Geo-metric: A Perceptual Dataset of Distortions on Faces

KRZYSZTOF WOLSKI, MPI Informatik & Meta, Germany
LAURA TRUTOIU, Meta, USA
ZHAO DONG, Meta, USA
ZHENGYANG SHEN, Meta, USA
KEVIN MACKENZIE, Meta, USA
ALEXANDRE CHAPIRO, Meta, USA

Reference	High frequency noise	Low frequency noise	Simplification	Smoothing
	-2.220 JOD probability: 0.93	-0.924 JOD probability: 0.73	-2.875 JOD probability: 0.97	-0.789 JOD probability: 0.70

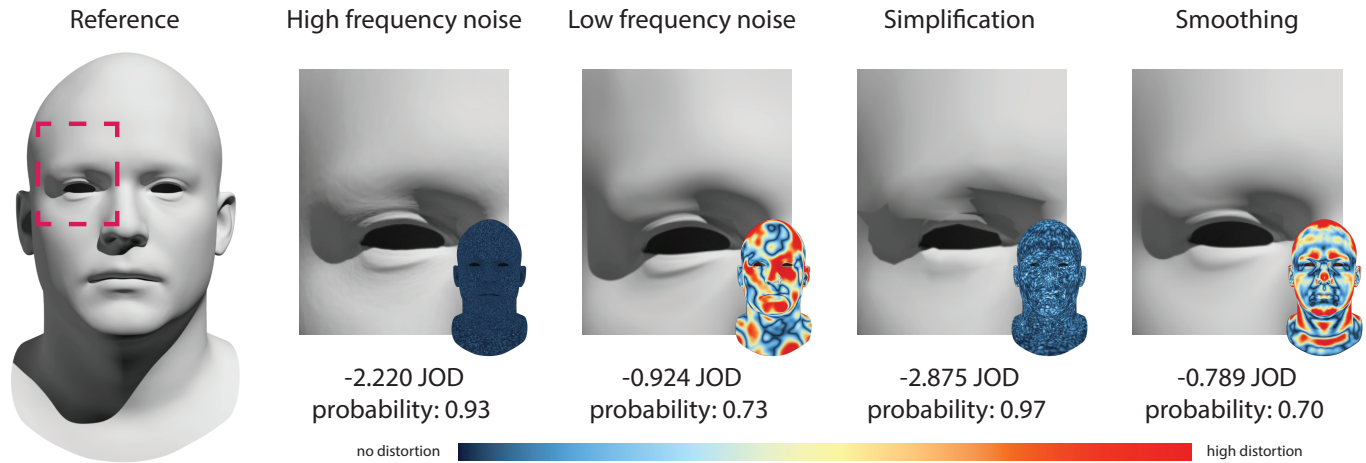no distortion ▬▬▬▬▬ high distortion

Fig. 1. A sample of the artifacts studied presented in our subjective study. Heatmaps show the scale of each distortion, expressed as Euclidean distance to the reference. Just-objectionable-difference (JOD) values give an estimate of the perceived magnitude, and can be converted to a probability of selection, shown below.

In this work we take a novel perception-centered approach to quantify distortions on 3D geometry of faces, to which humans are particularly sensitive. We generated a dataset, composed of 100 high-quality and demographically-balanced face scans. We then subjected these meshes to distortions that cover relevant use cases in computer graphics, and conducted a large-scale perceptual study to subjectively evaluate them. Our dataset consists of over 84,000 quality comparisons, making it the largest ever psychophysical dataset for geometric distortions. Finally, we demonstrated how our data can be used for applications like metrics, compression, and level-of-detail rendering.

CCS Concepts: • **Computing methodologies → Perception**.

Additional Key Words and Phrases: perception, psychophysics, geometry processing, faces

Authors' addresses: Krzysztof Wolski, kwolski@mpi-inf.mpg.de, MPI Informatik & Meta, Germany; Laura Trutoiu, auract@fb.com, Meta, USA; Zhao Dong, zhaodong@fb.com, Meta, USA; Zhengyang Shen, zyshen@live.unc.edu, Meta, USA; Kevin MacKenzie, kjmacken@fb.com, Meta, USA; Alexandre Chapiro, alex@chapiro.net, Meta, USA.

1 INTRODUCTION

Subjective quality datasets are a required component to develop models and algorithms which follow users' preferences and generate high quality results. Creating these datasets is a challenging task, which requires comprehensive understanding of perception, psychophysics, statistical analysis, and modeling. Although large-scale perceptual databases such as LIVE [2005] and TID [2015] exist for image and video applications, many other areas of graphics lack this support.

In this work, we study the perception of geometric distortions on faces. Human characters or avatars are ubiquitous in gaming, entertainment, and social applications. Faithful reproduction of faces is a particularly important and challenging task [Zhao et al. 2003], as people are known to be especially sensitive to distortions of facial features. Although these distortions can be measured objectively, this does not provide a sense of how users perceive their impact. Quantifying this is important because perceptually-driven methods in geometry processing can result in better improved outcomes for key applications by selecting optimal cost vs. quality tradeoffs.

We begin our work by creating a novel dataset consisting of 100 high-quality face meshes, taking special care to maintain demographic balance. We proceed to apply a set of carefully chosen geometric distortions, as outlined in Figure 1. We measure the perceived magnitudes of these artifacts in a large-scale user study, which gathered over 84,000 subjective judgements. We used modern statistical methods to convert our study's results into a standardized database of perceptual quality, which can be used directly for relevant applications.

We demonstrate some promising use cases for our perceptual database. First, we evaluate the performance of existing geometry metrics against our ground-truth data. Next, we propose a simple data-driven retargeting scheme to augment these metrics. Finally, we leverage the augmented metrics to perceptually tune algorithms like mesh compression and level-of-detail rendering (LOD).

Our contributions are as follows:

- A large-scale perceptual database of the visibility of geometric distortions on faces
- A dataset of meshes representing human faces that follows the demographic distribution of the US census [2021], ensuring diversity
- Applications of our database, such as evaluating metrics and setting perceptually meaningful thresholds for tasks like mesh compression and LOD rendering.

Our perceptual database and benchmark dataset of reference and distorted faces are made available[1] with the goal of encouraging further research into perceptual aspects of computational geometry.

2 RELATED WORK

2.1 Quality metrics

Visual difference metrics for images or videos are widely available, from popular metrics like PSNR and SSIM [2010] to complex metrics that model components of the human visual system like HDR-VDP [2011] and Fov-Video-VDP [2021]. These image and video difference metrics operate in "pixel space", i.e. the objects of comparison are the final rendered images, and do not address geometric distortions on underlying meshes, if present. In contrast to this, it is often desirable to evaluate the quality of 3D meshes at an earlier stage of the graphics pipeline, while staying agnostic to illumination and texture, long before the final image is rendered.

3D mesh quality metrics tackle this problem. A popular geometry metric is Metro [Cignoni et al. 1998], which uses a Hausdorff distance as a base component. Metro is mathematically motivated, and does not aim to reproduce perception. In contrast, metrics like MSDM [Lavoué et al. 2006] are inspired by perceptually-motivated image metrics like SSIM, and use mean curvature as input—informing on the smooth vs. bumpy profile of the studied object. Local distortions are measured as a difference of Gaussian-weighted statistics over a local spherical neighborhood, with the final score being computed through Minkowski pooling. MSDM2 [Lavoué 2011] improves upon the previous version by projecting the curvature value between two objects and using a multiscale

approach, allowing users to compare meshes with different topologies. More advanced perceptual models inspired the authors of DAME [Váša and Rus 2012], which takes visual masking into account. Following the intuition that distortions introduced to bumpy areas are less visible than ones added to smooth surfaces, the authors calculate dihedral angles and weigh them accordingly. This metric is also significantly faster than MSDM, making it more suitable for computationally intensive tasks like optimization. Another fast roughness-based method is FMPD [Wang et al. 2012], which computes the roughness of each vertex on the test and reference meshes using local Gaussian curvature, modulated by a power function to model visual masking. The final score is computed by taking the difference between normalized surface integrals of each mesh. This method is unsuitable for low-frequency distortions, as their impact can be underestimated.

TPDMPW [Feng et al. 2018] proposes a novel spatial pooling method to assess mesh quality, based on applying a percentile weighting strategy to existing metrics to emphasize local regions with severe distortions. Nouri et al. [2016] propose a multi-scale saliency map as a basis for local statistics on 3D meshes, computing differences in mean local roughness to model visual masking. Nader et al. [2015] propose a just-noticeable-difference-based 3D geometry distortion metric, developed for flat-shaded geometries. Authors employ a sensitivity paradigm, which models perceptually relevant parameters like distance and masking. Unfortunately, we were not able to obtain working implementations of these methods, stopping us from analyzing them further in this work.

2.2 Perceptual data and scaling

Several metrics described in the previous section are based on existing perceptual studies of geometric distortion. The DAME [Váša and Rus 2012] dataset consists of 65 distorted meshes, ≈2,800 perceptual data points. The MSDM/LIRIS [Lavoué et al. 2006] datasets range from 11-88 meshes, and ≈1,000 data points. In contrast, our distortion dataset is more than 30 times larger at 2,500 distorted meshes, and consists of 84,000 measurements. We are inspired by perceptual data in fields like imaging, where large-scale collection is ubiquitously used (e.g. the popular TID2013 [Ponomarenko et al. 2015] dataset consists of over 3,000 images, ≈1,500,000 data points).

