box of the model. Then multi-scale local distortion measure $\delta_M(v)$ is computed as: $\delta_M(v) = \frac{1}{n}\sum_{i=1}^{n}\delta_v^{h_i}$. n is the number of scales (3 in our study). Finally, we consider a root mean square pooling of these local measurements to obtain our global SDCD (Standard Deviation of Curvature Difference) value:

$$SDCD(M_d, M_r) = (\frac{1}{|M_d|}\sum_{v \in M_d}\delta_M(v)^2)^{\frac{1}{2}}. \tag{6.3}$$

This metric captures local roughness variations, whereas more global changes (e.g. a global shrink of the model) are not considered as perceptually significant. The performance of this metric is evaluated in the following section.

6.3.2 Mesh and Image metric evaluation

Since most of distorted models from our dataset are associated to a single attack (either on geometry or texture), their subjective scores may be used as ground truth to evaluate existing metrics respectively for geometry quality assessment only and image quality assessment only (applied on texture image) and consequently determine the most appropriate metrics (resp. for geometry and texture image) for their further combination. For each of our reference models, we split the dataset into two groups according to the distortion type: geometry ad texture. We select several commonly used perceptual geometric and image metrics. For geometry, we select the Root Mean Square Error (RMSE) computed on geometry, MSDM2

[56], which is one of the best performing perceptually-motivated metrics, and our newly proposed metric SDCD. For texture, we select the RMSE on image pixels, SSIM [127] and MS-SSIM [129] (top performing metrics on natural images). Tables 6.4 and 6.5 detail the Spearman and Pearson correlations between the objective metrics and the subjective scores for geometry and texture quality assessment, respectively.

	Squirrel		**Hulk**		**Statue**		**Sport Car**		**Dwarf**		**Average**	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
Geometry RMSE	0.72	**0.90**	0.37	0.61	0.07	0.00	**0.65**	0.28	0.02	0.28	0.37	0.41
MSDM2 [56]	**0.78**	0.76	**0.88**	**0.85**	0.15	0.15	0.53	**0.52**	0.86	**0.90**	0.64	0.64
SDCD (Our Metric)	0.60	0.55	0.84	0.80	**0.64**	**0.64**	0.40	0.41	**0.89**	0.86	**0.67**	**0.65**

Table 6.4: Performance comparison (Pearson r_p and Spearman r_p correlations) of several geometric metrics on our subjective database (geometric distortions only). Rendering with shading.

	Squirrel		**Hulk**		**Statue**		**Sport Car**		**Dwarf**		**Average**	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
Image RMSE	0.73	0.86	0.21	0.28	0.39	0.50	**0.93**	**1.00**	0.65	0.73	0.58	0.67
SSIM [127]	0.27	0.17	0.60	0.77	0.70	0.69	0.90	0.93	0.73	0.88	0.64	0.69
MS-SSIM [129]	**0.83**	**0.98**	**0.67**	**0.81**	**0.70**	**0.76**	0.86	0.86	**0.86**	**0.95**	**0.78**	**0.87**

Table 6.5: Performance comparison (Pearson r_p and Spearman r_p correlations) of several image metrics on our subjective database (texture distortions only). Rendering with shading.

Previous studies dedicated to quality assessment of 3D meshes [24] and natural images [136] have shown that geometry and image RMSEs are not good predictors of the visual quality. It is interesting to see that these results are confirmed for our database for which they are outperformed by perceptual metrics. This result is interesting because, in our case, texture and geometry involve complex masking effects. Indeed, texture artifacts may be masked by the geometric mapping and vice versa.

Table 6.4 shows that, on certain models, MSDM2 performs better than SDCD. However, for the *Statue*, SDCD demonstrates a significant improvement. Regarding the performance, we will consider both MSDM2 and SDCD for our further combined quality metric for textured mesh. For texture metric (see Table 6.5), MS-SSIM provides the best overall performance and will thus be chosen as texture metric for our further optimal combination.

6.3.3 Toward an optimal combination

We propose to assess the visual quality of a textured mesh as a simple linear combination of its geometry quality and its texture quality, respectively assessed by a 3D mesh metric Q_G

(MSDM2 or SDCD) and an image metric Q_T (MS-SSIM). We will see in the results, that this simple scheme is able to provide very good results. Our combined metric is thus defined as follows:

$$CM = \alpha Q_G + (1-\alpha) Q_T \tag{6.4}$$

where α is an optimal weight determined through a greedy optimization in a 5-folds cross-validation way. For each model (resp. Squirrel, Statue, Sport Car, Hulk and Dwarf) we compute the optimal weights as the one which maximize the Spearman correlation over the 4 other models. We propose two versions of our combined metric: CM_1 (the optimal combination of MSDM2 and MS-SSIM) and CM_2 (the optimal combination of SDCD and MS-SSIM). We compare our metrics CM_1 and CM_2 to several state of the art metrics:

- FQM [113, 114], a metric especially designed for textured mesh quality assessment. It is defined as a weighted combination of two simple error measures: the mean squared error over mesh vertices and the mean squared error over texture pixels. Optimal weights are computed using cross-validation, as for our metric.
- Several video quality metrics applied to the rendered videos: The DCT-based Video Quality Metric from F. Xiao [30], PSNR applied on all frames and averaged, and MS-SSIM applied on all frames and averaged. These metrics have been computed using the MSU Video Quality Measurement Tool [1].

The performance of these metrics are evaluated using the Spearman and Pearson correlations between the objective metric's values and the subjective scores, as well as the root-mean-squared error (RMS). The Pearson correlation and the RMS are computed after a logistic regression which provides a non-linear mapping between the objective and subjective scores. Results are shown in Tables 6.6 and 6.7, respectively for the shaded and non-shaded rendering. Scatter plots of subjective scores versus metric's values are presented in Figure 6.10.

As illustrated in the tables, our metrics CM_1 and CM_2 outperform the other for most of the models. Given the fact that, for a given model, the weighting factor α is learnt using the other ones, this good performance demonstrates an excellent inter-model robustness. It is interesting to see that the performance of our metrics is generally better for the non-shaded rendering than for the shaded rendering. The reason is that shading involves complex masking interactions between texture and geometry that are not considered in our metrics since they evaluate geometry and texture separately. These interactions are extremely limited in the non-shaded rendering, which accounts for the improved results of our metrics.

---

[1] http://compression.ru/video/quality_measure/video_measurement_tool_en.html

	Squirrel			**Hulk**			**Statue**			**Sport Car**			**Dwarf**		
	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS
Video-DCT	0.12	0.09	4.76	0.18	0.36	4.99	0.26	0.30	5.04	0.62	0.67	3.97	0.24	0.25	4.75
Video-PSNR	0.22	0.26	4.68	0.33	0.36	4.79	0.21	0.26	5.12	0.67	0.70	3.68	0.31	0.32	4.63
Video-MSSSIM	0.24	0.39	4.64	0.17	0.41	5.00	0.25	0.42	5.05	0.67	**0.72**	3.64	0.38	0.40	4.50
FQM	0.80	**0.85**	2.83	0.41	0.56	4.55	0.26	0.18	5.07	0.67	0.47	3.86	0.38	0.41	4.53
CM_1 (Our Metric)	**0.82**	0.82	**2.24**	**0.81**	**0.81**	**2.92**	0.36	0.30	4.93	**0.70**	0.60	**3.58**	0.56	0.70	3.95
CM_2 (Our Metric)	0.78	0.72	2.80	0.73	0.72	3.48	**0.62**	**0.68**	**3.94**	0.68	0.51	3.72	**0.58**	**0.70**	**3.87**

Table 6.6: Performance comparison (Pearson r_p and Spearman r_p correlations, RMSE of the residuals) of several textured mesh quality metrics on our subjective database. Rendering with shading.

	Squirrel			**Hulk**			**Statue**			**Sport Car**			**Dwarf**		
	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS	r_p	r_s	RMS
Video-DCT	-0.21	-0.26	4.76	0.22	0.25	4.72	0.34	0.24	4.74	0.58	0.42	4.23	0.24	0.21	4.89
Video-PSNR	-0.13	-0.04	4.82	0.35	0.39	4.50	0.13	0.22	4.97	0.51	0.56	4.32	0.31	0.29	4.78
Video-MSSSIM	-0.05	0.14	4.86	0.28	0.61	4.64	0.34	0.37	4.72	0.67	0.69	3.77	0.65	0.63	3.80
FQM	0.69	0.71	3.47	0.46	0.62	4.17	0.48	0.31	4.44	0.73	0.82	3.34	0.56	0.69	4.15
CM_1 (Our Metric)	**0.80**	**0.88**	**2.38**	0.69	0.80	3.29	0.47	0.41	4.53	**0.77**	**0.90**	**2.86**	0.73	0.81	3.23
CM_2 (Our Metric)	0.79	0.81	2.54	**0.72**	**0.82**	**2.93**	**0.63**	**0.69**	**3.81**	0.74	0.88	3.12	**0.76**	**0.88**	**3.04**

Table 6.7: Performance comparison (Pearson r_p and Spearman r_p correlations, RMSE of the residuals) of several textured mesh quality metrics on our subjective database. Rendering without shading.

All combined metrics (CM_1, CM_2 and FQM) have lower performances for the Statue. The reason is visible on Figure 6.10. Indeed, in the corresponding plots, we observe that the visual impact of simplification distortions (green dots) is underestimated by these metrics (which provide good quality scores for simplified models). The reason is that the poor subjective quality of these distorted versions is due to damage on the texture seams, which considerably alters the visual appearance but is totally unpredictable by combined metrics that do not take into account the texture mapping.he same underestimation of the simplification impact is observed for the Dwarf model, which also exhibits complex texture seams. For this latter model, the geometric metrics also underestimate the impact of quantization which is particularly harmful for such a high resolution model.

One last observation is the superiority of video-based metrics for the Sport Car model (shaded rendering). The reason is as follows: this model exhibits several interior parts (e.g. seats, radio) whose texture maps are severely damaged by JPEG and Sub-sampling distortions. However, these interior parts are almost invisible due to the position of the camera in the videos. Hence, the subjective scores are rather good for these distorted models whereas the image metrics predict very low quality values. On the contrary, video-based metrics only take into account the visible parts and thus provide correct results (in particular Video-MSSSIM). To verify this effect, we adopted a slightly different viewpoint for this particular model, in the non-shaded rendering, which improves the visibility of these interior parts. As expected, combined metrics results are much better (see table 6.7).

It is interesting to see that the performance of our metrics is overall better for the reflectance-only rendering than for the shaded rendering. The reason is that the shading involves complex masking interactions between texture and geometry that are not considered in our metrics since they evaluate geometry and texture separately. These interactions are very limited in the reflectance-only rendering, which explains the improved results of our metrics.

Table 6.8 details the α values for each model (computed using our 5-folds cross validation) for both rendering settings. As illustrated by the table, α are consistently smaller for the rendering without shading. As explained in section 6.2.6, when only the reflectance of the surface is taken into account, the visual impact of the geometric distortions is highly reduced. Hence, the parameter α may be adjusted/calibrated according to the lighting, material and other rendering parameters since they will emphasize either texture or geometry distortions. Still, it is interesting to notice that if we learn the α values on the subjective scores coming from the shaded rendering and use them to predict subjective score from the reflectance only dataset, the obtained correlation remain correct (less than 10% decrease of performance).

