FaceMap: Distortion-Driven Perceptual Facial Saliency Maps

ZHONGSHI JIANG, Meta, USA
KISHORE VENKATESHAN, Meta, USA
GILJOO NAM, Meta, USA
MEIXU CHEN, Meta, USA
ROMAIN BACHY, Meta, USA
JEAN-CHARLES BAZIN, Meta, USA
ALEXANDRE CHAPIRO, Meta, USA

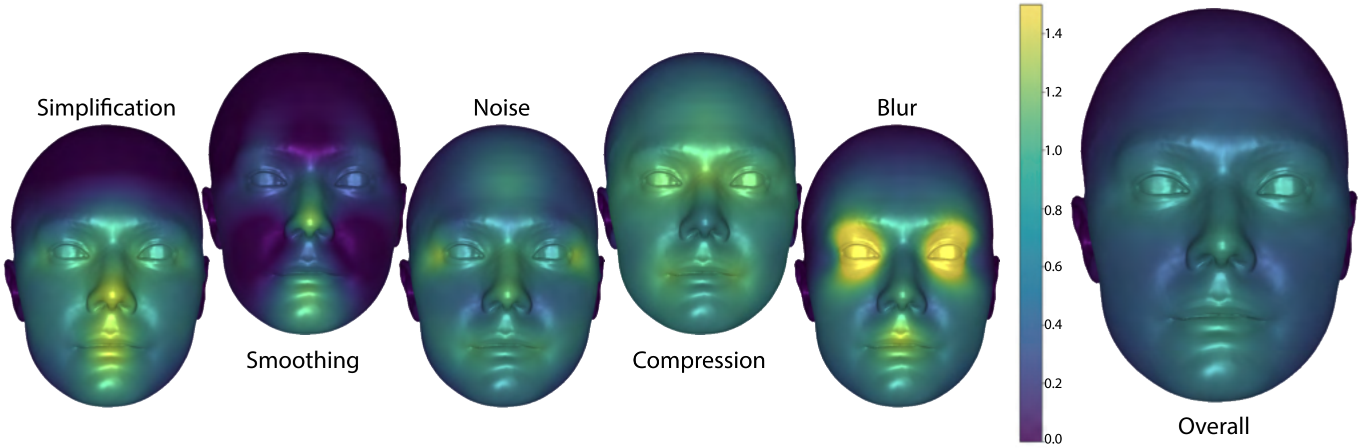


Fig. 1. A visual representation of the facial saliency maps produced in our work. On the left side, we show the subjective saliency of face regions for each of the studied distortion types, including geometry (simplification, smoothing, and noise) and texture (compression and blur). Note the nontrivial variations - such as the nose and mouth region being most important for simplification, yet, the eyes are most important for blur. On the right side, the face's aggregate general-use saliency map is shown across all distortions.

Humans are uniquely sensitive to faces. Recognizing fine detail in faces plays an important role in social cognition, identity; and it is key to human interaction. In this work, we present the first quantitative study of the relative importance of face regions to human observers. We created a dataset of 960 unique models featuring localized geometry and texture distortions relevant to visual computing applications. We then conducted an extensive subjective study examining the perceptual saliency of facial regions through the lens of distortion visibility. Our study comprises over 18,000 comparisons and indicates non-trivial preferences across distortion types and facial areas. Our results provide relevant insights for algorithm design, and we demonstrate our data's value in model compression applications.

CCS Concepts: • **Computing methodologies → Perception**; *Shape analysis.*

Authors' Contact Information: Zhongshi Jiang, Meta, USA; Kishore Venkateshan, Meta, USA; Giljoo Nam, Meta, USA; Meixu Chen, Meta, USA; Romain Bachy, Meta, USA; Jean-Charles Bazin, Meta, USA; Alexandre Chapiro, Meta, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SA Conference Papers '24, December 3–6, 2024, Tokyo, Japan
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1131-2/24/12
https://doi.org/10.1145/3680528.3687639

Additional Key Words and Phrases: Perception, Psychophysics, Facial Animation, Geometric Modeling

ACM Reference Format:
Zhongshi Jiang, Kishore Venkateshan, Giljoo Nam, Meixu Chen, Romain Bachy, Jean-Charles Bazin, and Alexandre Chapiro. 2024. FaceMap: Distortion-Driven Perceptual Facial Saliency Maps. In *SIGGRAPH Asia 2024 Conference Papers (SA Conference Papers '24), December 3–6, 2024, Tokyo, Japan.* ACM, New York, NY, USA, Article 39, 19 pages. https://doi.org/10.1145/3680528.3687639

1 INTRODUCTION

The human face holds immense importance in various domains, ranging from psychology and neuroscience to computer vision and graphics. Faces serve as a primary medium for social communication, conveying a wealth of information through expressions, gestures, and other facial cues. However, there has been an absence of quantitative studies that specifically examine which areas of a face are most important for different tasks and perceptual processes.