The output of the 3D geometry metrics in the previous section are arbitrarily scaled, and cannot be easily contrasted against the results of a subjective study (with the exception of the work of Nader et al. [2015]). To make our dataset more useful as a tool for metric calibration, our subjective data is scaled into units of *just objectionable difference* (JOD) [Perez-Ortiz et al. 2019], closely related to the psychophysical concept of a *just noticeable difference.* An advantage of using a JOD scale is that it provides a universal perceptual scale of quality that can be directly related to the probability a distorted mesh will be chosen over another in an experiment, as shown in Figure 2. We expect this type of perceptually-meaningful scale to be more interpretable and to lend itself to setting useful perceptual thresholds in applications, as this type of scaling is already popular for image and video difference metrics [Mantiuk et al. 2011, 2021].

---

[1]https://github.com/facebookresearch/Geo-metric

Fig. 2. This plot shows how pairwise comparison preferences are represented in *just objectionable difference* (JOD) scores. For example, if a distorted mesh **A** is 1 JOD away from the reference, this represents a 50% decrease in likelihood of selection, i.e. the reference will be chosen over **A** 75% of the time. More detail can be seen in Sec. 3.9 of the work of Mantiuk et al. [2021].

2.3 Face perception and datasets

Several datasets of 3D faces have been presented in the literature, categorized as real (captured) and synthesized. The former are typically obtained using 3D scanners. Cao et al. [2014] use a Kinect to scan 150 individuals and 20 expressions. Due to the low precision of the depth estimates, the detailed facial geometry such as wrinkles is lost. Zhang and colleagues [2013] use a Di3D dynamic face capturing system to capture a 3D video database of spontaneous facial expressions from 41 young adults. Bagdanov et al. [2011] capture 3D scans from 53 Caucasian adult individuals using the 3dMDFace system. Ranjan et al. [2018] released a database including 12 individuals with 12 extreme expression sequences. To explore lip synchronization and realistic facial animation, Cudeiro and colleagues [2019] released a 3D audio-visual dataset, captured from 12 individuals with 40 dynamic sequences for each. Zhu et al. [2021] present a large-scale detailed 3D face dataset, captured from 938 Asian participants, each with 20 specific expressions. Although these datasets provide good coverage over age and gender, they rarely take other demographic factors into consideration. Yin et at. [2006] aims to capture expressions of 100 subjects from several demographic groups, providing 2500 unique face meshes with high-quality textures and low-detail geometry. In a follow-up study [2008], the authors add time-varying data, providing 606 3D expression sequences with 100 frames each.

Unlike captured geometry, synthetic datasets are often generated using a 3D Morphable Model (3DMM) [1999] using either a statistical appearance model [Guo et al. 2018; Paysan et al. 2009], FACS [2011], or inverse rendering [Booth et al. 2018]. Such datasets are often orders of magnitude larger, and can cover more granular data groups. However, reconstructed faces often suffer from loss of detail, making them less accurate representations of reality. Our synthetic data is based on high-resolution scans gathered using a 3dMD scanner. In addition, our population distribution aims to reflect diverse US demographics, based on the Census [2021], which ensures some representative data is present for each group.

3 SYNTHETIC DATASET

To measure the perceptual effects of distortions on 3D facial geometry, we first need to prepare a representative benchmark dataset. We do this by leveraging a large-scale private collection of human head scans. Although we are not able to use these proprietary scans in this work, we draw on this database to generate a synthetic set which blends the original scans into new meshes, while maintaining desirable properties like high quality and demographic balance.

3.1 Non-public dataset

Proprietary head meshes were gathered using a commercially available high-quality 3D scanner: 3dMD Ltd's static 12-viewpoint 3dMD-head scanning system. This is a top of the line scanner which includes 36 machine vision cameras in 12 Modular Camera Units (or viewpoints), optimized specifically for full head and ear detail capture. The system automatically generates a continuous 3D textured surface mesh with variable resolution (typically around 180k vertices). The 3dMDvultus Software is then used to add 3D landmarks to the meshes. The scanned objects preserve the original scale, with 1 unit corresponding to 1 millimeter. According to 3dMD [2], this type of system has a linear accuracy range of 0.2mm or better on a frame-by-frame basis.

A set of 1,417 heads were scanned. The raw scanned meshes can contain sampling and reconstruction artifacts. To drive direct comparisons and maintain standardization, they are registered to a custom artist-generated template mesh of head and neck geometry consisting of 7,306 vertices, which was originally designed for expression details. Each mesh in the dataset and the registration target were annotated with 6 landmarks. R3dS Wrap software[3] was used to register all original scans to the template, aligned using an Orthogonal Procrustes model [1966].

3.2 Public synthetic heads generation

We proceed to leverage the proprietary dataset in order to create a public dataset of 100 synthetic faces, which maintains desirable properties such as high quality and demographic distribution. Similarly to Dai et al. [2019], we use a principal component model (PCA) to generate our meshes [Blanz and Vetter 1999]. Given M pre-registered meshes, each with N vertices and the same connectivity, we can denote each individual scan as $x_i \in \mathbb{R}^{3N}$. We can then compute the mean shape over the set $\overline{x}$, and recenter each subject around this value $\overline{x_i} = x_i - \overline{x}$, with the new set of shapes $\overline{X} \in \mathbb{R}^{M \times 3N}$. Since in this case $M << 3N$, the covariance matrix $\overline{X}^T \overline{X}$ is rank deficient.

We eigen-decompose the covariance matrix, from which we get the basis as well as the corresponding eigenvalues, which we assume are sorted in descending order of magnitude. We select the top K eigenvalues and their corresponding basis vectors $b = \{b_1, ..., b_K\}$, where $b_k \in \mathbb{R}^{3N}$, as principal components for our statistical model. We selected $K = 196$ which explains 99% of the variance in the input dataset.

To generate synthetic heads, we blend L registered meshes. We begin by inversely solving the coefficients for each registered head

---

[2] https://3dmd.com/products/
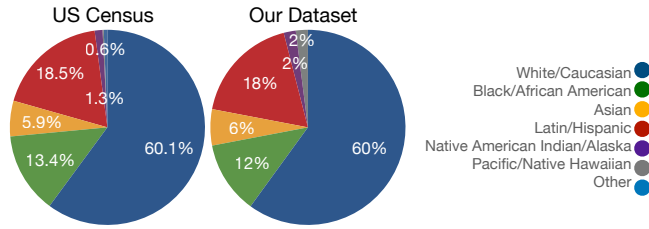[3] https://www.russian3dscanner.com/

Fig. 3. Comparison of the demographic distribution showing US census 2021 [2021] (L) and our dataset (R). Note that we ensure a minimum of one face is present for each demographic group and gender.
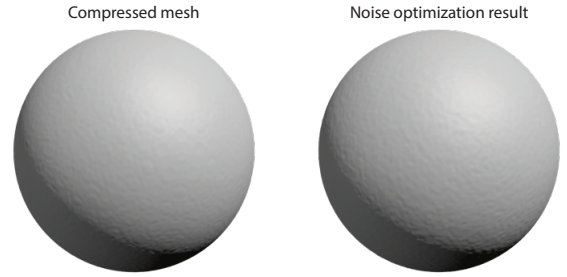

Fig. 4. Comparison of a compressed sphere (L) and the optimized noisy sphere match (R), which mimics the characteristics of the compression distortion using Perlin noise.

e_i, where $e_i \in \mathbb{R}^K$ refers to the coefficient vector of x_i over b. We generate a new synthetic set of coefficients $\hat{e} = \sum_{i=1}^{L} w_i e_i$, where $w_i \geq 0$, and $\sum_{i=1}^{L} w_i = 1$. The blended coefficients $\hat{e}$ are used to generate a new blended mesh $\hat{x} = b\hat{e} + \bar{x}$. To prevent shading artifacts, we generate a denser version of the blended mesh through subdivision by employing two applications of the Loop Subdivision Algorithm [1987], resulting in 116593 vertices per head.

Our PCA model was generated using 3D geometry from the entire private dataset of 1,417 participants (679F, 738M, aged between 18 and 65). For each of the 100 output synthetic heads, we blended $L = 3$ unique meshes (a total of 300), with the ratio set to a simple average $w_s = 1/3, \forall s$. These trios were set to belong in the same sex and race demographic category according to the US Census [2021], as shown in Figure 3.

3.3 Distortion Dataset

To study the perception of artifacts on geometry, we needed to generate a set of distorted meshes. We use the synthetic dataset composed of 100 heads described in Section 3.2 as our base. We augment this data by modifying it with three selected effects: *noise* (Section 3.3.1), *simplification* (Section 3.3.2), and *smoothing* (Section 3.3.3), as these artifacts are ubiquitous in applications [Botsch et al. 2007]. The parameters of each distortion were carefully selected based on perceptual and quantitative data. A sample set of distorted meshes from our dataset can be seen in Figure 6.