	Squirrel		**Hulk**		**Statue**		**Sport Car**		**Dwarf**		**Average**	
	αCM_1	αCM_2	αCM_1	αCM_2	αCM_1	αCM_2	αCM_1	αCM_2	αCM_1	αCM_2	αCM_1	αCM_2
With shading	0.086	0.117	0.108	0.184	0.103	0.133	0.086	0.132	0.061	0.111	0.089	0.135
Without shading	0.061	0.109	0.026	0.108	0.061	0.118	0.061	0.109	0.061	0.109	0.054	0.111

Table 6.8: α values for each model and each rendering setting.

6.3.4 Validation on compound distortions

In this section, we validate our objective metrics using a new set of compound geometry-texture distortions. We selected the Dwarf model, and we manually selected 36 distorted versions among the $12 \times 8 = 96$ possible combinations of the 12 geometry and 8 texture distortions detailed in Section 6.2.1, resulting in a new validation set of 36 models. Details about these mixed distortions are available in the supplementary material ; some examples are shown in Figure 6.4, bottom right. As before, we created two sets of videos (of 10 seconds duration) with and without shading, respectively. We performed a paired-comparison experiment to obtain the subjective scores. For this mixed-distortion setting, the hypothesis $Q_{D_i} > Q_{D_j}, \forall j > i$ from our sorting algorithm no longer holds. Hence, we implemented a more classical self-balancing binary tree, as in [77]. The average number of comparisons for sorting the 36 distorted models was 140 (instead of 630 for the full comparison). 20 observers took part to this new experiment, 19 rated 1 set and 1 rated both sets. We thus obtained 10 judgments for the shaded rendering and 11 for the non-shaded one. The average time to finish the experiment for one video set of 36 videos was 21 minutes. As for the previous experiment, the agreement is rather high (resp. 0.75 and 0.71 for shaded and non-shaded settings). Raw scores are detailed on the supplementary material.

To validate ou metrics, we computed their optimal weights based on the models from our previous experiment (we excluded the Dwarf). Results are detailed in Table 6.9 and Figure 6.9. Even for this difficult scenario (learning on single-type distortions and testing on compound distortions), our metrics offer excellent performance: 0.85 and 0.87 Spearman correlations for CM_2, for both rendering settings, hence demonstrating once again an excellent robustness.

	With shading			**Without shading**		
	r_p	r_s	RMS	r_p	r_s	RMS
Video-DCT	0.32	0.50	7.81	0.35	0.32	8.23
Video-PSNR	0.33	0.58	7.32	0.40	0.33	8.05
Video-MSSSIM	0.67	0.66	6.79	0.68	0.67	6.42
FQM	0.64	0.66	6.90	0.76	0.77	5.73
CM_1 (Our Metric)	0.74	0.77	6.01	0.85	0.86	4.47
CM_2 (Our Metric)	**0.80**	**0.85**	**5.33**	**0.86**	**0.87**	**4.44**

Table 6.9: Performance comparison on the compound distortion dataset.

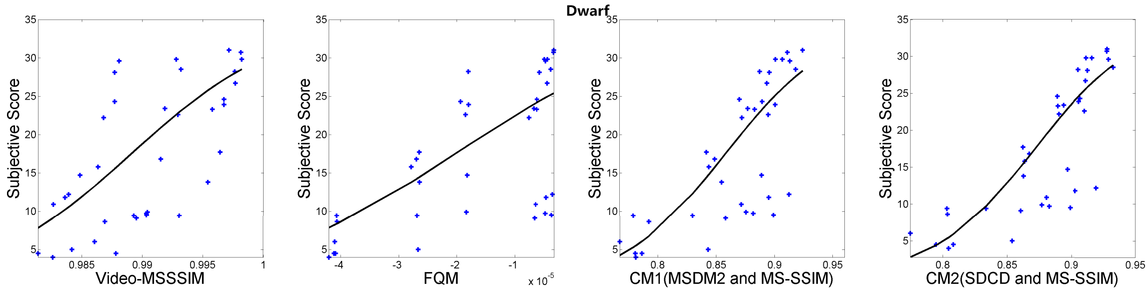


Figure 6.9: Scatter plots of subjective scores versus objective metric values for the compound distortion dataset. Each point represents one distorted model. Fitted logistic curves are represented in black.

6.4 Conclusion and Perspective

In this work, we have designed and constructed a new subjectively-rated database of textured 3D meshes. Our subjective study is based on a paired comparison protocol, and involved more than 100 subjects. The database contains 136 distorted models (subject to geometry and texture distortions), which were evaluated within 2 rendering settings. The subjective results allowed us to draw interesting conclusions regarding the influence of the shape and texture content as well as the rendering on the perceptual impact of distortions. We then proposed new objective metrics for visual quality assessment of textured meshes, as optimized linear combinations of mesh quality and texture quality. We used our subjective dataset to evaluate the performance of these metrics against state-of-the-art ones and to explain failure of metrics applied on the rendered images/videos. Such perceptually-validated metrics are of great interest for many applications such as 3D model simplification or compression, texture simplification and so on. We also proposed a new measure for geometry quality assessment. Note that our dataset, subjective scores and metric results will be made publicly available on-line.

While our proposed metrics showed that they outperformed their counterparts for the task of textured mesh quality assessment, there is still room for improvement. For instance, when evaluating geometry and texture quality, we need to integrate some visibility information. Indeed, the interior parts of a 3D model do not contribute to its rendered visual appearance; as another example, regions in very convex areas will only be visible in a few viewpoints and thus will have little impact on subjective opinion. Another issue is the variation in texel size within a texture; indeed, different regions of a texture map are not necessarily mapped with the same size in the screen, and this effect should be taken into account. The angular resolution (in pixels per degree of visual angle) of the rendered scene is also of importance and may be integrated as a scale factor in the metric. Finally, a major issue is to integrate texture coordinate distortions which are a common side effect of geometric changes. As we observed, even slight movements of texture seams may seriously harm visual appearance. Finally, we could conduct a comprehensive evaluation to determine, among the dozens of existing image metrics, the most adapted for texture evaluation.

Acknowledgements

We would like to thank Massimiliano Corsini and the Visual Computing Laboratory of ISTI-CNR for the Dwarf model, as well as Mark Pauly and the EPFL Computer Graphics and Geometry Laboratory for the Squirrel and Statue models. This work is supported in part by the China Scholarship Council.

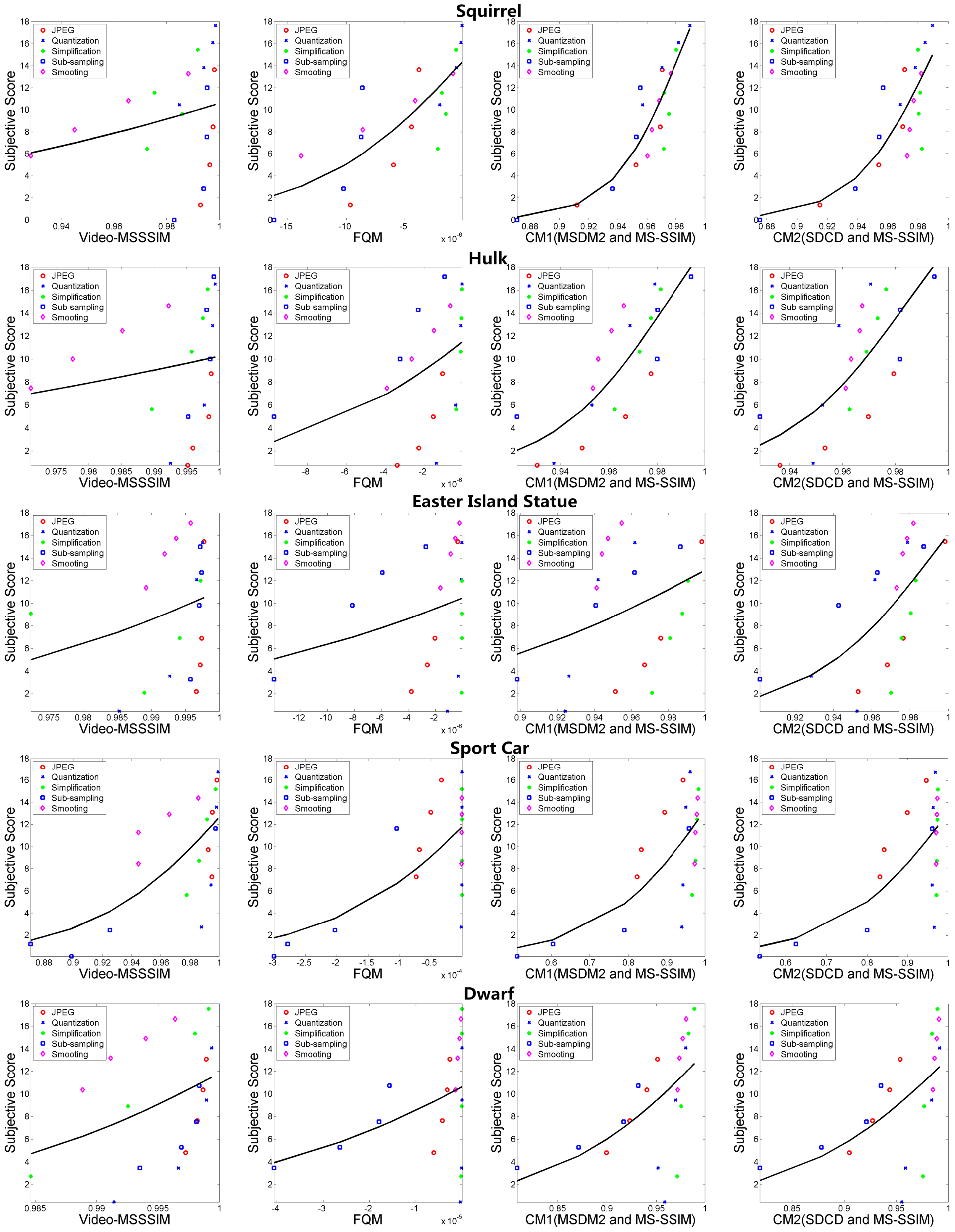


Figure 6.10: Scatter plots of subjective scores versus objective metric's values. Each point represents one distorted model. Fitted logistic curves are represented in black.

Chapter 7

Conclusion and Perspective

7.1 Summary of the contributions

In this manuscript, we have presented our research study on objective and subjective assessments of 3D models. Our main objectives were firstly to evaluate the performance of several visual quality metrics on predicting the local visibility of different distortions and providing perceptual quality index for textured 3D models with various artifacts and rendering scenarios, and then to propose new visual quality metrics for the texture mapped meshes. These objectives were accomplished through our designed subjective quality assessments on evaluating the local visibility of artifacts, and a variety of distorted textured meshes with different rendering protocols respectively. The two new quality metrics for textured meshes were determined through the optimal combinations of geometry and texture quality metrics. The experimental results have demonstrated the effectiveness of the proposed methods.
The contributions of this manuscript can be summarized as follows:

- **Designing a novel subjective experiment to collect information about local visibility of complex artifacts**
  We utilized the obtained distortion probability maps to illustrate several functionalities of the human visual system, and to quantitatively evaluate a large set of 3D mesh attributes as well as recent mesh visual quality metrics on predicting perceived local distortions. The experimental results showed that curvature-based attributes (particularly curvedness) demonstrate a much better performance than others (e.g. dihedral angles, normals, Laplacian, saliency), while Hausdorff geometric distance is a poor predictor of the perceived local distortion. Meanwhile, the recent perceptual quality metrics, originally designed for global quality estimation, also provides excellent results for the task of predicting locally perceived distortion

- **A public dataset of distortion probability maps of complex artifacts**
  This dataset contributes to a further understanding of the HVS mechanisms on perceiving geometric artifacts, and assists researchers to propose more efficient and effective visual quality metrics in future. In particular this dataset will be very useful to evaluate and/or to train future metrics.