In perception literature, visual saliency (or salience) is the perceptual quality that makes some items in the world stand out from their neighbors and grab our attention. As this concept became integrated into computational frameworks, methods that may be inspired by biological principles, but geared empirically towards applications in computer vision and graphics have become popular [Perazzi et al.

2012]. Within this framework, we define a *facial saliency map* as one that indicates which parts of the face are perceptually important to an observer in a visual computing context. Intuitively we may expect some regions of the face to be especially prominent in this way. In applied scenarios this intuition is recognized in practice by artists and animators, who have historically given special attention to features like the eyes or mouth for critical tasks like facial blend shape based expression generation or sampling. While there is ample literature on both face perception and saliency, no quantitative data on visual sensitivity to face regions in this context is available.

We set out to build a subjective map of the face. In particular, our goal is to generate a numerical descriptor that can be used in computer graphics applications related to face rendering, learning, and geometry processing. As such, we focus on mapping human sensitivity by how noticeable distortions are when present on different areas of a face. We gathered a large-scale ($N = 72$) psychophysical dataset for localized facial distortions, consisting of over 18, 000 subjective evaluations in a carefully planned experimental procedure. Notably, as face-rendering technology is constantly evolving, our study is designed in a way as to make it independent of any one specific rendering technique, instead choosing to focus on commonly occurring artifacts. As our study parameters were unusually broad, spanning 3 base models, 5 different artifacts types at 2 magnitudes, and 32 local regions (for a total of 960 unique combinations), modern active sampling techniques were used to ensure optimal coverage and maximize information gain with each trial.

Finally, we demonstrated and subjectively validated use cases leveraging our facial saliency maps to face-related applications in compression for both geometry and texture, and Gaussian Splatting-based rendering [Kerbl et al. 2023]. Understanding localized importance maps of the face opens the way for future work studying expression, identity, and other perceptual priors for visual computing. Our face saliency maps, distortion dataset, and subjective experiment data are made available to the community[1].

2 RELATED WORK

2.1 Saliency

The concept of saliency was originally introduced in the cognitive neuroscience community and transferred over to a computational framework by Itti et al. [1998]. In this context, saliency algorithms attempt to model the mechanisms of visual attention in humans. This task is important, because knowing which area of a scene is likely to be attended by viewers can enable applications that optimally allocate rendering or scanning resources, and is particularly important in cases where limited means are available, like image and video compression, or on mobile devices with limited computation.

In image and video processing, saliency has become a major topic of research, with thousands of published works available. For more information, we refer the reader to the survey by Borji et al. [2019]. Typically, computational techniques are bottom-up and rely on low-level image features like edges and texture. While some methods use face-detection as a high-level importance-boosting signal [Cerf et al. 2007], they are not concerned with the relative importance of regions on the face, but rather treat the entire region containing it (as obtained by a bounding box, or similar techniques) uniformly.

Lee et al. [2005] extended the notion of image saliency to 3D geometry, using a multi-scale geometric curvature computation. Subsequent research extends the method to compute saliency, with spectral global geometry analysis [Song et al. 2014] or viewing region information [Leifman et al. 2016]. Applications include visualization, mesh simplification [Gal and Cohen-Or 2006; Shilane and Funkhouser 2007], and mesh watermarking [Lavoué et al. 2006]. It is important to point out that while these descriptors of shape can be relevant for computational geometry applications, they do not necessarily result in perceived importance. For example, in terms of curvature eyes are relatively flat, while ears are intricate, but our intuition points to eyes often being more salient than ears. Song et al. [2021] leveraged image saliency models to produce 3D mesh saliency. This method evaluates un-textured 3D meshes, which are not suitable for studies of the face.