3.3.1 Noise. The first artifact we examine is *noise*, which is a common distortion that can stem from many sources in application. Noise can be parametrized by magnitude (how large the deviations are) and frequency (the size of the distortions relative to the mesh). As we want our dataset to be application-agnostic, we chose to add synthetic noise rather than noise stemming from a particular source. We use Perlin noise [2002] due to its simplicity, and because both frequency and magnitude can be easily controlled. We used a three-dimensional noise implementation provided in the SideFX Houdini[4] software suite.

Despite the flexibility of the Perlin noise framework, we need to select a discrete number of frequencies and magnitudes to evaluate in our study. To do this, we begin by searching for an upper bound on the noise frequency. In practice, high-frequency noise often stems from compression algorithms, and is especially notable in methods using direct coordinate quantization of vertices [Deering

---

[4]https://www.sidefx.com/

1995]. To estimate the frequency profile of this type of noise, we run an optimization which approximates a compressed sphere with a sphere distorted using Perlin noise at varying frequencies. Compression is done using the open-source Draco library[5]. As meshes typically consist of locally varying vertex densities, we run several comparisons on spheres with different, but uniform vertex density. Simple cost functions like RMSE do not converge, as features caused by compression and noise do not coincide spatially. Instead, we crafted a custom cost function based on the variance of Gaussian curvature (see details in the supplemental material). A result of this process is shown in Figure 4. We found that Perlin noise ranging in frequency from 0.5-2 cycles/mm corresponded to plausible fits across all examined densities. We proceed to set our frequency upper bound to 2 cycles/mm. A lower bound of 0.01 cycles/mm was selected manually based on a comparison with the spectral compression algorithm presented by Sorkine and colleagues [2003]. Their work specifically aims at preferentially distorting lower frequencies of a mesh, which are considered by the authors to be less visible. We proceed to interpolate between these two bounds in log space to obtain the four sample points used in our experiment (0.01, 0.06, 0.34, and 2 cycles/mm). A sample set of each frequency and amplitude of noise can be seen in Figure 6.

Next, we need to set the noise amplitudes. Our aim is to select amplitudes that are neither obvious nor invisible, as either would yield only trivial data in the study. To begin, we manually selected a noise amplitude that is clearly visible. We then generate 101 intermediary values by linearly interpolating between this value and the noise-free baseline. Next, we ran a small-scale pilot user study (N=2) for each noise frequency. The experimental setup was identical to the one used in our main study (see Section 4.2). We used a Minimum Expected Entropy Staircase experiment procedure [Saunders and Backus 2006], as implemented in Psychtoolbox[6], to find the amplitude that corresponds to the perceptual threshold (i.e. 1 just-objectionable-difference, JOD, away from the reference, as explained in Section 4.4). In addition, the staircase procedure generates a psychometric curve (shown in Figure 5). We observe that the curve saturates at approximately 3 JODs. We sample four linearly spaced points between this saturation point and the reference, which guarantees a perceptually significant range of values will be presented

---

[5]https://google.github.io/draco/
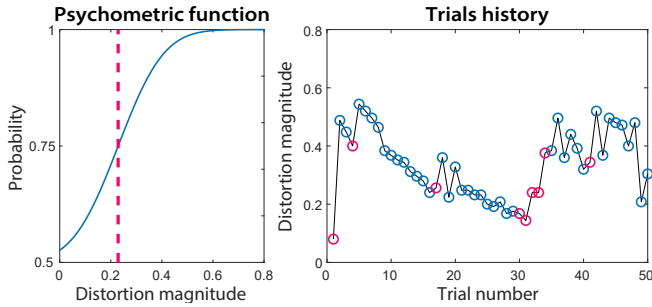[6]http://psychtoolbox.org/

Fig. 5. A pilot experiment was used to set the distortion magnitudes for our study. We show the user's responses converging towards the threshold in a staircase-type procedure on the right. Blue circles denote a correct answer, while pink ones show mistakes. On the left, a psychometric curve (continuous line) is fitted to the data, with a dashed line marking the threshold value of 1 JOD.

to participants. In total, this brings the number of noise trials to 16 (as each frequency is studied at four unique amplitude levels).

3.3.2 Simplification. The next effect we explored was *simplification*, which is relevant in the context of key applications like decimation, or level-of-detail rendering algorithms.

To select magnitudes of distortion, we followed the simple rule-of-thumb used in many LOD algorithms—reducing triangle count by half at every step [Luebke et al. 2003]. Next, we performed a staircase experiment identical to the one described in Section 3.3.1. We found that a single reduction level (i.e. 50%) was well-below the perceptual threshold. As a consequence, we selected the first level of simplification as 25%, and followed the halving rule to set four magnitudes of simplification (25%, 12.5%, 6.25%, and 3.125%).

3.3.3 Smoothing. The third and final artifact we explored was *smoothing*. In applications, it can stem from simplification, errors during mesh reconstruction, mesh decimation, volume-based remeshing algorithms, etc. To simulate different magnitudes of smoothing, we employed a Laplacian Smoothing algorithm [Witkin 1987] with fixed boundaries, as implemented in SideFX Houdini. The default value of $\lambda = 0.5$ was used.

In order to select the magnitude of smoothing, we once again employed a staircase pilot experiment (identical to Section 3.3.1). The expert-led experiment showed even a single application of Laplacian smoothing as above the threshold of noticeability. However, unlike other distortions, further testing with naïve users found they were significantly less sensitive to the artifact, which led us to set the first level of magnitude at 2 repetitions. As more aggressive smoothing is needed to produce changes on a smoother surface, we set the remaining values based on extensive pilot experiments. The final values used in our main study are 2, 6, 12, 20 repetitions of Laplacian smoothing.

3.3.4 Overview. We begin with 100 reference meshes of faces, which are distorted using 6 different types of artifacts (4 types of noise, smoothing and simplification). Each artifact has 4 levels of magnitude, resulting in 24 distorted meshes per reference, or 2400 distorted

Table 1. Parameters of the Experimental Conditions

Distortion	Magnitude measure	Magnitude			
		Level1	Level2	Level3	Level4
Noise - lowest freq.	amplitude	5.00	8.33	11.66	15.00
Noise - low freq.	amplitude	0.21	0.35	0.49	0.63
Noise - high freq.	amplitude	0.0675	0.1125	0.1575	0.2025
Noise - highest freq.	amplitude	0.0740	0.1233	0.1727	0.222
Simplification	% of remaining triangles	25.0	12.5	6.25	3.125
Smoothing	# iterations	2	6	12	20

meshes in total. A breakdown of the exact distortions used can be seen in Table 1.

4 USER STUDY

In order to gather perceptual data, we ran a large-scale user study, detailed in this section. This is the largest psychophysical collection of geometric distortions ever obtained, and is composed of 84,000 subjective comparisons.

4.1 Participants

In observance of restrictions for in-person work during the COVID pandemic, this study was conducted remotely. 50 male and 50 female paid participants were recruited (aged between 18 and 55 years). Our study received ethical approval through a third party institutional review board[7], and each of the 100 paid participants signed an informed consent form. Experimental output was anonymized.

Each experiment session consisted of 280 randomly selected trials over a batch of 10 unique heads. Our dataset of 100 synthetic head meshes was randomly split into 10 batches of 10 meshes each, with 10 participants assigned to each batch. In pilots, participants needed approximately 10 seconds per trial, resulting in an estimated 47 minutes required to finish the main portion of each session, which was deemed adequate to avoid fatigue. Each participant completed 3 experimental sessions each on different days.

4.2 Stimulus rendering

Our experiment aims to measure the perceived strength of distortions, but additional factors like illumination and texture can be expected to influence the result [O'Shea et al. 2008; Serrano et al. 2021]). As there are too many possible configurations of texture and illumination, we choose to perform our experiment using representative conditions that are expected to maximize the subjects' overall sensitivity, following guidelines set in the literature [Lavoué et al. 2016; Rogowitz and Rushmeier 2001]. We use a distant top-right light source, mimicking natural environmental conditions [O'Shea et al. 2008; Sun and Perona 1998]. Because this can leave the lower portion of the head unlit, we add a lower intensity light from the bottom-left direction, avoiding unnaturally strong shadows. We do not use any texture, with the heads rendered with a single uniform neutral gray color. This is expected to maximize sensitivity to distortions, as overlapping signals from texture can cause masking [Ferwerda et al. 1997].