- **A new subjective rating database of textured 3D meshes**
  Based on a paired comparison protocol, this subjective study enables us to draw remarkable conclusions regarding the perceptual impacts of the interactions between distortions and rendering protocols.

- **New objective quality metrics of textured 3D meshes based on optimal combination of mesh quality and texture quality**
  We used our subjective dataset to evaluate the performance of this metric against state-of-the-art ones. Such metric can play an efficient role in many applications, such as 3D model simplification or compression, texture simplification and so on. We also proposed a new metric for geometry quality assessment.

In a nutshell, an important characteristic of this thesis work is that we have made several creative efforts on exploiting the mechanisms of visual perception in complex quality evaluation tasks, and proposed some notable mathematical assessments to evaluate the visual quality of 3D meshes. Indeed, we believe that the future research on visual quality assessment of 3D objects can be improved by emending machine-learning mechanism, and integrating human vision processes (such like a model of contrast sensitivity function).

7.2 Perspective

Several research perspectives appear at the end of this thesis.

- **Improvement of the mesh visual quality metrics based on Machine-learning mechanism**
  Machine learning provides a good opportunity to construct the perceptual quality metrics based on our new ground-truths (i.e. distortion probability maps). Such metrics have potential applications of 3D mesh artifact localization.

- **Improvement of the visual quality metrics for textured 3D objects by integrating visibility information**
  Considering the effects of less visible regions (e.g., interior or convex areas of a model,

etc.), texture mapping information and the complex masking interactions between geometry and texture, this visibility information could improve the efficiency of the future quality metrics for texture mapped meshes.

More challenges are still waiting in the fields of visual quality assessment for 3D graphical data. Particularly, many progresses still remain to be achieved in understanding the complex interactions between geometry, material and lighting information, and the hidden connections between these interactions and functionalities of HVS. All these factors should be considered to devise efficient quality metrics in future. In all, we believe that the study on this research area has a very promising future.

7.3 Relevant Publications

International Journal

- Jinjiang Guo, Vincent Vidal, Irene Cheng, Anup Basu, Atilla Baskurt, and Guillaume Lavoue, Subjective and Objective Visual Quality Assessment of Textured 3D Meshes, ACM Transactions on Applied Perception, Association for Computing Machinery, 2016, 14 (2), pp.11.

International Conference

- Jinjiang Guo, Vincent Vidal, Atilla Baskurt, and Guillaume Lavoue. Evaluating the local visibility of geometric artifacts. In Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, pages 91-98. ACM, 2015.

Bibliography

[1] Alliez, P., Cohen-Steiner, D., Devillers, O., Lévy, B., and Desbrun, M. (2003). Anisotropic polygonal remeshing. *ACM Transactions on Graphics*, 22(3):485.

[2] Alliez, P., Tayeb, S., and Wormser, C. (2009). AABB Tree. CGAL 3.5 edition.

[3] Aydin, T. O., Čadík, M., Myszkowski, K., and Seidel, H.-P. (2010). Video quality assessment for computer graphics applications. *ACM Transactions on Graphics*, 29(6):1.

[4] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. a., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics (Oxford, England)*, 16(5):412–424.

[5] Belyaev, A. and Ohtake, Y. (2003). A Comparison of Mesh Smoothing Methods. In *Israel-Korea Bi-national conference on geometric ...*, volume 2, pages 1–5. Citeseer.

[6] Benhabiles, H., Lavoúe, G., Vandeborre, J. P., and Daoudi, M. (2011). Learning boundary edges for 3D-mesh segmentation. *Computer Graphics Forum*, 30(8):2170–2182.

[7] Bolin, M. R. and Meyer, G. W. (1998). A perceptually based adaptive sampling algorithm. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 299–309. ACM.

[8] Bulbul, A., Capin, T., Lavoue, G., and Preda, M. (2011). Assessing visual quality of 3-D polygonal models. *IEEE Signal Processing Magazine*, 28(6):80–90.

[9] Čadík, M., Herzog, R., and Mantiuk, R. (2012). New measurements reveal weaknesses of image quality metrics in evaluating graphics artifacts. *ACM Siggraph*.

[10] Čadík, M., Herzog, R., Mantiuk, R., Mantiuk, R., Myszkowski, K., and Seidel, H. P. (2013). Learning to predict localized distortions in rendered images. In *Computer Graphics Forum*, volume 32, pages 401–410.

[11] Chan, R. H., Ho, C.-W., and Nikolova, M. (2005). Salt and Pepper Noise Removal by Median Type Noise Detectors and Detail-Preserving Regularization. *IEEE Transactions on Image processing*, 14(10):1479–1485.

[12] Chandler, D. M. and Hemami, S. S. (2007). VSNR : A Visual Signal-to-Noise Ratio for Natural Images. *IEEE Transactions on Image Processing*, 16(9):2284–2298.

[13] Charrier, C., Lézoray, O., and Lebrun, G. (2012). Machine learning to design full-reference image quality assessment algorithm. *Signal Processing: Image ...*, 27(3):209–219.

[14] Chen, K., Wu, C., Chang, Y., and Lei, C. (2009). A Crowdsourceable QoE evalutation framework for multimedia content. In *ACM International Conference on Multimedia*, pages 491–500. ACM.

[15] Chen, W., Fournier, J., Barkowsky, M., and Le Callet, P. (2010). New requirements of subjective video quality assessment methodologies for 3dtv. In *Video Processing and Quality Metrics 2010 (VPQM)*.

[16] Chen, X., Saparov, A., Pang, B., and Funkhouser, T. (2012). Schelling points on 3D surface meshes. *ACM Transactions on Graphics*, 31(4):1–12.

[17] Cho, J., Prost, R., and Jung, H. (2007). An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms. *IEEE Transactions on Signal Processing*, 55(1):142–155.

[18] Chou, P. H. and Meng, T. H. (2002). Vertex data compression through vector quantization. *IEEE Transactions on Visualization and Computer Graphics*, 8(4):373–382.

[19] Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. (2008). Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, volume 2008, pages 129–136.

[20] Cignoni, P., Rocchini, C., and Scopigno, R. (1998). Metro: Measuring Error on Simplfied Surfaces. In *Compututer Graphics Forum*, volume 17, pages 167–174. Wiley Online Library.

[21] Cohen-Steiner, D., Alliez, P., and Desbrun, M. (2004). Variational shape approximation. *ACM Transactions on Graphics*, 23(3):905.

[22] Cohen-Steiner, D. and Morvan, J.-M. (2003). Restricted delaunay triangulations and normal cycle. In *Proceedings of the nineteenth conference on Computational geometry - SCG '03*, page 312.

[23] Corsini, M., Gelasca, E. D., Ebrahimi, T., and Barni, M. (2007). Watermarked 3-D mesh quality assessment. *IEEE Transactions on Multimedia*, 9(2):247–255.

[24] Corsini, M., Larabi, M. C., Lavoué, G., Petřík, O., Váša, L., and Wang, K. (2013). Perceptual metrics for static and dynamic triangle meshes. In *Computer Graphics Forum*, volume 32, pages 101–125. Wiley Online Library.

[25] Daly, S. (1993). Visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, volume 1666, pages 179–206. International Society for Optics and Photonics.

[26] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., and Bovik, A. C. (2000). Image quality assessment based on a degradation model. *Image Processing, IEEE Transactions on*, 9(4):636–650.

[27] E, C. O. M. L. S., Chisholm, S., and Q, R. (2008). Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. *Group*, 28(February 2005):15–25.

[28] Eckert, M. P. and Bradley, A. P. (1998). Perceptual quality metrics applied to still image compression. *Signal Processing*, 70(3):177–200.

[29] Eggers, J. J., Bäuml, R., Tzschoppe, R., and Girod, B. (2003). Scalar Costa scheme for information embedding. *IEEE Transactions on Signal Processing*, 51(4):1003–1019.

[30] F. Xiao (2000). DCT-based video quality evaluation. Technical report, Stanford University.

[31] Faisman, A. and Langer, M. S. (2013). How does lighting direction affect shape perception of glossy and matte surfaces? In *Proceedings of the ACM Symposium on Applied Perception - SAP '13*, page 9. ACM.

[32] Farooque, M. A. and Rohankar, J. S. (2013). Survey on Various Noises and Techniques for Denoising the Color Image. *International Journal of Application or Innovation in Engineering and Management*, 2(11):217–221.

[33] Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. In *ACM Transactions on Graphics*, volume 21, pages 249–256. ACM.

[34] Ferwerda, J. A., Pattanaik, S., Shirley, P., and Greenberg, D. P. (1997). A Model of Visual Masking for Computer Graphics. In *Siggraph*, pages 143–152. ACM Press/Addison-Wesley Publishing Co.

[35] Fleming, R. (2014). Visual perception of materials and their properties. *Vision research*, 94(24):62–75.

[36] Gall, D. L. (1991). MPEG: A Video Compression Standard for Multimedia Applications. *Association for Computing Machinery.Communications of the ACM*, 34(4):46.

[37] Garland, M. and Heckbert, P. S. (1997). Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216. ACM Press/Addison-Wesley Publishing Co.

[38] Gelasca, E. D., Ebrahimi, T., Corsini, M., and Barni, M. (2005). Objective evaluation of the perceptual quality of 3D watermarking. In *Proceedings - International Conference on Image Processing, ICIP*, volume 1, pages 241–244. IEEE.

[39] Girod, B. (1993). What's wrong with mean-squared error? In *Digital images and human vision*, pages 207–220. MIT press.

[40] Goodman, J. W. (1976). Some fundamental properties of speckle. *J. Opt. Soc. Am.*, 66(11):1145–1150.

[41] Guennebaud, G. and Gross, M. (2007). Algebraic point set surfaces. In *ACM Transactions on Graphics*, volume 26, page 23. ACM.

[42] Guo, J., Vidal, V., Baskurt, A., and Lavou, G. (2015). Evaluating the local visibility of geometric artifacts. In *Symposium in Applied Perception*.

[43] Guskov, I. and Wood, Z. (2001). Topological noise removal. *Proc. Graphics Interface*, pages 19–26.

[44] Hadhoud, M. M., Dessouky, M. I., and El-Samie, F. E. A. (2003). Adaptive image interpolation based on local activity levels. In *Proceedings of the Twentieth National Radio Science Conference (NRSC'03)*, pages C4– 1–8. IEEE.

[45] Herzog, R., Čadík, M., Aydčin, T. O., Kim, K. I., Myszkowski, K., and Seidel, H.-P. (2012). NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis. *33rd Annual Conference of the European Association for Computer Graphics*, 31(2pt3):545–554.