Finally, an important subset of saliency exploration deals with attention as defined by explicit tracking of the users' gaze. Kim and colleagues [2010] validated the relationship between mesh saliency and human eye fixation and found a positive correlation. Lavoue et al. [2018] studies further to create a mesh fixation benchmarking dataset. Wang and colleagues [2018; 2016] used 3D printing and correlate gaze on the physical object to the digital 3D models. While gaze-based solutions can provide important information about a scene, it is often difficult to disambiguate the data from the experimental task, which usually consists of free-viewing experiences. In contrast, we chose an active detection task, i.e. participants are instructed to seek out visible distortions on faces, which allows for a better degree of confidence in the applicability of the results to improve the performance of visual computing applications.

2.2 Face Datasets and Perceptual Quality

Many 3D datasets of faces are present in the literature. Datasets focused on facial expressions range from ones using low-cost hardware like the Kinect [Cao et al. 2013] to high quality captures using the Di3D dynamic face capture system [Zhang et al. 2013]. Large datasets focusing on individual differences exist, such as Zhu et al. [2021] (938 participants) and Yin et at. [2006] (100 subjects). As our work focuses on regions of the face, it naturally results in a very large number of variables, which grows further when analysing multiple artifacts. To avoid an unfeasible study size, we chose to focus on a small number of base models collected using a high-resolution 3dMD scanner, as described in Section 3.

A number of datasets on perceptual quality are available in the literature, for both imaging and geometry. Imaging datasets are often used to calibrate perceptual quality metrics, as discussed by Mantiuk et al. [2021]. Similarly, in the geometry domain perceptual metrics like DAME [Váša and Rus 2012] and MSDM [Lavoué 2011; Lavoué et al. 2006] are calibrated on quality datasets. Nehmé et al. [2023] ran a large-scale study of distorted textured meshes, used to train a difference metric. Zerman et al. [2020] performed a subjective study where mesh-based and point-cloud based methods are compared for volumetric video of full body performances. However, none of these datasets or metrics deals specifically with faces.

[1] Source code and data at https://github.com/facebookresearch/FaceMap

Wolski et al. [2022] collected a dataset quantifying the visibility of geometric distortions on faces. As they focus exclusively on geometry, non-textured meshes were employed to avoid reducing artifact visibility due to masking. As our work is focused on both texture and geometry distortions, we are interested in studying the visibility of artifacts in realistic face models, and thus use fully textured meshes. In addition, our study focuses on localized distortions in order to derive a map of relative importance of face regions, while Geo-metric applies distortions evenly on the entire head model, making this kind of analysis impossible. McDonnell et al. [2021] examine the visibility of expressions on static textured faces by rating different activation levels of blendshapes via a Likert scale experiment. This work is similar to ours in that stimuli have local characteristics on the face (depending on the expression), but differs in that it does not examine the visibility of distortions stemming from artifacts. The researchers also explored race and sex differences, but found no strong evidence of an effect. We selected a diverse range of subjects in terms of gender and race for our studies, but leave deeper exploration of these factors to future work.

2.3 Subjective Studies

To explore the relative importance of regions of the face, we selected a side-by-side study design (termed 2-alternative-forced-choice, or 2AFC). This method was selected instead of direct ratings (such as a Likert scale or mean-opinion-scoring) as pairwise comparisons simplify the task in each trial to a binary choice, and have been shown to produce more accurate results [Zerman et al. 2018].

The output of the 2AFC study can be converted from pairwise comparison data into perceptually meaningful just-objectionable-difference units (JODs, a concept closely related to just-noticeable differences, or JNDs) using the method of Perez-Ortiz et al. [2017]. Scaling our results in JOD units presents several benefits, such as being able to extend our study in the future without relying on a user's internal Likert scale, which is strongly dependent on the dataset being used. In addition, JOD scores can be directly converted to probability preference in a 2AFC comparison (see supplementary Fig. 13, and the work of Mantiuk et al. [2021] for detailed analysis).

3 LOCAL FACIAL DISTORTION DATASET

We need a suitable dataset to study the relative importance of facial regions in terms of artifact visibility. We describe the process of creating our local distortion dataset in this section.