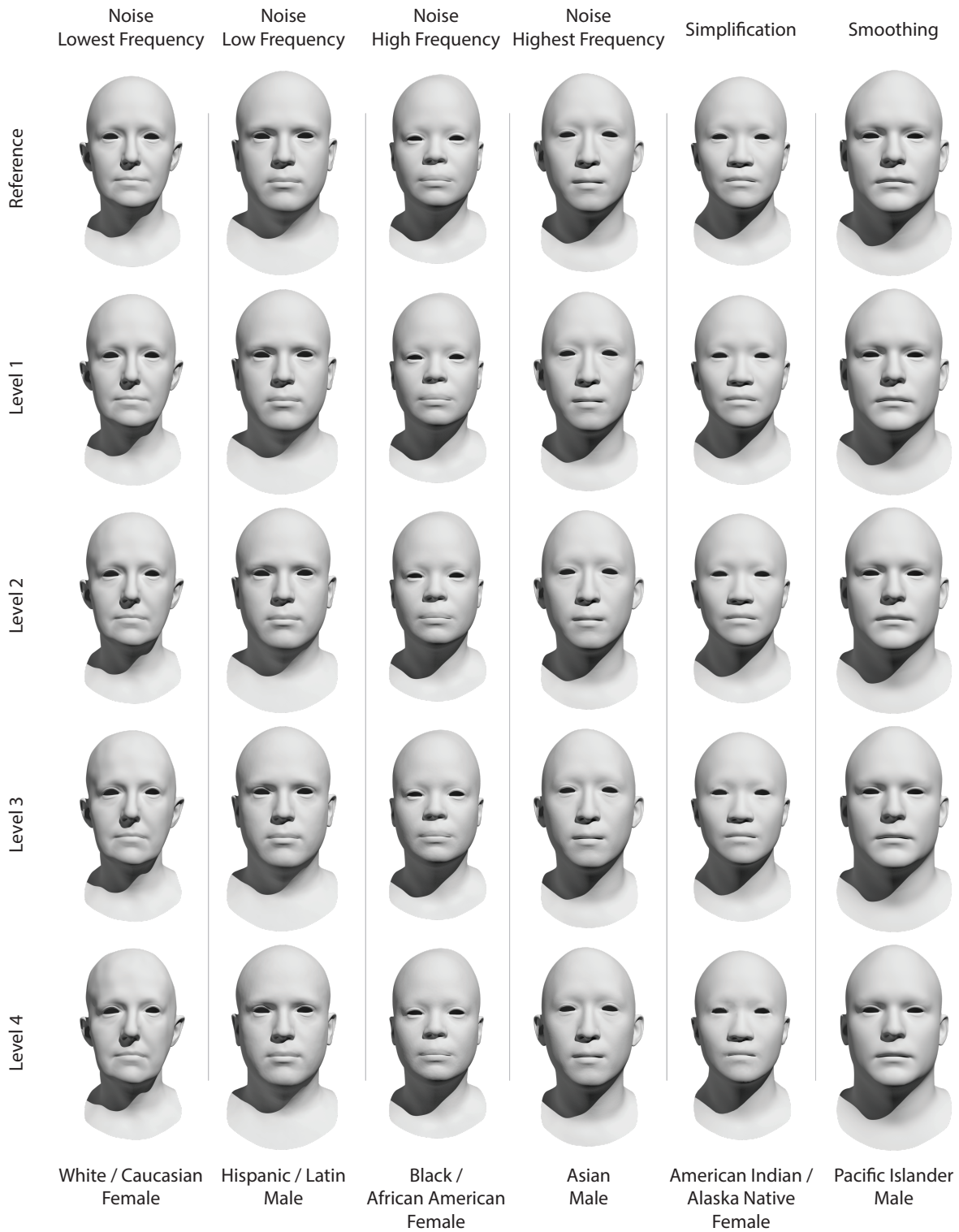
[7]Link to be added after acceptance

Fig. 6. Sample of distortions and demographic groups from our dataset. Each column represents a different distortion type and a different demographic group. Rows show sequentially increasing magnitude levels of distortion, starting from an undistorted reference (top), and ending with the highest magnitude of distortion (bottom). Please zoom in to fully appreciate the distortions. Note that in our study an individual face filled about two-thirds of the subjects' displays vertically.

Our experimental software was implemented using the Unity game engine[8] and rendered in real-time on the participants' machine using pre-processed meshes. Participants were allowed to freely rotate the models around the vertical axis, and were instructed to make a holistic decision based on the distortions on the entire mesh. All three meshes rotated in synchrony to enable direct comparisons. When no participant interaction was detected for a few seconds, the meshes began to slowly rotate, as pilot studies suggested this helps users detect distortions.

In addition to the artifacts described in Section 3.3, we decided to study two additional ancillary variables. The first is *distance*. Although distance is known to play a large role in the perception of visual artifacts [Daly 1992; Ye et al. 2019], it is not modeled by most existing geometry metrics. We study two distance conditions, "close" and "far" (subtending approx. 18 and 9 visual degrees respectively, assuming a standard distance of 2.5 picture heights away from the display as instructed in training). The second variable relates to human sensitivity to faces, in contrast to other objects. Prior art points to humans being especially sensitive to deformations on faces [Zhao et al. 2003], and evidence suggests a higher order holistic process takes place. In order to quantify the sensitivity differential between faces and general objects, we need to compare identical objects, but have one object be perceived as a face, while the other is not. This led us to the influential Thatcher Illusion [Thompson 1980], which demonstrated that the recognition of significant distortions of a face, such as rotating the eyes and mouth 180°, can go undetected when the face is presented upside-down. This effect has previously been successfully exploited in the computer graphics community to increase a subjects' sensitivity to colormap calibration [Kindlmann et al. 2002], where users presented with upside-down faces were found to be significantly less sensitive to distortion. We follow this template, and present our experimental subjects with two conditions: "normal" and "flipped", where the latter case shows the face upside-down. In this way, for each condition described in Section 3.3, we study four cases (close/up, close/flipped, far/up, and far/flipped).

4.3 Experimental procedure

We chose to employ a two-alternative forced choice (2AFC) procedure, as this is considered the most accurate way to measure threshold distortion visibility [Perez-Ortiz et al. 2019]. A sample trial of our study is shown in Fig. 7. The participant is presented with an unmodified reference face, selected from the synthetic dataset described in Section 3.2, and two variations subject to a the same artifact at different magnitudes (as described in Section 3.3). The subject is prompted to select the head that is less distorted compared to the reference by pressing on the left or right arrow keys.

To avoid obvious comparisons which do not add information to the set, only distortion magnitudes of up to 2 steps away from each other were compared [Perez-Ortiz et al. 2019], as shown in Table 2.

Before beginning each experimental session the subjects were trained. A representative subset of comparisons were shown containing all distortion types with exaggerated magnitudes, simplifying the task while avoiding over-training. During training, participants were given feedback on whether their answer was accurate to help

Table 2. Comparison Matrix

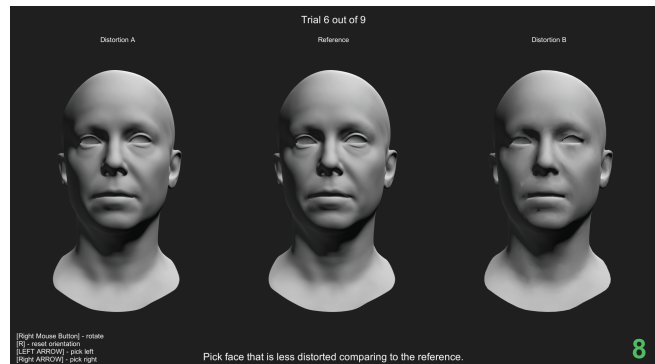|	Reference	Level 1	Level 2	Level 3	Level 4
Reference		✓	✓		
Level 1	✓		✓	✓	
Level 2	✓	✓		✓	✓
Level 3		✓	✓		✓
Level 4			✓	✓	

Fig. 7. This is a screenshot of the software used in our subjective study. The observer is presented with 3 heads at a time: a reference placed in the center, and two distorted heads randomly assigned to either side. The task is to pick the head that is less distorted in relation to the reference by pressing the left or right arrow keys.

accustom them to the experimental procedure. A separate set of meshes that was not repeated in the main study was used for training, and training results were discarded from further analysis.

Participants performed the experiment at home, and were instructed to sit comfortably at an approximate distance of 2 to 3 picture heights from their display. Breaks were allowed during the study, and it was possible to pause or exit the experiment and pick progress up at a later point if necessary. In order to avoid users getting stuck on difficult comparisons, we implemented a 30 second timer on the bottom right, and instructed participants to try not to exceed this time limit per trial, although this rule was not enforced. A minimum trial time of 2 seconds was enforced in order to avoid accidental responses. Once a session was completed, the anonymized results were automatically uploaded to our servers.