[46] Hoppe, H., DeRose, T., and Duchamp, T. (1993). Mesh optimization. In *Proceedings of the 20th ...*, volume d, pages 19–25. ACM.

[47] ITU (2002). Methodology for the subjective assessment of the quality of television pictures. *Itu-R Bt.500-11*, 211(BT.500-11):1–48.

[48] Karni, Z. and Gotsman, C. (2000). Spectral compression of mesh geometry. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00*, pages 279–286. ACM Press/Addison-Wesley Publishing Co.

[49] Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Josa*, 69(10):1340–1349.

[50] Kendall, M. G. (1975). Rank Correlation Methods. *London Griffin*, 34(1):108.

[51] Kendall, M. G. and Smith, B. B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*, 10(3):275–287.

[52] Kendall, M. G. and Smith, B. B. (1947). On the method of paired comparisons. *Biometrika*, 34(Pt 3-4):324–345.

[53] Koenderink, J. J. and van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–564.

[54] Kripendorff, K. (1980). *Content Analysis. AN Introduction to Its Methodology*. Thousand Oaks.

[55] Lavoué, G. (2009). A local roughness measure for 3D meshes and its application to visual masking. *ACM Transactions on Applied Perception*, 5(4):1–23.

[56] Lavoué, G. (2014). A multiscale metric for 3D mesh visual quality assessment. In *Computer Graphics Forum*, volume 30, pages 1427–1437. Wiley Online Library.

[57] Lavoué, G., Cheng, I., and Basu, A. (2013). Perceptual quality metrics for 3D meshes: Towards an optimal multi-attribute computational model. In *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, pages 3271–3276. IEEE.

[58] Lavoue, G. and Corsini, M. (2010). A comparison of perceptually-based metrics for objective evaluation of geometry processing. *IEEE Transactions on Multimedia*, 12(7):636–649.

[59] Lavoue, G., Drelie Gelasca, E., Dupont, F., Baskurt, A., and Ebrahimi, T. (2006). Perceptually driven 3D distance metrics with application to watermarking. In *Spie*, volume 6312, pages 63120L–63120L–12. International Society for Optics and Photonics.

[60] Lavoue, G., Larabi, M. C., and Vasa, L. (2016). On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models. *IEEE Transactions on Visualization and Computer Graphics*, 22(8):1987–1999.

[61] Lavoué, G. and Mantiuk, R. (2015). Quality assessment in computer graphics. In *Visual Signal Quality Assessment: Quality of Experience (QoE)*, pages 243–286. Springer.

[62] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H. (2005). Evaluation of tone mapping operators using a High Dynamic Range display. *ACM Transactions on Graphics*, 24(3):640.

[63] Lee, C. H., Varshney, A., and Jacobs, D. W. (2005). Mesh saliency. In *ACM Transactions on Graphics*, volume 24, page 659. ACM.

[64] Lee, E. S. and Ko, H. S. (2000). Vertex data compression for triangular meshes. In *Proceedings - Pacific Conference on Computer Graphics and Applications*, volume 2000-January, pages 225–234. IEEE.

[65] Lee, J. S. (1981). Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17(1):24–32.

[66] Lee, J. S. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics and Image Processing*, 24(2):255–269.

[67] Lee, J.-S., De Simone, F., and Ebrahimi, T. (2011). Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding. *IEEE Transactions on Multimedia*, 13(5):882–893.

[68] Levoy, M., Rusinkiewicz, S., Ginzton, M., Ginsberg, J., Pulli, K., Koller, D., Anderson, S., Shade, J., Curless, B., Pereira, L., Davis, J., and Fulk, D. (2000). The Digital Michelangelo Project: 3D Scanning of Large Statues. In *Conference on Computer Graphics and Interactive Techniques*, pages 131–144. ACM Press/Addison-Wesley Publishing Co.

[69] Li, C. and Bovik, A. C. (2010). Content-weighted video quality assessment using a three-component image model. *Journal of Electronic Imaging*, 19(1):011003.

[70] Lin, W. and Kuo, C. C. J. (2011). Perceptual visual quality metrics A survey-ref. *Journal of Visual Communication and Image Representation*, 22(4):297.

[71] Lindstrom, P. (2000). *Model Simplification using Image and Geometry-Based Metrics*. PhD thesis, Citeseer.

[72] Lindstrom, P. and Turk, G. (2000). Image-driven simplification. *ACM Transactions on Graphics (ToG)*, 19(3):204–241.

[73] Lubin, J. (1995). A visual discrimination model for image system design and evaluation. *E. Peli, Ed. Singapore: World Scientific in Visual Models for Target Detectionand Recognition*, 2:207–220.

[74] Maglo, A. (2013). 3D mesh compression : survey , comparisons and emerging trends. *ACM Computing Surveys (CSUR)*, 9(4):44.

[75] Majid Rabbani, R. J. (2002). An overview of the JPEG2000 still image compression standard. *Signal Processing: Image Communication 17 (2002)*, 17(1):3–48.

[76] Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2004). Visible difference predicator for high dynamic range images. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 3, pages 2763–2769. IEEE.

[77] Mantiuk, R., Tomaszewska, A., and Mantiuk, R. (2012). Comparison of four subjective methods for image quality assessment. In *Computer graphics forum*, volume 31, pages 2478–2491. Wiley Online Library.

[78] Narwaria, M. and Lin, W. (2012). SVD-based quality metric for image and video using machine learning. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 42(2):347–64.

[79] O'Shea, J., Banks, M., and Agrawala, M. (2008). The assumed light direction for perceiving shape from shading. *symposium on Applied perception in graphics and visualization*.

[80] P.910, I.-T. R. (2008). Subjective video quality assessment methods for multimedia applications.

[81] Pan, Y., Cheng, I., and Basu, A. (2005). Quality metric for approximating subjective evaluation of 3-D objects. *Multimedia, IEEE Transactions on*, 7(2):269–279.

[82] Parker, J. a., Kenyon, R. V., and Troxel, D. E. (1983). Comparison of interpolation methods for image resampling. *IEEE Transactions on Medical Imaging*, 2(1):31–39.

[83] Pearson, K. (2006). Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London (1854-1905)*, 58:240–242.

[84] Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America. A, Optics and image science*, 7(10):2032–2040.

[85] Pinson, M. and Wolf, S. (2004). "A New Standardized Method for Objectively Measuring Video Quality," IEEE Transactions on Broadcasting, vol. 50, no. 3, pp. 312-322. *Broadcasting, IEEE Transactions on*, 50(3):312–322.

[86] Qu, L. and Meyer, G. W. (2008). Perceptually guided polygon reduction. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1015–1029.

[87] Ramponi, G. (1999). Warped distance for space-variant linear image interpolation. *IEEE Transactions on Image Processing*, 8(5):629–639.

[88] Rickett, A. A. (1977). *Wang Kuo-wei's Jen-chien Tz'u-hua: A Study in Chinese Literary Criticism*, volume 1. Hong Kong University Press.

[89] Rogowitz, B. E. and Rushmeier, H. E. (2001). Are image quality metrics adequate to evaluate the quality of geometric objects? *Photonics West 2001-Electronic Imaging*, pages 340–348.

[90] Rohaly, A. M., Corriveau, P. J., Libert, J. M., Webster, A. A., Baroncini, V., Beerends, J., ... & Winkler, S. (2000). Video quality experts group: Current results and future directions. In *SPIEVisual Communication and Image Processing*, pages 742–753. International Society for Optics and Photonics.

[91] Ronfard, R. and Rossignac, J. (1996). Full-range approximation of triangulated polyhedra. In *Computer Graphics Forum*, volume 15, pages 67–76. Wiley Online Library.

[92] Rossignac, J. and Borrel, P. (1993). *Multi-resolution 3D approximations for rendering complex scenes*. Springer.

[93] Rubinstein, M., Gutierrez, D., Sorkine, O., and Shamir, A. (2010). A comparative study of image retargeting. *SIGGRAPH ASIA '10: SIGGRAPH Asia 2010 papers*, 29(6):1.

[94] Rushmeier, H., Rogowitz, B., and Piatko, C. (2000). Perceptual issues in substituting texture for geometry. In *Spie*, pages 372–383. International Society for Optics and Photonics.

[95] Salkind, N. J. (2006). *Encyclopedia of measurement and statistics*. Sage Publications.

[96] Schroeder, W. J., Zarge, J. A., and Lorensen, W. E. (1992). Decimation of triangle meshes. In *ACM Siggraph Computer Graphics*, volume 26, pages 65–70. ACM.

[97] Sedgewick, R. and Flajolet, P. (2013). *An Introduction to the Analysis of Algorithms*. Addison-Wesley.

[98] Seshadrinathan, K. and Bovik, A. C. (2009). Motion-based perceptual quality assessment of video. In *Proceedings of SPIE*, pages 72400X–72400X–12. International Society for Optics and Photonics.

[99] Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441.

[100] Sheikh, H. R. and Bovik, A. C. (2006). Image information and visual quality. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 15(2):430–444.

[101] Sheikh, H. R., Bovik, A. C., and De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128.

[102] Silverstein, D. A. and Farrell, J. E. (2001). Efficient method for paired comparison. *Journal of Electronic Imaging*, 10(2):394–398.

[103] Skodras, A., Christopoulos, C., and Ebrahimi, T. (2001). The JPEG 2000 Still Image. *IEEE Signal Processing Magazine*, 18(September):36–58.

[104] Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C., and Seidel, H.-P. (2004). Laplacian Surface Editing. In *Eurographics Symposium on Geometry Processing*, pages 175–184. ACM.

[105] Sorkine, O., Cohen-Or, D., and Toldeo, S. (2003). High-pass quantization for mesh encoding. In *Eurographics Symposium on Geometry Processing*, pages 42–51. Citeseer.

[106] Spearman, C. (1904). The proof and measurement of association between two rings. *Ameri J Psychol*, 5(15):72–101.

[107] Ström, J. and Akenine-Möller, T. (2005). i PACKMAN: high-quality, low-complexity texture compression for mobile phones. *ACM SIGGRAPH/EUROGRAPHICS Conference on Graphics Hardware*, pages 177–182.

[108] Sun, J. and Perona, P. (1998). Where is the sun? *Nature neuroscience*, pages 183–184.

[109] Taubin, G. (1995a). Curve and surface smoothing without shrinkage. In *Proceedings of IEEE International Conference on Computer Vision*, pages 852–857. IEEE.

[110] Taubin, G. (1995b). A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 351–358. ACM.

[111] Thévenaz, P., Blu, T., and Unser, M. (2009). Image interpolation and resampling. *Handbook of Medical Image Processing and Analysis*, pages 465–493.

[112] Thurstone, L. L. (1927). A law of comparative judgments. *Psychological Review*, 34:273–286.

[113] Tian, D. and AlRegib, G. (2004). Fqm: a fast quality measure for efficient transmission of textured 3d models. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 684–691. ACM.

[114] Tian, D. and Alregib, G. (2008). BaTex3: Bit allocation for progressive transmission of textured 3-D models. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1):23–35.

[115] Torkhani, F., Wang, K., and Chassery, J. M. (2012). A curvature tensor distance for mesh visual quality assessment. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7594 LNCS, pages 253–263. Springer.