3.1 Base meshes

We use facial scans of human subjects as bases for our dataset. 3dMD Ltd's static 12-viewpoint 3dMDhead scanning system, a commercially available 3D scanner, was used to obtain the data. This high-end scanner[2] is composed of 36 machine vision cameras in 12 viewpoints, which is optimized for detailed captures of the head. Continuous 3D textured surface meshes are generated at varying levels of detail (approximately 45k vertices in our case). Three volunteering and informed subjects' scans were used in this study.

3.2 Distortion Locations

To study different regions on the face, we need a way to localize distortions. To do this, we defined points-of-interest in an empirically-guided fashion. These points were chosen based on an approximately even distribution across the face while also ensuring labels were placed on semantically significant regions like the eyes, mouth, nose, and other facial features. Six landmarks were selected along the center line of the face, and 13 more were placed symmetrically on each of the left and right sides. Because face outlines can be significantly affected by geometric distortions, points were also placed along the jawline. An illustration of all the landmark locations and semantic labels can be seen in Figure 9. A validation of these landmarks against randomized points is given in Supplementary B.

To avoid an excessive experiment size, we did not place samples in regions like the hair, neck, interior of the mouth, or back of the ears or head, restricting our study to regions of the face only. To further reduce the study duration, we randomly presented landmarks not on the center line on either the right or left side, which was not expected to affect results due to the symmetric nature of the face and the non-sided nature of the experiment design. This assumption was confirmed in data analysis during piloting, and the sidedness of the landmarks was shown to not be a significant factor during statistical analysis of our study data (see Section 4.7).

3.3 Distortion Types

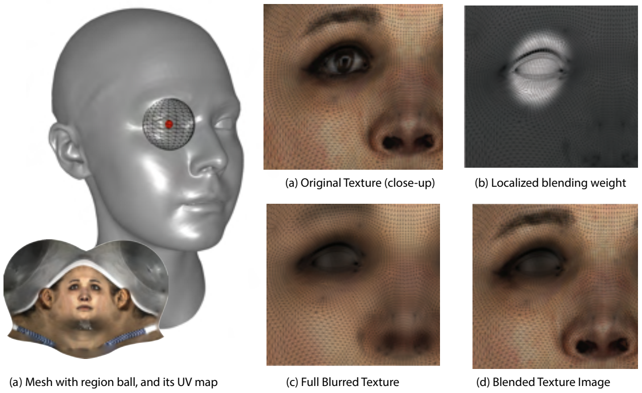
In order to study the visibility of distortions in localized face regions, we must select a set of distortion modalities to apply to our base meshes. Our dataset consists of both geometry and texture distortions. To help make our data more widely applicable, we focus on general distortion types commonly found in visual computing applications, as specific implementations may change over time. Each artifact was presented in two different magnitudes, with artifact strength defined during pilot experiments to subtend approximately 1 JOD per level in the aggregate, which was deemed optimal to avoid low-value conditions where artifacts are either obvious or invisible.

3.3.1 Texture distortions. Texture distortions were generated by first processing the whole input texture image T_i globally, obtaining a distorted texture T_d. Vertices within a spherical radius of 5% from a given landmark L were then marked as belonging to the local neighborhood of L (Figure 2). A localized radial basis function f as defined by Schaback et al. [2001] was then used to blend this area with its neighbors to avoid a noticeable hard edge as follows: $(1 - f) * T_i + f * T_d$. Two image-based distortions were studied:

Blur. A common distortion across visual computing applications is *blur.* It may be introduced in many situations - faulty scanning, naïve compression, or as a byproduct of training a neural network that is unable to reproduce fine detail of the texture. We generate blur by employing a Pyramid decomposition as implemented in OpenCV [2000]. The texture image is downsampled in 4 and 6 pyramid levels for artifact intensities 1 and 2, respectively, and then upsampled via bicubic interpolation back to its original size.