4.4 Data analysis

To allow for uniform comparisons across distortion types and magnitudes, we convert users' 2AFC responses using perceptual scaling (see Section 2.2). To do this, we employ the software developed by Perez-Ortiz et al. [2019] for each of the studied distortions. This results in a per-artifact JOD scale, which can be directly used for perceptually meaningful comparisons and predictions. This method also allows for automatic statistical outlier removal, which flagged 7 participants whose results were discarded from further analysis.
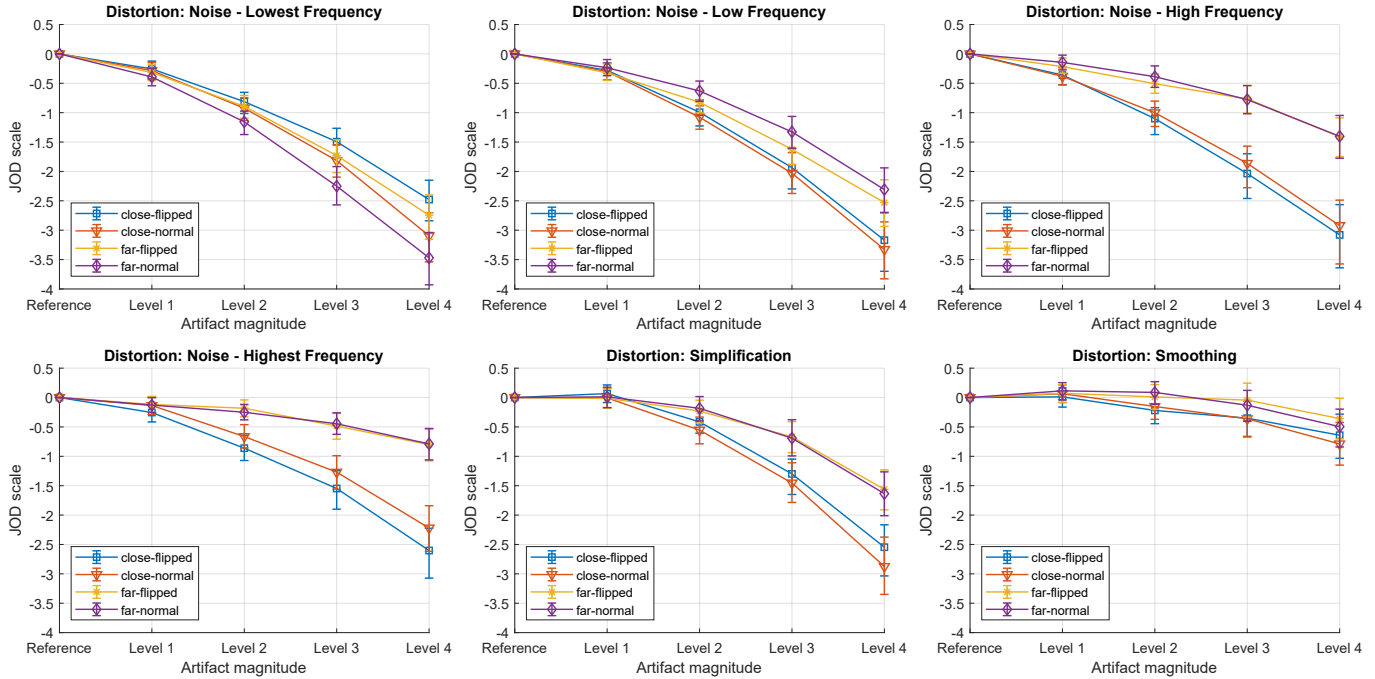
Fig. 8. This figure shows the results of our subjective study. Each plot represents one of six distortions we measured. Every line correspond to a separate condition. Note the strong impact of distance for the three highest noise frequencies and simplification. Please see our analysis in Section 5.

5 EXPERIMENT RESULTS

The results of our experiment are shown in Figure 8. A separate plot is presented for each of the six distortions studied (4 frequencies of noise, simplification and smoothing). On the horizontal axis, we show progressively larger magnitudes of distortion. On the vertical axis, we show a JOD-scale distance from the reference. The four lines represent the results of each combination of ancillary conditions (far/near, normal/flipped). Vertical bars show 95% confidence intervals, calculated using bootstrapping, as described by Perez-Ortiz et al. [2019]. Each plot shows the mean results across all participants. Below, we discuss the results for each distortion.

Noise. The results in the first four plots (top row and bottom left) correspond to the distortion caused by *noise*. First, we can focus on the top left plot, showing the lowest frequency. Note that, as expected, the JOD distance from the reference increases gradually as higher magnitudes of distortion are examined. In an effect only seen at the lowest frequency, users were more sensitive to the *far* conditions. We theorize this is due to the human visual system being less sensitive to stimuli at both very low and very high frequencies [Campbell and Robson 1968]. As the pattern moves away from the viewer, it subtends a smaller area resulting in an increase in perceived frequency as experienced by the observer, and consequently an increase in sensitivity for this distortion. This effect was previously exploited by Oliva and colleagues in their work on "Hybrid Images" [2006]. As the noise frequency is increased in the subsequent plots (top middle, top right, and bottom left), we see the near and far conditions reverse, making the far condition less visible

than near, with the separation becoming more apparent at higher frequencies. This follows the same insight, with sensitivity falling off for high frequencies which appear even higher to the observer due to increased distance, resulting in lowered sensitivity.

Simplification. The bottom middle plot shows the result for *simplification*. As expected, a significant difference between the close/far conditions is observed. This data follows the perceptual insight leveraged for popular level-of-detail rendering methods, which simplify meshes that lie further from the observer.

Smoothing. The result is shown in the bottom right plot. *Smoothing* was much less visible to our observers than the other distortions, with a maximum mean JOD distance of only -0.79. This is likely due to the localized nature of the distortion, which is only visible in areas of the mesh that contain high frequencies. In addition, our dataset of relatively smooth head meshes may also have led to more difficult comparisons.

Overview. We generated a perceptually meaningful scale of distortion for each of six studied artifacts. The effect of *distance* was strongly present in all conditions, either enhancing or obscuring the visibility of distortions following expected relations from contrast sensitivity models [Nader et al. 2015]. Conversely, the normal/flipped conditions did not impact responses. We hypothesize this is due to the unlimited length of presentation, which may have allowed participants to mentally cancel the effect of presenting a face upside-down [Lewis 2001]. We confirm these findings through a statistical analysis of the results using the effect size *d* [David 1963]

(see supplementary). As a consequence, in the following sections results are averaged over the normal/flipped conditions.

6 DISTORTION METRIC ANALYSIS

We begin by analyzing the performance of popular metrics on 3D geometry: RMS, Metro [Cignoni et al. 1998], DAME [Váša and Rus 2012], MSDM [Lavoué et al. 2006], MSDM2 [Lavoué 2011] and FMPD [Wang et al. 2012].

As artifacts differ significantly in their subjective characteristics, it is unlikely that any metric is able to successfully explain the entire dataset simultaneously. We confirm this by calculating Pearson's Linear Correlation Coefficient (PLCC) and Spearman's Rank Correlation Coefficient (SROCC) compared to the results of our subjective study (see supplementary).

In order to obtain more targeted performance indicators, we proceed to study the correlation of each metric to each of the studied distortions separately. DAME and MSDM are excluded from the *simplification* trials, as these metrics require vertex-to-vertex correspondence. We begin the analysis using only the results of the "close" condition, shown in Table 3. FMPD shows the best performance for *noise*, with top PLCC results in every case. Metro, MSDM, and RMS obtain the best SROCC for low, high, and highest frequency noise, respectively. Surprisingly, the simplest metric RMS performs best for *simplification* and *smoothing*, with the latter artifact having the worst overall scores across all metrics.

Table 4 shows the same analysis performed for the "close" and "far" conditions together. As none of the metrics model distance, which is a significant factor to the subjective judgements, performance drops happen for most comparisons highlighting the need for metrics that incorporate this perceptual factor. DAME now has the best PLCC performance for lowest-frequency noise, while FMPD maintains best performance for low, high and highest-frequency noise. RMS remains the best PLCC performing metric for *simplification* and *smoothing*.

7 APPLICATIONS

In this section, we show some practical applications that demonstrate the utility of our perceptual dataset.

7.1 Retargeting metrics

We build on the analysis of geometry metrics performed in Section 6 by retargeting these methods to fit our perceptual data. These metrics produce results in different, non-perceptual units. This makes it difficult to use their output to perform commonly encountered engineering tasks, such as setting meaningful thresholds or stopping criteria, as the output values have no perceptual reference. Typically, this type of parameter can be set through experiment, but relating the results of an experiment to the metric output can be challenging, and thresholds often do not generalize beyond the original results. This highlights the importance of perceptually scaled data, such as our dataset, to achieve practical tasks. We give a concrete example of this scenario for mesh compression in Section 7.2.

We also observe that distance, despite being a major perceptual factor to the visibility of artifacts, is not modeled by any of the examined metrics, which produce identical responses regardless of the users' position in relation to the object. We explore an extension of metrics to distance using our dataset for a LOD application in Section 7.3.

New perceptually-aware metrics for geometry distortion could be developed to address the points above, but generating such a metric is beyond the scope of this work. We simplify the problem by instead retargeting the responses of the metrics we examined in Section 6, allowing us to inject some perceptual insights. After experimenting with different functional forms (see supplementary material), we observed that an empirical remapping of the metrics' output to the perceptually-motivated JOD units by applying scaling in the form $Y = a \cdot X^b$ produces the best results of all examined models. This produced good fits of metrics' responses to our subjective data for a single distance, where X is a raw metric response, Y is metric retargeted to the JOD scale, and a and b are optimized fitting parameters.

We also explicitly incorporate distance as a factor to the metric responses. We do this by modulating the results using the ratio of visual angles subtended by the object. We set the size of the stimuli shown in the *close* condition of our study as the reference α, and the current size of the object as β. We then modify our retargeting formula to $Y = \frac{\beta}{\alpha} \cdot a \cdot X^b$. As is often practical in application, we can reformulate this to use distance relative to the observer. We define m_e as the maximum extent of the object (half the maximum size of the bounding box), d_1 as the distance corresponding to the default visual angle α, and d_2 as the distance for which we are querying (schematized in Figure 9). This results in:

$$Y = \frac{\arctan(m_e/d_2)}{\arctan(m_e/d_1)} \cdot a \cdot X^b \qquad (1)$$

The values of the parameters a and b are fitted to the dataset, minimizing error between the remapped metric response and JOD values found in our experiment for each stimulus. In addition, we add a boundary condition so that large distances ($\alpha < 0.11°$) have a score of 0 JOD as distortions are not visible at that scale. Based on the conclusions of Section 6, we perform this fitting separately for each metric and distortion type. The values of the fitted parameters a and b can be found in Table 3 of the supplemental material.