[116] Touma, C. and Gotsman, C. (1998). Triangle mesh compression.

[117] Vallet, B. and Lévy, B. (2008). Spectral geometry processing with manifold harmonics. In *Computer Graphics Forum*, volume 27, pages 251–260. Wiley Online Library.

[118] van Dijk, A., Martens, J., and Watson, A. (1802). Quality assessment of coded images using numerical category scaling. In *Advanced Networks and Services*, volume 2451, pages 90–101. International Society for Optics and Photonics.

[119] Váša, L. and Rus, J. (2012). Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. In *Computer Graphics Forum*, volume 31, pages 1715–1724. Wiley Online Library.

[120] Vidal, V. (2011). *Développement de modèles graphiques probabilistes pour analyser et remailler les maillages triangulaires 2-variétés*. PhD thesis, INSA de Lyon.

[121] Voran, S. and Catellier, A. (2011). Gradient ascent subjective multimedia quality testing. In *Eurasip Journal on Image and Video Processing*, volume 2011, pages 133–138. IEEE.

[122] Wallace, G. K. (1992). The JPEG Still Picture Compession Standard. *Architecture*, 38(1):xviii—-xxxiv.

[123] Wang, K., Lavou, G., Denis, F., and Baskurt, A. (2011). Robust and blind mesh watermarking based on volume moments. *Computers and Graphics (Pergamon)*, 35(1):1–19.

[124] Wang, K., Luo, M., Bors, A. G., and Denis, F. (2009). Blind and robust mesh watermarking using manifold harmonics. In *Proceedings - International Conference on Image Processing, ICIP*, number 3, pages 3657–3660.

[125] Wang, K., Torkhani, F., and Montanvert, A. (2012). A fast roughness-based approach to the assessment of 3D mesh visual quality. *Computers and Graphics (Pergamon)*, 36(7):808–818.

[126] Wang, Z. and Bovik, A. C. (2006). *Modern Image Quality Assessment*, volume 2. Morgan {&} Claypool Publishers.

[127] Wang, Z., Bovik, a. C., Sheikh, H. R., and Simmoncelli, E. P. (2004). Image quality assessment: form error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.

[128] Wang, Z. and Li, Q. (2011). Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198.

[129] Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. Ieee.

[130] Watson, A. B. (1993). DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images. In *Sid International Symposium Digest of Technical Papers*, volume 24, pages 946–949. SOCIETY FOR INFORMATION DISPLAY.

[131] Watson, A. B., Yang, G. Y., Solomon, J. A., and Villasenor, J. (1997). Visibility of wavelet quantization noise. *Image Processing, IEEE Transactions on*, 6(8):1164–1175.

[132] Watson, B. (2001). Measuring and predicting visual fidelity. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 213–220.

[133] Wills, J., Agarwal, S., Kriegman, D., and Belongie, S. (2009). Toward a perceptual space for gloss. *ACM Transactions on Graphics (TOG)*, 28(4):103.

[134] Wu, J. H., Hu, S. M., Tai, C. L., and Sun, J. G. (2001). An effective feature-preserving mesh simplification scheme based on face constriction. In *Proceedings - Pacific Conference on Computer Graphics and Applications*, volume 2001-January, pages 12–21.

[135] Yeganeh, H. and Wang, Z. (2013). Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667.

[136] Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2012). A comprehensive evaluation of full reference image quality assessment algorithms. *Proceedings of the 19th IEEE International Conference on Image Processing*, pages 1477–1480.

[137] Zhou Wang, L. L. and Bovik, A. C. (2004). Video Quality Assessment Based on Structural Distortion Measurement. *Signal Processing: Image Communication*, 19(2):121–132.

[138] Zhu, Q., Zhao, J., Du, Z., and Zhang, Y. (2010). Quantitative analysis of discrete 3D geometrical detail levels based on perceptual metric. *Computers and Graphics (Pergamon)*, 34(1):55–65.

Appendix A

Supplementary materials for Chapter 5

In this supplement, we report additional images, tables and experimental results.

A.1 Additional details on the ground-truth data

The 11 subjective distortion maps:

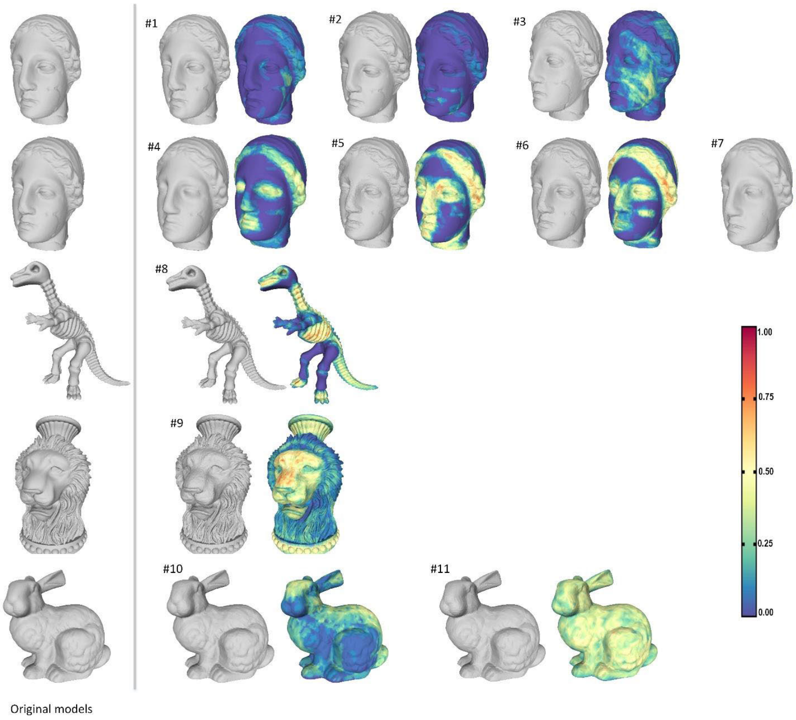


Figure A.1: Original models, distorted models and observer maps (color-coded probabilities of artifact detection) for each stimulus from our dataset.

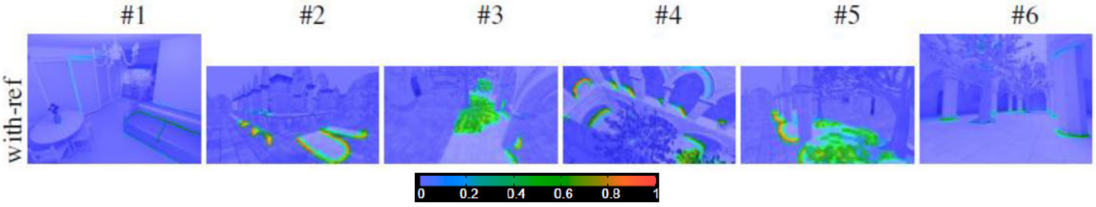


Figure A.2: Observer maps of the 6 first stimuli from the dataset of Cadik et al., SIGGRAPH ASIA 2012.

The Figure A.2 illustrates the subjective distortion maps from the 2D image experiment from Cadik et al., SIGGRAPH ASIA 2012. We can observe that the amplitudes and distributions of the detection probabilities are similar to our experiments. For instance, the most distorted areas are usually marked by 50 to 80% of the observers. This illustrates the fact that the agreements of observers are similar in both experiments (also shown by the Kendall u coefficients).

The agreement of the observers is also illustrated in the Figure A.3 which presents the painting results of the 20 observers for stimulus #8.

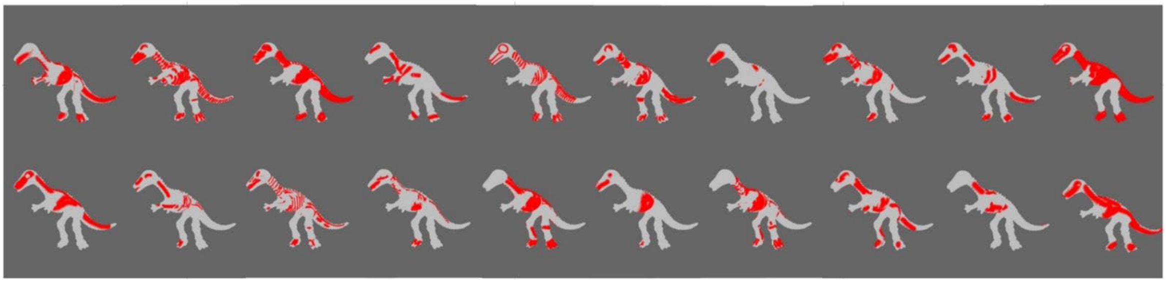


Figure A.3: Example of raw binary maps from 20 observers for stimulus #8 (Kripp. α = 0.162).

A.2 Additional results

Complete AUC (area under the ROC curve) results for all the tested features and for all filters. Mean values are also presented, averaged over the 11 objects and averaged over 9 objects (excluding #2 and #11). Features are ranked according to their “mean 9 objects” performance:

Table A.1: AUC values for all the tested features and for all filters.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	Mean 11 objects	Mean 9 objects
MSDM2	0,97	0,97	0,85	0,84	0,90	0,88	0,81	0,77	0,79	0,71	0,54	0,82	0,84
Curvedness(2,S)	0,95	0,97	0,86	0,89	0,91	0,90	0,85	0,71	0,67	0,70	0,54	0,82	0,83
MaxCurv(2,S)	0,95	0,97	0,85	0,90	0,91	0,89	0,86	0,71	0,66	0,70	0,54	0,81	0,83
Curvedness(2,N)	0,94	0,96	0,87	0,89	0,85	0,82	0,82	0,65	0,78	0,73	0,58	0,81	0,82
MaxCurv(2,N)	0,94	0,96	0,86	0,89	0,85	0,82	0,82	0,65	0,77	0,73	0,57	0,81	0,81
Curvedness(1,N)	0,94	0,83	0,84	0,84	0,88	0,84	0,80	0,76	0,73	0,68	0,55	0,79	0,81
MeanCurv(2,S)	0,93	0,96	0,83	0,89	0,90	0,87	0,86	0,67	0,66	0,68	0,51	0,80	0,81
MaxCurv(1,N)	0,93	0,65	0,83	0,84	0,87	0,84	0,80	0,75	0,74	0,68	0,55	0,77	0,81
Curvedness(1,S)	0,90	0,68	0,80	0,85	0,90	0,88	0,82	0,80	0,61	0,66	0,55	0,77	0,80
MaxCurv(1,S)	0,88	0,48	0,79	0,85	0,90	0,88	0,82	0,79	0,62	0,66	0,57	0,75	0,80
FMPD(2,S)	0,96	0,95	0,87	0,83	0,82	0,79	0,78	0,70	0,82	0,61	0,56	0,79	0,80
FMPD(2,N)	0,96	0,95	0,87	0,83	0,82	0,79	0,78	0,70	0,82	0,61	0,56	0,79	0,80
MeanCurv(1,N)	0,93	0,76	0,80	0,83	0,87	0,83	0,79	0,74	0,70	0,68	0,56	0,77	0,80
MeanCurv(2,N)	0,92	0,96	0,85	0,87	0,83	0,79	0,82	0,61	0,77	0,70	0,54	0,79	0,80
MeanCurv(1,S)	0,87	0,54	0,77	0,84	0,89	0,87	0,82	0,78	0,59	0,66	0,52	0,74	0,79
Laplacegauss(2,S)	0,88	0,66	0,71	0,86	0,92	0,91	0,81	0,77	0,72	0,50	0,56	0,75	0,79
Laplacegauss(2,N)	0,88	0,66	0,71	0,86	0,92	0,91	0,81	0,77	0,72	0,50	0,56	0,75	0,79
FMPD(1,S)	0,90	0,78	0,81	0,79	0,83	0,79	0,76	0,72	0,77	0,62	0,60	0,76	0,78
FMPD(1,N)	0,90	0,78	0,81	0,79	0,83	0,79	0,76	0,72	0,77	0,62	0,60	0,76	0,78
ShapeIndex(1,S)	0,86	0,67	0,69	0,81	0,86	0,82	0,78	0,72	0,71	0,63	0,53	0,73	0,76
Normal(1,S)	0,90	0,72	0,59	0,85	0,88	0,85	0,84	0,77	0,61	0,58	0,57	0,74	0,76
Normal(1,N)	0,90	0,72	0,59	0,85	0,88	0,85	0,84	0,77	0,61	0,58	0,57	0,74	0,76
Laplacegauss(1,S)	0,76	0,89	0,63	0,85	0,90	0,89	0,83	0,78	0,64	0,54	0,58	0,75	0,76
Laplacegauss(1,N)	0,76	0,89	0,63	0,85	0,90	0,89	0,83	0,78	0,64	0,54	0,58	0,75	0,76