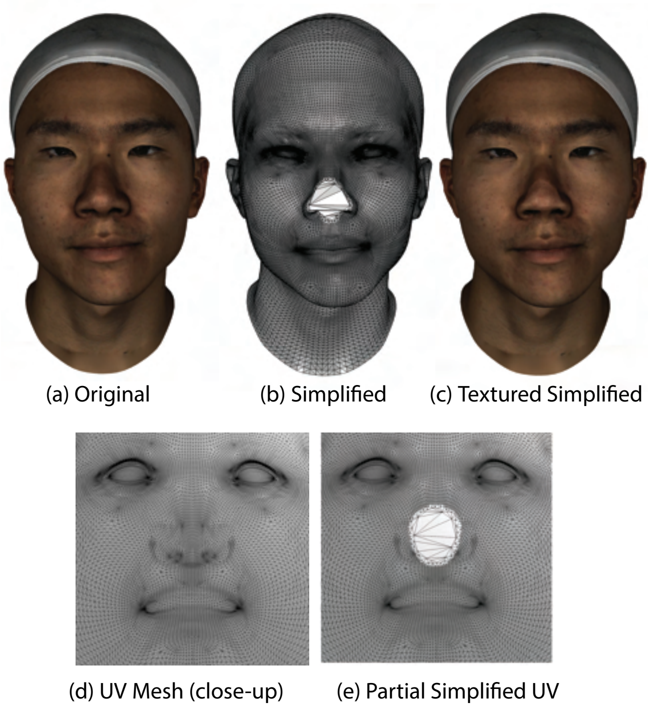
Compression. Texture compression is a popular method to reduce file size and aid storage and transmission. Unlike the other

[2] Linear accuracy range below 200 micrometers on a frame-by-frame basis claimed by the manufacturer, see https://3dmd.com/products/



(a) Mesh with region ball, and its UV map (a) Original Texture (close-up) (b) Localized blending weight (c) Full Blurred Texture (d) Blended Texture Image

Fig. 2. This figure illustrates the method used to generate localized distortions on texture for the *blur* and *compression* artifacts. A spherical neighborhood of a landmark is analyzed, and the original and distorted textures are blended via a radial function (see Section 3.3.1 for detail).



(a) Original (b) Simplified (c) Textured Simplified

(d) UV Mesh (close-up) (e) Partial Simplified UV

Fig. 3. Regional simplification for meshes. Inside a ball region of a given landmark (nose tip as illustrated), we remove the triangles inside and regenerate the triangulation with a smaller vertex budget. Then we sample the original UV map to reconstruct the shape and texture.

distortions studied in this work, it is difficult to abstract image compression from a specific algorithm. We use the popular JPEG format [1992] as implemented in OpenCV [2000], applied at compression ratios of 10% and 7% for artifact intensities 1 and 2.

3.3.2 Geometry distortions. We selected three geometry distortions, following the template set in the work of Wolski et al. [2022]. The *noise* distortion is applied locally in the same way as the texture distortions, while *smoothing* leverage a constrained optimizer, and *simplification* is using the method in Figure 3.

Noise. A common distortion that can be introduced in many parts of a visual computing pipeline is *noise*. We add Perlin noise [2002] along the normal directions for the mesh vertices. The frequency of the noise is 2 cycles/mm following the *highest* frequency version studied by Wolski et al. [2022], as low-frequency noise is not visible when applied locally. The amplitude is set to 10% and 15% for the intensity 1 and 2, respectively.

Smoothing. In applications, *smoothing* can be introduced by errors in simplification algorithms, mesh reconstruction or decimation, remeshing, and faulty scanning. In their work, Wolski and colleagues [2022] employed Laplacian smoothing [Witkin 1987], which is unsuitable for localized distortions [Jacobson et al. 2010]. Instead, we apply constrained biharmonic smoothing [Botsch and Kobbelt 2004]: outside a spherical region surrounding each landmark, vertices are fixed, and biharmonic weights are applied for the near-landmark region. We use this target as the high strength distortion (level 2) and then linearly blend (80%) with the input for the lower strength version (level 1).

Simplification. Finally, *simplification* is a technique commonly employed to reduce the polygon count of a geometric representation. This is useful for compression or when using distance-based adaptive techniques like level-of-detail rendering (LOD). To alter the mesh, simplification was carried out in the UV domain. For a sphere surrounding a given landmark, the UV mesh is locally retriangulated using the Triangle algorithm [Shewchuk 2005] with either 10% or 0% of the original internal vertices for artifact intensities 2 and 1, respectively.

3.4 Summary

In summary, our dataset consists of 3 base head scans. 2 texture and 3 geometry distortions are applied for a total of 5 unique artifacts, each of which can take 2 different strengths. The geometry noise distortion applied to each landmark is shown in Figure 9. All distortions applied to two representative landmarks for each unique face are shown in Figure 11, and the effect of different strengths on all the artifacts are shown in Figure 10. This results in a total of $3 \times 5 \times 2 \times 32 = 960$ unique meshes, not counting the reference.