Through Eq. 1, we can use legacy metrics to make predictions that account for distance, and approximate metric output to the JOD units used in our study. In the following sections, we apply a retargeted version of FMPD to compression and LOD rendering.
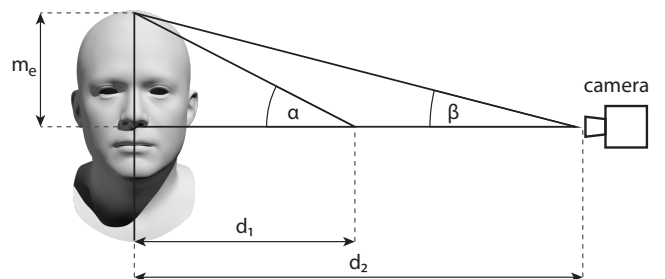

Fig. 9. The trigonometric relationship between visual angle, size and distance from the observer used in our model.

Table 3. Correlation Coefficients Computed for Each Distortion Type Separately (for Close Heads)

	Noise								Simplification		Smoothing	
	Lowest freq.		Low freq.		High freq.		Highest freq.					
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
METRO	-0.83786	-0.85531	-0.85159	-0.8886	-0.80548	-0.85085	-0.73139	-0.76049	-0.81164	-0.80853	-0.54	-0.50158
DAME	-0.83952	-0.86269	-0.83788	-0.87335	-0.78928	-0.81867	-0.72004	-0.74596	NA	NA	-0.54911	-0.53409
FMPD	-0.85387	-0.87492	-0.87582	-0.87283	-0.83075	-0.8595	-0.73603	-0.75207	-0.82596	-0.7967	-0.43741	-0.4882
MSDM	-0.81545	-0.86174	-0.75198	-0.87583	-0.71318	-0.83951	-0.61431	-0.77115	NA	NA	-0.50298	-0.52468
MSDM2	-0.7756	-0.86293	-0.8146	-0.88512	-0.76524	-0.84845	-0.64805	-0.75457	-0.66364	-0.80923	-0.55012	-0.52745
RMS	-0.8377	-0.85725	-0.85082	-0.88765	-0.80674	-0.87288	-0.72701	-0.73285	-0.82752	-0.79958	-0.55147	-0.54268

Table 4. Correlation Coefficients Computed for Each Distortion Type Separately (for Both Distances)

|	Noise								Simplification		Smoothing	
	Lowest freq.		Low freq.		High freq.		Highest freq.					
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
METRO	-0.84552	-0.87899	-0.80067	-0.83565	-0.64015	-0.65671	-0.56912	-0.58581	-0.71671	-0.70817	-0.47613	-0.44695
DAME	-0.8503	-0.88571	-0.78221	-0.80956	-0.63214	-0.6478	-0.55242	-0.55489	NA	NA	-0.47313	-0.46808
FMPD	-0.83415	-0.87444	-0.8242	-0.82314	-0.66238	-0.66313	-0.57095	-0.57861	-0.73287	-0.70548	-0.3836	-0.45222
MSDM	-0.81724	-0.86826	-0.70719	-0.818	-0.56928	-0.66113	-0.47542	-0.58426	NA	NA	-0.43856	-0.4754
MSDM2	-0.77799	-0.86893	-0.76497	-0.8251	-0.61003	-0.65798	-0.50469	-0.59126	-0.57988	-0.71234	-0.47943	-0.46324
RMS	-0.84133	-0.86611	-0.79922	-0.83484	-0.64533	-0.67854	-0.56143	-0.54794	-0.73604	-0.70532	-0.49102	-0.4832
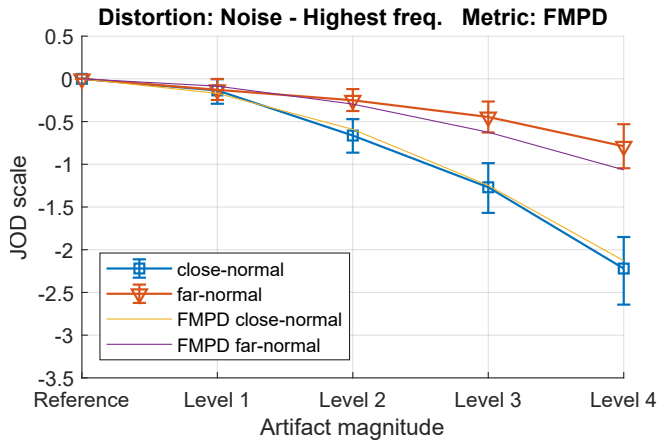
Fig. 10. We retarget FMPD as described in Section 7.1 to assess distortions caused by compression algorithms. Orange and blue lines represent the results of our study for the close and far conditions, while yellow and purple lines correspond to the remapped output of FMPD.

7.2 Perceptually bound compression

When employing mesh bitrate compression, a cost vs. quality trade-off is present—the goal is to compress at the lowest rate that presents acceptable subjective quality. To test the performance of the metric retargeting performed in Section 7.1, we ran a validation study.

Stimuli. Although our main study consisted only of human faces, we wanted to explore the hypothesis that our learnings can generalize by selecting a varied set of test cases. These consisted of shapes used by Lavoué et al. [2006] (venus, armadillo, rocker), as well as one face from our dataset. The former present challenges, as

these shapes are very different from those we studied. In particular, noisy references, such as the armadillo mesh, can result in reduced visibility of artifacts due to masking, which was not modeled in our work. For each object, we generate a progression of quantized meshes using the Draco library. Note that some low-compression-factor meshes were excluded during piloting, as they were deemed visually identical to the reference.

Procedure. We used a setup identical to that described in our main experiment (Section 4.3), with 8 additional participants. Participants' responses were converted to JOD units using the method of Perez-Ortiz et al. [2019]. We compared these values to the retargeted version of the FMPD metric, shown in Figure 10.

Results. The results of our validation can be seen in Figure 11. Note that, unlike traditional metrics, our retargeted metric's output can be directly compared to the results of the subjective study. The metric matches the perceptual outcome well—with identical selections for 3 out of 4 meshes at a threshold level of 1 JOD.

7.3 Perceptually driven automatic LOD

Level-of-detail rendering is a popular real-time graphics technique, swapping out meshes for lower complexity alternatives when objects are distant in order to improve performance. Despite its ubiquity, most practical implementations rely on distance swapping criteria set manually by artists. Traditional geometry metrics such as those examined in Section 6 are not distance-aware, preventing their use for this application. Conversely, a metric trained on our distance-aware data could estimate visibility based on distance, generating an automated quality-cost trade-off prediction. To exemplify this, we use a retargetted version of FMPD (as described in Section 7.1).
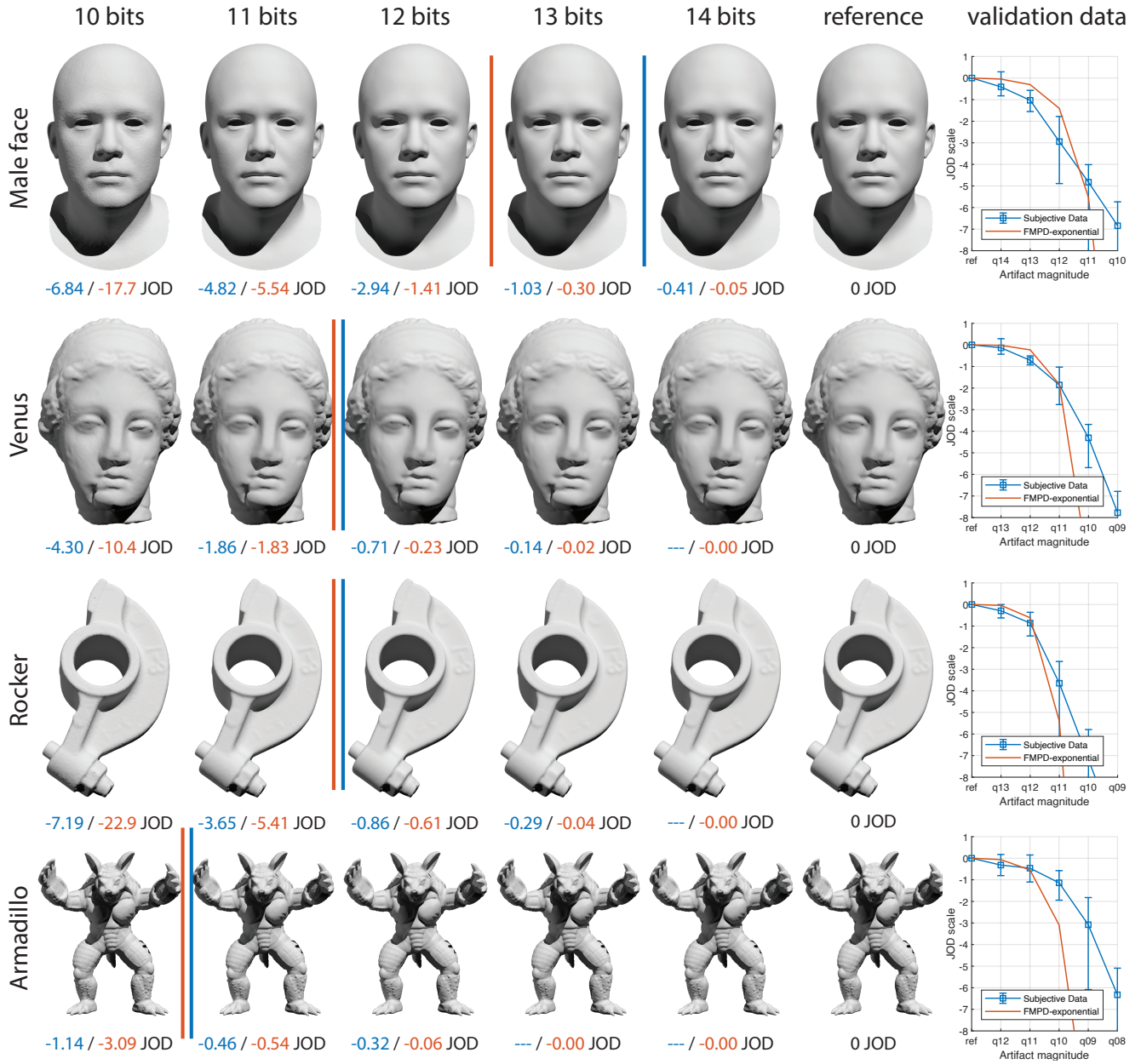
Fig. 11. A look at our compression application. Each column represents a mesh, encoded with a static number of bits per coordinate. The rightmost column shows the results of our validation experiment (Section 7.2) contrasted to the output of the retargeted FMPD metric, also shown under each mesh. Blue and orange separators denote the perceptual threshold, set to 1 JOD (corresponding to a 75% probability of preference), according to the study (blue) and metric (orange). Note that in contrast to our work, setting a meaningful threshold in arbitrary metric units is a non-trivial task. Please zoom in to observe the distortions.