Table A.1: (continued)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	Mean 11 objects	Mean 9 objects
ShapeIndex(1,N)	0,87	0,69	0,69	0,80	0,85	0,82	0,77	0,72	0,69	0,60	0,52	0,73	0,76
Curvedness(0,N)	0,83	0,78	0,72	0,80	0,85	0,83	0,78	0,75	0,59	0,65	0,56	0,74	0,75
GausCurv(1,N)	0,86	0,51	0,68	0,81	0,87	0,84	0,79	0,76	0,55	0,61	0,54	0,71	0,75
MeanCurv(0,N)	0,83	0,77	0,71	0,79	0,85	0,83	0,78	0,75	0,59	0,63	0,55	0,73	0,75
Position(2,S)	0,90	0,83	0,68	0,82	0,88	0,86	0,77	0,78	0,56	0,50	0,53	0,74	0,75
Position(2,N)	0,90	0,83	0,68	0,82	0,88	0,86	0,77	0,78	0,56	0,50	0,53	0,74	0,75
MaxCurv(0,N)	0,81	0,76	0,71	0,80	0,85	0,83	0,78	0,75	0,59	0,64	0,54	0,73	0,75
Position(1,S)	0,84	0,83	0,66	0,83	0,87	0,86	0,81	0,79	0,59	0,50	0,53	0,74	0,75
Position(1,N)	0,84	0,83	0,66	0,83	0,87	0,86	0,81	0,79	0,59	0,50	0,53	0,74	0,75
Normal(2,S)	0,91	0,93	0,52	0,87	0,86	0,84	0,87	0,72	0,61	0,56	0,60	0,75	0,75
Normal(2,N)	0,91	0,93	0,52	0,87	0,86	0,84	0,87	0,72	0,61	0,56	0,60	0,75	0,75
ShapeIndex(0,S)	0,81	0,61	0,69	0,79	0,85	0,82	0,78	0,75	0,63	0,62	0,52	0,71	0,75
Normal(0,S)	0,84	0,49	0,61	0,83	0,87	0,85	0,82	0,77	0,58	0,57	0,59	0,71	0,75
Normal(0,N)	0,84	0,49	0,61	0,83	0,87	0,85	0,82	0,77	0,58	0,57	0,59	0,71	0,75
MiniCurv(1,S)	0,81	0,67	0,60	0,83	0,87	0,84	0,80	0,75	0,58	0,62	0,54	0,72	0,74
Laplacian(1,S)	0,90	0,52	0,53	0,82	0,87	0,84	0,78	0,76	0,54	0,67	0,54	0,71	0,74
Laplacian(1,N)	0,90	0,52	0,53	0,82	0,87	0,84	0,78	0,76	0,54	0,67	0,54	0,71	0,74
MiniCurv(0,N)	0,82	0,59	0,67	0,79	0,85	0,83	0,78	0,76	0,54	0,59	0,53	0,71	0,74
GausCurv(2,N)	0,79	0,89	0,73	0,84	0,88	0,84	0,80	0,63	0,51	0,62	0,54	0,73	0,74
Dihedral(0,S)	0,80	0,45	0,68	0,80	0,86	0,84	0,79	0,76	0,52	0,57	0,57	0,70	0,74
Dihedral(0,N)	0,80	0,45	0,68	0,80	0,86	0,84	0,79	0,76	0,52	0,57	0,57	0,70	0,74
MeanCurv(0,S)	0,76	0,58	0,63	0,81	0,86	0,84	0,79	0,76	0,54	0,63	0,54	0,70	0,74
GausCurv(2,S)	0,73	0,91	0,68	0,84	0,90	0,86	0,81	0,70	0,50	0,60	0,57	0,74	0,74
Dihedral(1,S)	0,89	0,50	0,52	0,82	0,86	0,83	0,79	0,74	0,54	0,61	0,55	0,70	0,74

Table A.1: (continued)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	Mean 11 objects	Mean 9 objects
Dihedral(1,N)	0,89	0,50	0,52	0,82	0,86	0,83	0,79	0,74	0,54	0,61	0,55	0,70	0,74
MiniCurv(1,N)	0,88	0,68	0,65	0,79	0,85	0,81	0,77	0,75	0,52	0,57	0,53	0,71	0,73
ShapeIndex(0,N)	0,77	0,63	0,65	0,79	0,84	0,82	0,77	0,75	0,62	0,59	0,51	0,70	0,73
Curvedness(0,S)	0,73	0,53	0,63	0,81	0,86	0,84	0,79	0,76	0,54	0,62	0,56	0,70	0,73
Laplacian(0,S)	0,79	0,65	0,60	0,80	0,86	0,85	0,77	0,77	0,50	0,65	0,54	0,71	0,73
Laplacian(0,N)	0,79	0,65	0,60	0,80	0,86	0,85	0,77	0,77	0,50	0,65	0,54	0,71	0,73
MiniCurv(0,S)	0,77	0,52	0,62	0,81	0,86	0,84	0,79	0,77	0,54	0,60	0,53	0,69	0,73
MaxCurv(0,S)	0,72	0,49	0,63	0,81	0,86	0,84	0,79	0,76	0,55	0,61	0,57	0,69	0,73
GausCurv(1,S)	0,66	0,76	0,58	0,83	0,89	0,87	0,81	0,79	0,52	0,56	0,56	0,71	0,73
FMPD(0,S)	0,75	0,62	0,56	0,80	0,85	0,82	0,79	0,74	0,63	0,60	0,60	0,70	0,73
Laplacian(2,S)	0,87	0,59	0,52	0,81	0,83	0,79	0,76	0,70	0,53	0,69	0,63	0,70	0,72
Laplacian(2,N)	0,87	0,59	0,52	0,81	0,83	0,79	0,76	0,70	0,53	0,69	0,63	0,70	0,72
GausCurv(0,N)	0,73	0,60	0,55	0,81	0,87	0,85	0,80	0,77	0,52	0,60	0,54	0,69	0,72
MiniCurv(2,S)	0,73	0,92	0,68	0,82	0,85	0,81	0,79	0,65	0,51	0,63	0,51	0,72	0,72
FMPD(0,N)	0,73	0,55	0,60	0,77	0,82	0,77	0,76	0,71	0,68	0,63	0,66	0,70	0,72
ShapeIndex(2,S)	0,68	0,88	0,74	0,78	0,83	0,79	0,75	0,64	0,60	0,62	0,52	0,71	0,71
GausCurv(0,S)	0,65	0,63	0,57	0,81	0,87	0,85	0,78	0,77	0,57	0,55	0,58	0,69	0,71
Saliency(1,N)	0,84	0,66	0,64	0,77	0,81	0,77	0,74	0,65	0,54	0,63	0,53	0,69	0,71
Laplacegaus(0,N)	0,65	0,73	0,55	0,82	0,88	0,86	0,81	0,77	0,51	0,56	0,61	0,70	0,71
MiniCurv(2,N)	0,79	0,89	0,71	0,79	0,81	0,76	0,75	0,63	0,56	0,58	0,52	0,71	0,71
ShapeIndex(2,N)	0,68	0,85	0,73	0,78	0,83	0,78	0,74	0,64	0,57	0,60	0,51	0,70	0,71
Dihedral(2,S)	0,85	0,59	0,50	0,80	0,76	0,77	0,76	0,65	0,58	0,65	0,57	0,68	0,70
Dihedral(2,N)	0,85	0,59	0,50	0,80	0,76	0,77	0,76	0,65	0,58	0,65	0,57	0,68	0,70
Laplacegaus(0,S)	0,56	0,75	0,58	0,82	0,84	0,82	0,76	0,77	0,58	0,56	0,66	0,70	0,70

Table A.1: (continued)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	Mean 11 objects	Mean 9 objects
Position(0,S)	0,76	0,83	0,60	0,79	0,83	0,80	0,76	0,73	0,50	0,50	0,56	0,70	0,70
Position(0,N)	0,76	0,83	0,60	0,79	0,83	0,80	0,76	0,73	0,50	0,50	0,56	0,70	0,70
Saliency(1,S)	0,72	0,45	0,53	0,79	0,83	0,79	0,76	0,65	0,56	0,60	0,54	0,66	0,69
Saliency(2,N)	0,85	0,91	0,63	0,75	0,71	0,68	0,68	0,67	0,54	0,61	0,50	0,69	0,68
Saliency(2,S)	0,83	0,90	0,55	0,74	0,71	0,68	0,69	0,62	0,53	0,62	0,57	0,68	0,66

Mean and global AUC values when considering the 11 objects (instead of 9, as presented in Chapter 5):

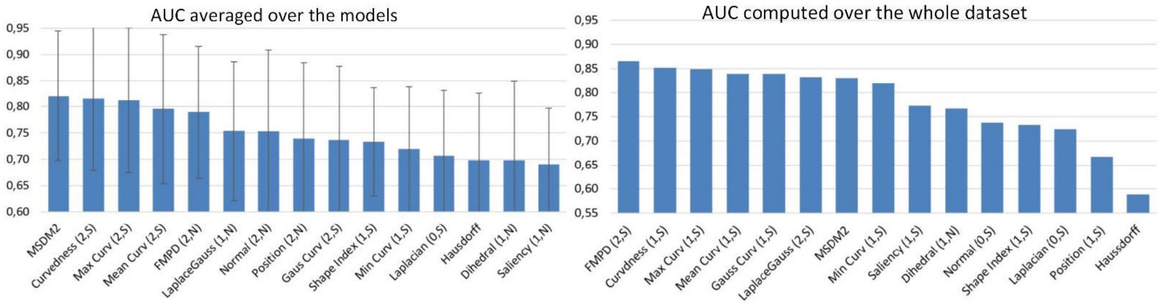


Figure A.4: Left: Mean AUC values for all the tested features. Error bars denote standard deviations. Right: AUC values when computing ROC curves on the whole dataset.