4 USER DATA COLLECTION

4.1 Experiment participants

Before the main portion of the study, several rounds of piloting were employed to tune in experimental procedures, hardware setup, and meaningful distortion parameters (three pilot studies, N=10, 9, and 12). Finally, 72 naïve participants took part in the main portion of our study over the course of 6 weeks. All subjects were externally recruited by a specialized firm, forming a demographically balanced pool, signed informed consent forms, and were financially compensated for their work. Our data collection effort was approved by a third-party ethical review board.

4.2 Hardware Setup

The study was conducted in a dedicated experimental room. Stimuli were shown on a 31-inch professional reference monitor (Eizo CG3146). The display was set to a maximum luminance of 300 cd/m^2, with an sRGB EOTF, P3 color primaries, and a 60 Hz refresh rate. The display was calibrated daily via its built-in colorimeter to ensure stable performance. Participants were seated comfortably at approximately 3 picture heights from the display, and ambient light was dimmed to avoid reflections on the picture. Participant responses were recorded using a compact 3-button keyboard, with the leftmost and rightmost buttons used to record their subjective assessment.

4.3 Experimental Software

Our study was implemented in the Unity game engine[3], and rendered in real-time using distorted models that were generated offline, using the software packages libigl [Jacobson et al. 2013], OpenCV [Bradski 2000], and Triangle [Shewchuk 2005]. Subjects were permitted to rotate the models horizontally by up to 60° on each side, in which case both the test meshes and reference rotated in the same manner to enable direct comparisons. Models first appeared at the maximal rotation angle for a randomly selected orientation to avoid a left-right bias. If participants did not interact with the experimental suite, the meshes began to slowly rotate after a short delay, which was identified in piloting as being helpful to users.

4.4 Stimulus Rendering

In addition to the models themselves, lighting must be defined to generate a stimulus. As our goal is to allow for a clean experimental setup where subjects are able to clearly see all portions of the stimulus, we adopt the approach of the recent work by Wolski et al. [2022]. A distant top-right source, similar to natural outdoor conditions, is used as the main illuminant, and a secondary lower-intensity light is added from the bottom-left direction to avoid strong shadows. Unlike this work, however, our subjects are rendered with full color, as textures were found in testing to be a key component driving saliency and necessary to faithfully represent key facial regions like the eyes and mouth. An undistorted reference was rendered between the two test models to provide a better basis for comparison.

4.5 Experimental Procedure

To simplify the experimental task, we employed a 2-alternative-forced-choice procedure (2AFC), in which the subject's goal is to select which of two stimuli is less distorted (more similar) in relation to the reference. This type of experiment was found to achieve higher accuracy when measuring threshold visibility [Perez-Ortiz et al. 2019], which is closely related to our task. Notably, this also mitigates the effects of model-specific imperfections, as these would be present in both reference and test conditions. Prior to the study, participants received a comprehensive briefing explaining the setup and experimental task. Following a monitored training procedure, participants performed the main portion of the experiment, which consisted of 250 trials. Sessions lasted an average of 49.6 minutes, including a midpoint break and followed by a post-experiment qualitative survey. A still screen from the study is shown in Figure 4. A timer was added to the bottom right of the frame to help participants track their timing, but duration constraints were not enforced.

4.6 Sampling

Section 3.4 outlines the total number of unique meshes present in our distortion dataset. In order to obtain a single comparable scale, we need to compare different regions, distortion types, and magnitudes to each other. If tested naïvely, this would result in $\binom{960}{2} = 460320$ pairwise comparisons, which would be impossible to perform in a reasonable time frame. To avoid this problem, we employed an active sampling method, *ASAP* [Mikhailiuk et al. 2021], which uses expected information gain maximization to optimally schedule the next trial based on all previously collected data. Notably, it is sufficient for *ASAP* to run just 1 comparison per unique distortion per user. To further reduce the number of required trials, we only study 1 base mesh at a time and treat sided distortions as equivalent to each other, selecting right or left sides at random during runtime (as detailed in Section 3.2). This results in $1 \times 5 \times 2 \times (6+13) = 190$ unique distortions. In piloting, we found that the value of the reference is significant for an accurate interpretation of the scaled results, so we empirically added 5 additional instances of the undistorted reference to the sample set. An additional 55 ($\approx 30\%$) trials are added at the tail end of the experiment, for a total of 250 comparisons per participant, resulting in a manageable study duration. 72 participants performed the study (28 for head 1, 23 for head 2, and 21 for head 3).