For a given mesh L_0, we begin by pre-computing a set of simplified versions $L_{1..N}$. We then solve an inverse problem, finding the distance d_i at which the difference between the reference and each of the simplified meshes lies beyond a given perceptual threshold value δ at each step, i.e. $d_i = \mathrm{argmin}_d(|L_0 - L_i|_d \le i\delta)$.

We compare the prediction of our retargeted metric on a head from our dataset to a manual setting chosen by an artist in Unity, who was instructed to pick the earliest point that avoided visible artifacts. We show the results of our method in Figure 12. Note that our model is built using side-by-side comparisons, while the visibility of popping artifacts in an LOD setting is more similar to
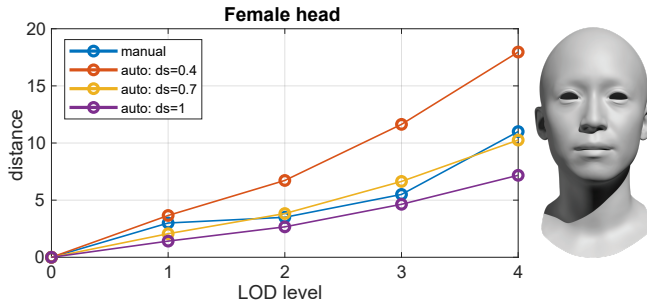
Female head

Fig. 12. Our retargeted metric can be used to automatically predict the distance at which an LOD swap should occur. Here, we compare our estimate to a selection made by an artist instructed to avoid visible popping artifacts (left) and present the mesh used for the evaluation (right).

a flicker comparison, which is expected to be significantly more sensitive [Simons and Rensink 2005]. As a result, using the same threshold of 1 JOD as Section 7.2 leads to swaps at lower distances than those selected by the artist. However, a lower threshold of 0.7 JOD produces a similar swapping profile. A perceptual data-driven model like the one presented here can be used by artists to automatically set an application-wide subjective threshold, reducing manual labor in selecting LOD settings.

8 CONCLUSIONS

In this work, we explore the perceptual impact of distortions on 3D geometry. We created a novel synthetic dataset, consisting of 100 head scans which mimic the demographic distribution of the US. We identified geometric distortions of interest to relevant applications, and ran carefully crafted pilot studies and numerical simulations to set the appropriate parameters for our perceptual study. We gathered a large-scale dataset of subjective responses for geometric distortion, including relevant perceptual parameters, such as distance. We used the results of our study to analyze the performance of geometric metrics. We also explored applications of our dataset to use cases like metric retargeting, automating stopping criteria for mesh compression, and LOD rendering.

Limitations and future work. Although we collected a large dataset of perceptual data on geometry, we were only able to sample a limited number of parameters and modalities for each distortion. In particular, it would be interesting to further explore the impact of distance on perception by sampling more in this dimension. Our dataset was composed of meshes representing human faces, which we measured in a normal and flipped orientation with the intent of comparing the perceptual impact of human sensitivity to faces. Although we found no difference in sensitivity between the two, it would be interesting to add a time-limited component to these trials to make it more difficult for users to mentally invert the rotation. Early piloting of our study with meshes representing arbitrary objects indicates that increased sensitivity for faces is present, and quantifying this numerically could generate valuable insights. Further, although masking is known to be a major effect for the visibility of perceptual artifacts, we did not explore it in this study. Notably, our study was done on meshes without texture, which is a likely source of masking for the distortions studied in this work. Future efforts should be made to collect psychophysical data on geometric distortions while actively controlling for masking due to properties of the reference mesh and texture.

REFERENCES

Andrew D. Bagdanov, Alberto Del Bimbo, and Iacopo Masi. 2011. The Florence 2D/3D Hybrid Face Dataset. In *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding.* 79–80. https://doi.org/10.1145/2072572.2072597

Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99).* ACM Press/Addison-Wesley Publishing Co., USA, 187–194. https://doi.org/10.1145/311535.311556

James Booth, Anastasios Roussos, Evangelos Ververas, Epameinondas Antonakos, Stylianos Ploumpis, Yannis Panagakis, and Stefanos Zafeiriou. 2018. 3D Reconstruction of "In-the-Wild" Faces in Images and Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 11 (2018), 2638–2652. https://doi.org/10.1109/TPAMI.2018.2832138

Mario Botsch, Mark Pauly, Leif Kobbelt, Pierre Alliez, Bruno Lévy, Stephan Bischoff, and Christian Rossl. 2007. Geometric Modeling Based on Polygonal Meshes. In *ACM SIGGRAPH Course Notes.* https://doi.org/10.1145/1281500.1281640

Fergus W. Campbell and John G. Robson. 1968. Application of Fourier Analysis to the Visibility of Gratings. *The Journal of physiology* 197, 3 (1968), 551–566. https://doi.org/10.1113/jphysiol.1968.sp008574

Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2014. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2014), 413–425. https://doi.org/10.1109/TVCG.2013.249

P. Cignoni, C. Rocchini, and R. Scopigno. 1998. Metro: Measuring Error on Simplified Surfaces. *Computer Graphics Forum* 17, 2 (1998), 167–174. https://doi.org/10.1111/1467-8659.00236

Darren Cosker, Eva Krumhuber, and Adrian Hilton. 2011. A FACS Valid 3D Dynamic Action Unit Database with Applications to 3D Dynamic Morphable Facial Modeling. *Proceedings of the IEEE International Conference on Computer Vision,* 2296–2303. https://doi.org/10.1109/ICCV.2011.6126510

D. Cudeiro, T. Bolkart, C. Laidlaw, A. Ranjan, and M. J. Black. 2019. Capture, Learning, and Synthesis of 3D Speaking Styles. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE Computer Society, Los Alamitos, CA, USA, 10093–10103. https://doi.org/10.1109/CVPR.2019.01034

Hang Dai, Nick Pears, William Smith, and Christian Duncan. 2019. Statistical Modeling of Craniofacial Shape and Texture. *International Journal of Computer Vision* 128, 2 (09 Nov 2019), 547–571. https://doi.org/10.1007/s11263-019-01260-7

Scott J. Daly. 1992. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity. In *Human Vision, Visual Processing, and Digital Display III,* Vol. 1666. International Society for Optics and Photonics, 2–15.

Herbert Aron David. 1963. *The Method of Paired Comparisons.* Vol. 12. London.