A supplementary illustration of subjective map and attributes:

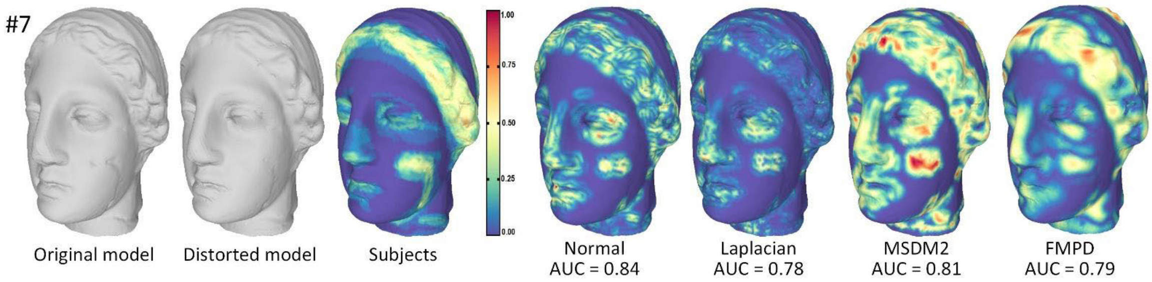


Figure A.5: Reference and distorted models from our dataset (objects #7), observer data and distortion maps from different attributes and metrics (colors are mapped in the min-max range).

Appendix B

Supplementary materials for Chapter 6

B.1 Wireframes and texture seams of our 3D models

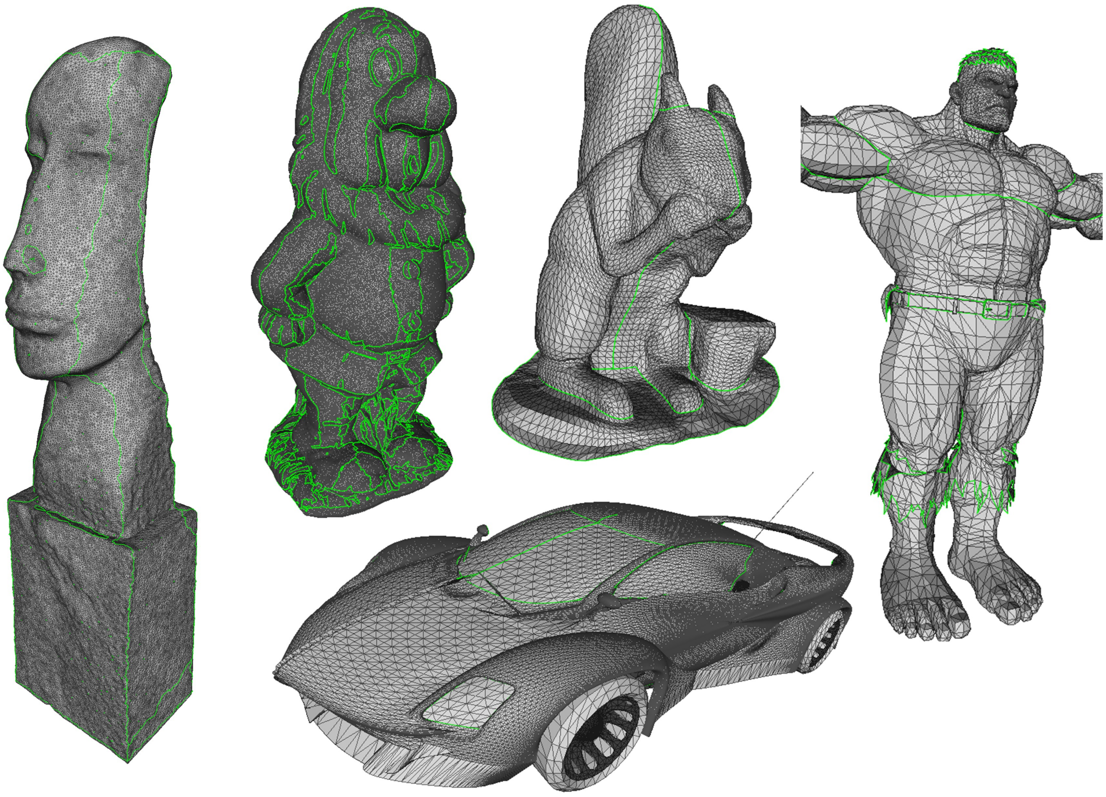


Figure B.1: Wireframes and texture seams (in green)

B.2 Distortions on the Dward model

ID	Texture Distortion		Geometric Distortion	
D1	JPEG	6% quality	Smoothing	25 iterations
D2	JPEG	6% quality	Smoothing	50 iterations
D3	JPEG	6% quality	Quantization	10 bits
D4	JPEG	6% quality	Quantization	8 bits
D5	JPEG	6% quality	Simplification	92% removed
D6	JPEG	6% quality	Simplification	98.7% removed
D7	JPEG	8% quality	Smoothing	25 iterations
D8	JPEG	8% quality	Smoothing	50 iterations
D9	JPEG	8% quality	Quantization	10 bits
D10	JPEG	8% quality	Quantization	8 bits
D11	JPEG	8% quality	Simplification	92% removed
D12	JPEG	8% quality	Simplification	98.7% removed
D13	JPEG	10% quality	Smoothing	25 iterations
D14	JPEG	10% quality	Smoothing	50 iterations
D15	JPEG	10% quality	Quantization	10 bits
D16	JPEG	10% quality	Quantization	8 bits
D17	JPEG	10% quality	Simplification	92% removed
D18	JPEG	10% quality	Simplification	98.7% removed
D19	Sub-sampling	3% sampled	Smoothing	25 iterations
D20	Sub-sampling	3% sampled	Smoothing	50 iterations
D21	Sub-sampling	3% sampled	Quantization	10 bits
D22	Sub-sampling	3% sampled	Quantization	8 bits
D23	Sub-sampling	3% sampled	Simplification	92% removed
D24	Sub-sampling	3% sampled	Simplification	98.7% removed
D25	Sub-sampling	5% sampled	Smoothing	25 iterations
D26	Sub-sampling	5% sampled	Smoothing	50 iterations
D27	Sub-sampling	5% sampled	Quantization	10 bits
D28	Sub-sampling	5% sampled	Quantization	8 bits
D29	Sub-sampling	5% sampled	Simplification	92% removed
D30	Sub-sampling	5% sampled	Simplification	98.7% removed
D31	Sub-sampling	8% sampled	Smoothing	25 iterations
D32	Sub-sampling	8% sampled	Smoothing	50 iterations
D33	Sub-sampling	8% sampled	Quantization	10 bits
D34	Sub-sampling	8% sampled	Quantization	8 bits
D35	Sub-sampling	8% sampled	Simplification	92% removed
D36	Sub-sampling	8% sampled	Simplification	98.7% removed

Table B.1: Details about the distortions applied on the *Dwarf* model.

B.3 Illustration of our sorting algorithm

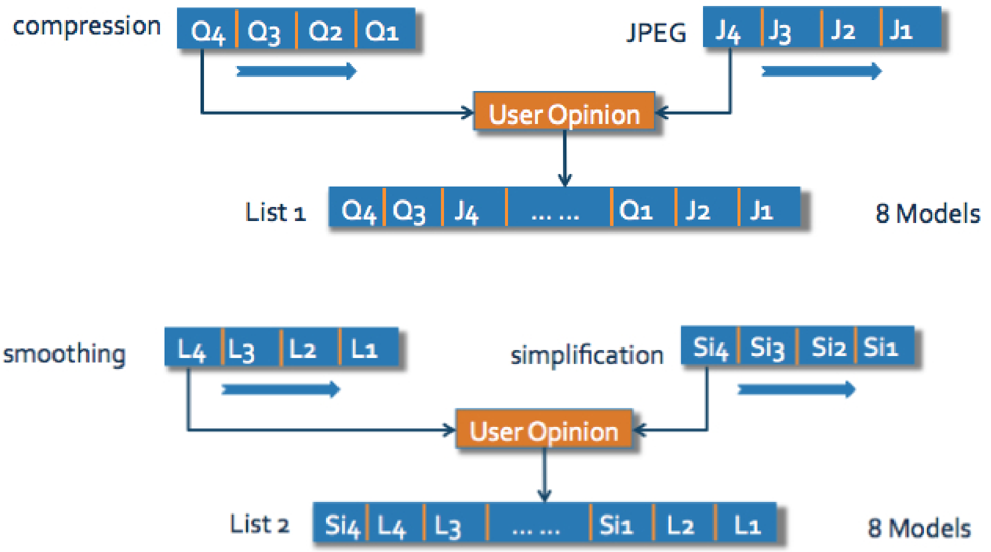


Figure B.2: **Step 1** Two groups (e.g., *Q* and *J*) are randomly chosen. The subject firstly compares *Q4* and *J4*. When the subject determines the model closer to the reference (*J4*), the index of the other model (*Q4*) will be pushed into a list (*List1*) as the worst quality model. Then the next two models will be shown. In this round, the selected model from previous round (*J4*) and a distorted model with a decreased strength from the other group (*Q3*) are shuffled and displayed to the subject. Then, following this way, we repeat the comparisons until all 8 models are sorted from the worst visual quality to the best into *List 1*. The same process is conducted between two other groups (e.g., *L* and *Si*), and another list (*List 2*) is obtained from these two groups.

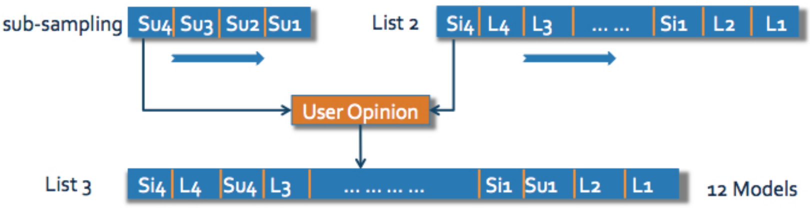


Figure B.3: **Step 2** One list among *List 1* and *List 2* is randomly chosen (e.g., *List 2*), in which all the models are already sorted by their visual qualities. This list will be merged with the remaining group (*Su*). Following the procedure detailed in step 1, these two sets of models are interleaved into *List 3*.

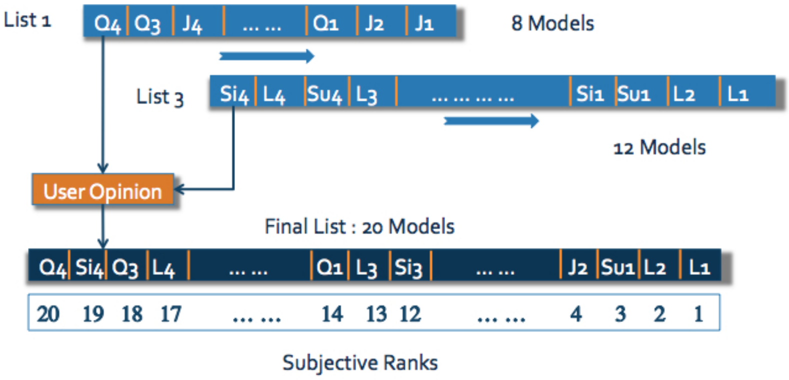


Figure B.4: **Step 3** Finally the remaining lists (*List 3*) and (*List 1*) are merged with the same process than before, to obtain the *Final List*, which contains the 20 models ranked from the worst visual quality to the best (1 means the best, and 20 means the worst).

B.4 Subjective scores

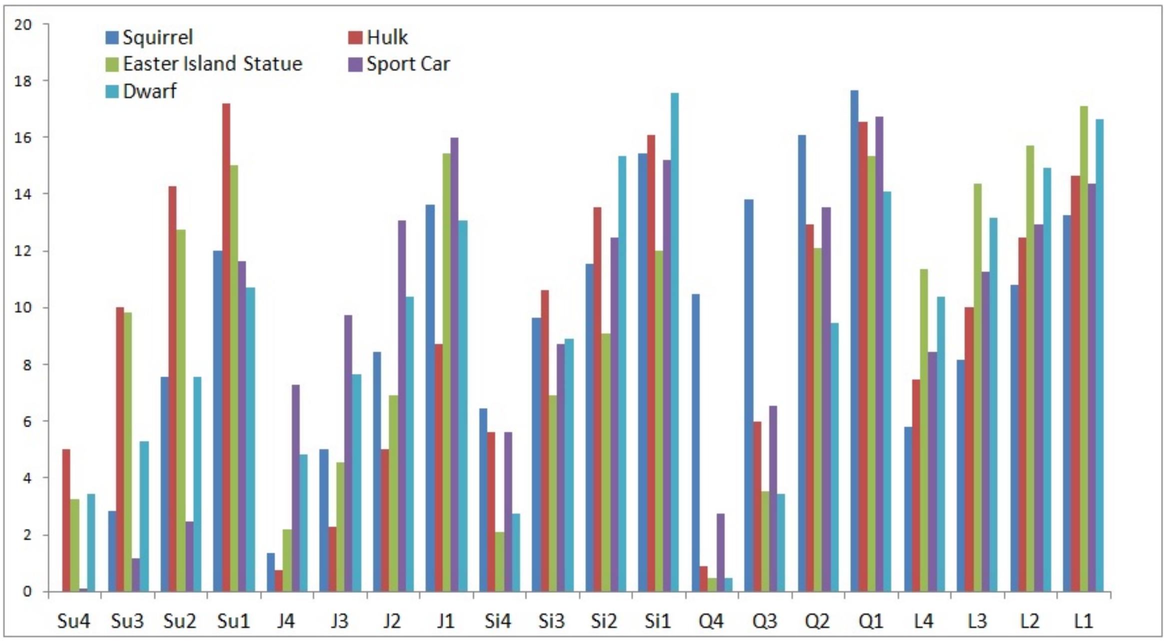


Figure B.5: Subjective vote scores for all the 100 distorted models (single-type distortion setting) for the rendering with shading. Higher scores mean better visual quality..

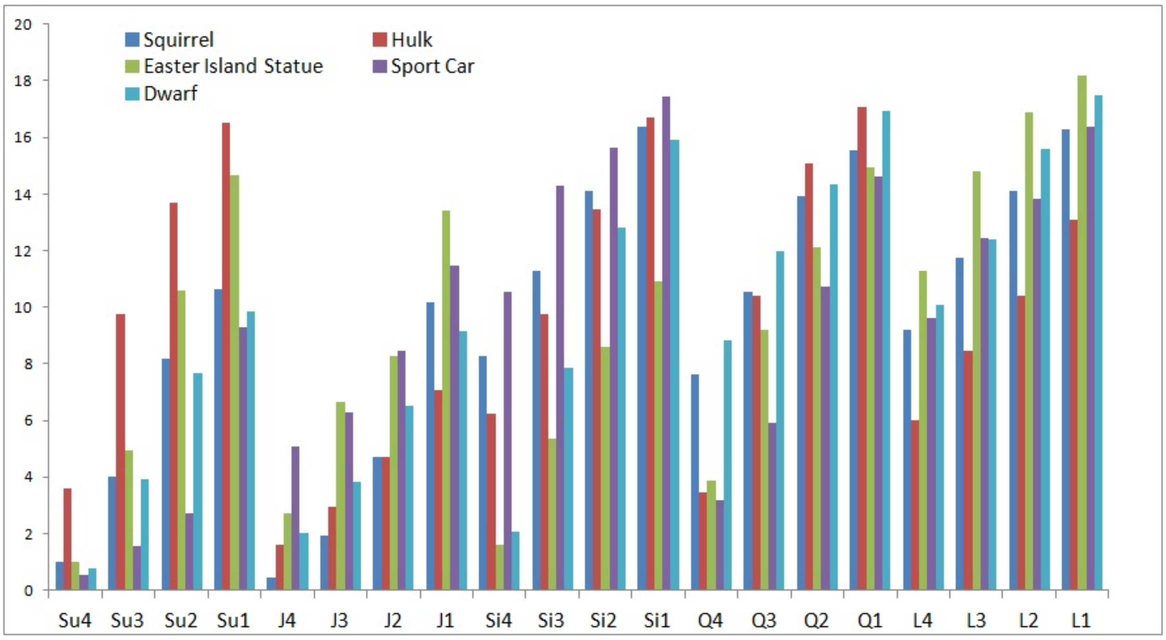


Figure B.6: Subjective vote scores for all the 100 distorted models (single-type distortion setting) for the rendering without shading. Higher scores mean better visual quality.

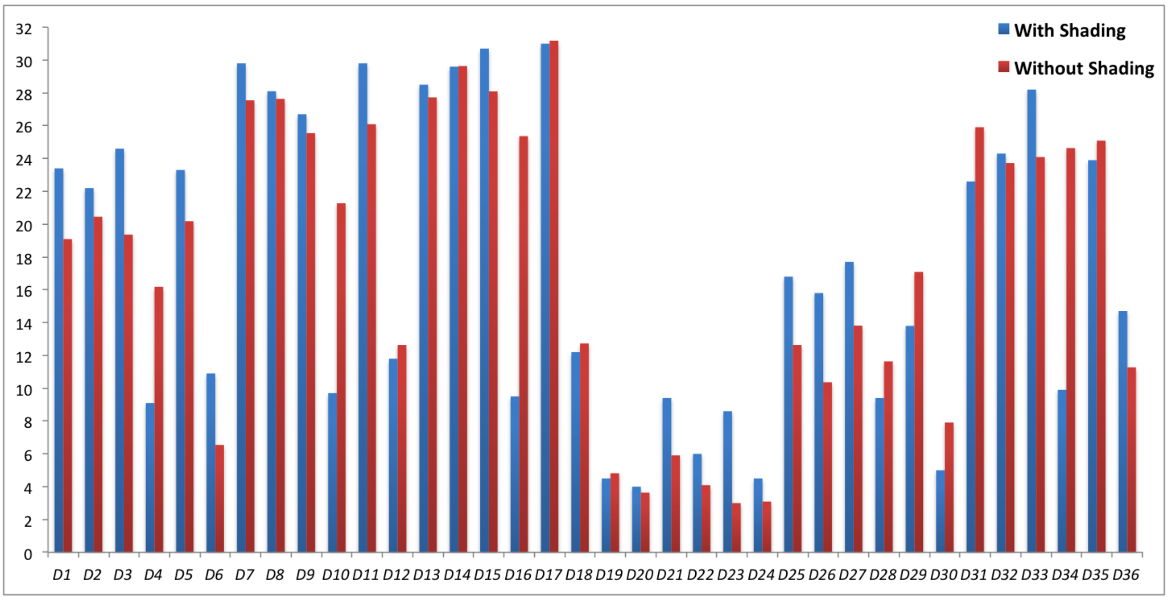


Figure B.7: Subjective vote scores for the 36 distorted *Dwarf* models (mixed-type distortion setting) for the renderings with shading and without shading respectively. Higher scores mean better visual quality.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : GUO DATE de SOUTENANCE : 06/10/2016
(avec précision du nom de jeune fille, le cas échéant)

Prénoms : Jinjiang

TITRE :
Contributions to objective and subjective visual quality assessment of 3D models

NATURE : Doctorat Numéro d'ordre : 2016LYSEI099

Ecole doctorale : INFORMATIQUE ET MATHEMATIQUES DE LYON

Spécialité : Informatique

RESUME :

Dans le domaine de l'informatique graphique, les données tridimensionnelles, généralement représentées par des maillages triangulaires, sont employées dans une grande variété d'applications (par exemple, le lissage, la compression, le remaillage, la simplification, le rendu, etc.). Cependant, ces procédés introduisent inévitablement des artefacts qui altèrent la qualité visuelle des données 3D rendues. Ainsi, afin de guider perceptuellement les algorithmes de traitement, il y a un besoin croissant d'évaluations subjectives et objectives de la qualité visuelle à la fois performantes et adaptées, pour évaluer et prédire les artefacts visuels. Dans cette thèse, nous présentons d'abord une étude exhaustive sur les différentes sources d'artefacts associés aux données numériques graphiques, ainsi que l'évaluation objective et subjective de la qualité visuelle des artefacts. Ensuite, nous introduisons une nouvelle étude sur la qualité subjective conçue sur la base de l'évaluations de la visibilité locale des artefacts géométriques, dans laquelle il a été demandé à des observateurs de marquer les zones de maillages 3D qui contiennent des distorsions visibles. Les cartes de distorsion visuelle collectées sont utilisées pour illustrer plusieurs fonctionnalités perceptuelles du système visuel humain (HVS), et servent de vérité-terrain pour évaluer les performances des attributs et des mesures géométriques bien connus pour prédire la visibilité locale des distorsions. Notre deuxième étude vise à évaluer la qualité visuelle de modèles 3D texturés, subjectivement et objectivement. Pour atteindre ces objectifs, nous avons introduit 136 modèles traités avec à la fois des distorsions géométriques et de texture, mené une expérience subjective de comparaison par paires, et invité 101 sujets pour évaluer les qualités visuelles des modèles à travers deux protocoles de rendu. Motivés par les opinions subjectives collectées, nous proposons deux mesures de qualité visuelle objective pour les maillages texturés, en se fondant sur les combinaisons optimales des mesures de qualité issues de la géométrie et de la texture. Ces mesures de perception proposées surpassent leurs homologues en termes de corrélation avec le jugement humain.

MOTS-CLÉS :
Informatique graphique, Maillages 3D, Artefact visuel, Maillage texturé, Évaluation de la Qualité Visuelle des Maillages, Évaluation de la Qualité des Images, Métriques perceptuelles, Étude subjective.

Laboratoire (s) de recherche :
Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)

Directeur de thèse:
LAVOUÉ Guillaume, Maître de Conférences, HDR, LIRIS, INSA Lyon

Président de jury :

Composition du jury :

DAOUDI Mohamed, Professeur des Universités, LIFL, TELECOM Lille 1
PUECH William, Professeur des Universités, LIRMM
DANIEL Marc, Professeur des Universités, LSIS, Polytech Marseille
LARABI Mohamed-Chaker, Maître de Conférences, XLIM-SIC
LAVOUÉ Guillaume, Maître de Conférences, HDR, LIRIS, INSA Lyon
VIDAL Vincent, Maître de Conférences, LIRIS, Université Lyon 1
BASKURT Atilla, Professeur des Universités, LIRIS, INSA Lyon