Michael Deering. 1995. Geometry Compression. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95).* Association for Computing Machinery, New York, NY, USA, 13–20. https://doi.org/10.1145/218380.218391

Xiang Feng, Wanggen Wan, Richard Yi Da Xu, Stuart Perry, Pengfei Li, and Song Zhu. 2018. A Novel Spatial Pooling Method for 3D Mesh Quality Assessment Based on Percentile Weighting Strategy. *Computers & Graphics* 74 (2018), 12–22. https://doi.org/10.1016/j.cag.2018.04.005

James A. Ferwerda, Peter Shirley, Sumanta N. Pattanaik, and Donald P. Greenberg. 1997. A Model of Visual Masking for Computer Graphics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97).* ACM Press/Addison-Wesley Publishing Co., USA, 143–152. https://doi.org/10.1145/258734.258818

Yudong Guo, Juyong Zhang, Jianfei Cai, Boyi Jiang, and Jianmin Zheng. 2018. CNN-Based Real-Time Dense Face Reconstruction with Inverse-Rendered Photo-Realistic Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (05 2018), 1–1. https://doi.org/10.1109/TPAMI.2018.2837742

Alain Horé and Djemel Ziou. 2010. Image Quality Metrics: PSNR vs. SSIM. In *2010 20th International Conference on Pattern Recognition*. 2366–2369. https://doi.org/10.1109/ICPR.2010.579

Gordon Kindlmann, Erik Reinhard, and Sarah Creem. 2002. Face-Based Luminance Matching for Perceptual Colormap Generation. In *IEEE Visualization, 2002*. 299–306. https://doi.org/10.1109/VISUAL.2002.1183788

Venkat Krishnamurthy and Marc Levoy. 1996. Fitting Smooth Surfaces to Dense Polygon Meshes. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 313–324. https://doi.org/10.1145/237170.237270

Guillaume Lavoué. 2011. A Multiscale Metric for 3D Mesh Visual Quality Assessment. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1427–1437. https://doi.org/10.1111/j.1467-8659.2011.02017.x

Guillaume Lavoué, Elisa Drelie Gelasca, Florent Dupont, Atilla Baskurt, and Touradj Ebrahimi. 2006. Perceptually Driven 3D Distance Metrics with Application to Watermarking. In *Applications of Digital Image Processing XXIX*, Vol. 6312. International Society for Optics and Photonics, 63120L. https://doi.org/10.1117/12.686964

Guillaume Lavoué, Mohamed Chaker Larabi, and Libor Váša. 2016. On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models. *IEEE Transactions on Visualization and Computer Graphics* 22, 8 (2016), 1987–1999. https://doi.org/10.1109/TVCG.2015.2480079

Michael B. Lewis. 2001. The Lady's Not for Turning: Rotation of the Thatcher Illusion. *Perception* 30, 6 (2001), 769–774. https://doi.org/10.1068/p3174

Charles Loop. 1987. Smooth Subdivision Surfaces Based on Triangles. *Master's thesis, University of Utah, Department of Mathematics* (1987). https://www.microsoft.com/en-us/research/publication/smooth-subdivision-surfaces-based-on-triangles/

David Luebke, Martin Reddy, Jonathan D. Cohen, Amitabh Varshney, Benjamin Watson, and Robert Huebner. 2003. *Level of Detail for 3D Graphics*. Morgan Kaufmann. https://doi.org/10.1016/B978-155860838-2/50002-9

Rafał Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Trans. Graph.* 30, 4, Article 40 (jul 2011), 14 pages. https://doi.org/10.1145/2010324.1964935

Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: A Visible Difference Predictor for Wide Field-of-View Video. *ACM Trans. Graph.* 40, 4, Article 49 (jul 2021), 19 pages. https://doi.org/10.1145/3450626.3459831

Georges Nader, Kai Wang, Franck Hétroy-Wheeler, and Florent Dupont. 2015. Just Noticeable Distortion Profile for Flat-Shaded 3D Mesh Surfaces. *IEEE transactions on visualization and computer graphics* 22, 11 (2015), 2423–2436. https://doi.org/10.1109/TVCG.2015.2507578

Anass Nouri, Christophe Charrier, and Olivier Lézoray. 2016. Full-Reference Saliency-Based 3D Mesh Quality Assessment Index. In *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1007–1011. https://doi.org/10.1109/ICIP.2016.7532509

Aude Oliva, Antonio Torralba, and Philippe G. Schyns. 2006. Hybrid Images. *ACM Trans. Graph.* 25, 3 (jul 2006), 527–532. https://doi.org/10.1145/1141911.1141919

James P. O'Shea, Martin S. Banks, and Maneesh Agrawala. 2008. The Assumed Light Direction for Perceiving Shape from Shading. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (Los Angeles, California) *(APGV '08)*. Association for Computing Machinery, 135–142. https://doi.org/10.1145/1394281.1394306

Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *IEEE international conference on advanced video and signal based surveillance*. 296–301. https://doi.org/10.1109/AVSS.2009.58

Maria Perez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, and Rafał K Mantiuk. 2019. From Pairwise Comparisons and Rating to a Unified Quality Scale. *IEEE Transactions on Image Processing* 29 (2019), 1139–1151. https://doi.org/10.1109/TIP.2019.2936103

Ken Perlin. 2002. Improving Noise. *ACM Trans. Graph.* 21, 3, 681–682. https://doi.org/10.1145/566654.566636

Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. 2015. Image Database TID2013: Peculiarities, Results and Perspectives. *Signal processing: Image communication* 30 (2015), 57–77. https://doi.org/10.1016/j.image.2014.10.009

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. 2018. Generating 3D Faces Using Convolutional Mesh Autoencoders. In *Proceedings of the European Conference on Computer Vision*. 704–720. http://arxiv.org/abs/1807.10267

Bernice E. Rogowitz and Holly E. Rushmeier. 2001. Are Image Quality Metrics Adequate to Evaluate the Quality of Geometric Objects?. In *Human Vision and Electronic Imaging VI (Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4299)*. 340–348. https://doi.org/10.1117/12.429504

Jeffrey A. Saunders and Benjamin T. Backus. 2006. Perception of Surface Slant From Oriented Textures. *Journal of Vision* 6, 9 (08 2006), 3–3. https://doi.org/10.1167/6.9.3

Peter H Schönemann. 1966. A Generalized Solution of the Orthogonal Procrustes Problem. *Psychometrika* 31, 1 (1966), 1–10. https://doi.org/10.1007/BF02289451

Ana Serrano, Bin Chen, Chao Wang, Michal Piovarči, Hans-Peter Seidel, Piotr Didyk, and Karol Myszkowski. 2021. The Effect of Shape and Illumination on Material Perception: Model and Applications. *ACM Trans. Graph.* 40, 4, Article 125 (jul 2021), 16 pages. https://doi.org/10.1145/3450626.3459813

Hamid R. Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik. 2005. LIVE Image Quality Assessment Database Release 2. (2005). http://live.ece.utexas.edu/research/quality

Daniel J Simons and Ronald A Rensink. 2005. Change Blindness: Past, Present, and Future. *Trends in Cognitive Sciences* 9, 1 (2005), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

Olga Sorkine, Daniel Cohen-Or, and Sivan Toledo. 2003. High-Pass Quantization for Mesh Encoding. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (Aachen, Germany) *(SGP '03)*. Eurographics Association, Goslar, DEU, 42–51. https://doi.org/10.2312/SGP/SGP03/042-051

Jennifer Sun and Pietro Perona. 1998. Where Is the Sun? *Nature neuroscience* 1, 3 (1998), 183–184. https://doi.org/10.1038/630

Peter Thompson. 1980. Margaret Thatcher: A New Illusion. *Perception* 9, 4 (1980), 483–484. https://doi.org/10.1068/p090483

United States Census Bureau. 2021. US census. https://www.census.gov/quickfacts/fact/table/US/PST045221 [Online; accessed 14-December-2021].

Libor Váša and Jan Rus. 2012. Dihedral Angle Mesh Error: A Fast Perception Correlated Distortion Measure for Fixed Connectivity Triangle Meshes. *Comput. Graph. Forum* 31, 5, 1715–1724. https://doi.org/10.1111/j.1467-8659.2012.03176.x

Kai Wang, Fakhri Torkhani, and Annick Montanvert. 2012. A Fast Roughness-Based Approach to the Assessment of 3D Mesh Visual Quality. *Computers & Graphics* 36, 7 (2012), 808–818. https://doi.org/10.1016/j.cag.2012.06.004

Andrew P. Witkin. 1987. Scale-Space Filtering. In *Readings in Computer Vision*. Elsevier, 329–332. https://doi.org/10.1016/B978-0-08-051581-6.50036-2

Nanyang Ye, Krzysztof Wolski, and Rafal K. Mantiuk. 2019. Predicting Visible Image Differences Under Varying Display Brightness and Viewing Distance. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5429–5437. https://doi.org/10.1109/CVPR.2019.00558

Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. 2008. A High-Resolution 3D Dynamic Facial Expression Database. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*. 1–6. https://doi.org/10.1109/AFGR.2008.4813324

Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato. 2006. A 3D Facial Expression Database For Facial Behavior Research. In *International Conference on Automatic Face and Gesture Recognition*. IEEE, 211–216. https://doi.org/10.1109/FGR.2006.6

Xing Zhang, Lijun Yin, Jeffrey F. Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, and Peng Liu. 2013. A High-Resolution Spontaneous 3D Dynamic Facial Expression Database. In *International Conference on Automatic Face and Gesture Recognition*. IEEE, 1–6. https://doi.org/10.1109/FG.2013.6553788

Wenyi Zhao, Rama Chellappa, P. Jonathon Phillips, and Azriel Rosenfeld. 2003. Face Recognition: A Literature Survey. *ACM computing surveys* 35, 4 (2003), 399–458. https://doi.org/10.1145/954339.954342

Hao Zhu, Haotian Yang, Longwei Guo, Yidi Zhang, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. 2021. FaceScape: 3D Facial Dataset and Benchmark for Single-View 3D Face Reconstruction. *arXiv preprint* (2021). https://doi.org/10.48550/arXiv.2111.01082