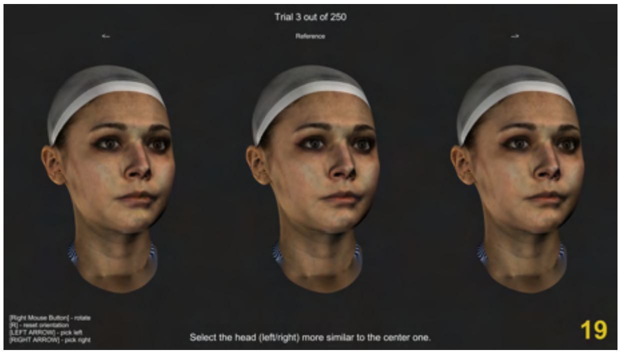
4.7 Data Processing

Data was converted to a perceptual JOD scale using the *pwcmp* library [Perez-Ortiz and Mantiuk 2017] (see Section 2.3). Outlier detection was performed as described by Perez-Ortiz and Mantiuk [2017], with a typical likelihood threshold of 1.5, resulting in the removal of 3 participants. Bootstrapping with 100 samples was used to calculate confidence intervals (Figure 6). Further, we performed N-way analysis of variance (ANOVA). As expected, significant variables included artifact type ($p \ll 0.01$), artifact strength ($p \ll 0.01$), and artifact location ($p \ll 0.01$). The side of the face on which a distortion appeared was found not to be significant ($p = 0.36$), validating our assumption that symmetrically mirrored distortions are analogous and are treated equivalently in the remainder of this work. The base model used was also not found to be significant ($p = 0.66$), which is an encouraging sign that the facial saliency map we built can generalize to other models. No significant between-factor interactions were found (see the supplementary material Section D for a full breakdown).

5 RESULTS

Perceptual scaling (as detailed in Section 4.7) was performed twice: once aggregating across all artifacts and again for each of the examined artifacts separately. The former is shown in Figure 5. Note that the reference is effectively one of the examined conditions, and we translate the dataset to ensure its value is set to 0 JODs by convention. Reinforcing our assumption that regions of the face have different perceptual weights, significant variations in the visibility of artifacts can be seen. Areas like the eyes, nose, and mouth show much higher values than those of the upper jaw, jaw joint,

[3] https://unity.com/



[Right Mouse Button] - rotate
[R] - reset orientation
[LEFT ARROW] - pick left
[RIGHT ARROW] - pick right

Select the head (left/right) more similar to the center one.

Fig. 4. Participants are asked to select the left or right mesh as being less distorted with respect to the reference, shown in the center.

and forehead. This is likely due to dual influences - both intrinsic perceptual importance, as well as practical visibility of artifacts in a given region. As expected, distortion strength increases the mean JOD output in all cases. A separate analysis per distortion is shown in Figure 6 and Figure 22 (in suppl.), and is discussed in detail below.

Geometric distortions. The top three rows of Figure 6 relate to geometry-based artifacts, namely *smoothing*, *noise*, and *simplification.* Note that these artifacts are generally stronger on the left side of the table, representing points in the center of the face. This is understandable, as the geometrically outstanding areas of the face that are also likely to be experientially important, like the nose and the lips, are located along the center axis. Curiously, the sensitivity on the outer eyes corner to geometric noise and simplification is high, likely due to changes in the face profile when rotated. Conversely, sensitivity in flatter regions like the cheeks and jaw is low.

Texture distortions. Figure 6 shows values for texture distortions in the bottom two rows. Notably, the sensitivity of the eye region to texture distortions is exceptionally high, suggesting both their unique perceptual importance and being strongly affected by texture artifacts due to the relatively detailed characteristics. Lips and eyebrows are also strongly affected, likely due to the presence of strong texture edges in these regions. Alternatively, the nose tip is less affected by texture distortions than geometric ones. The Jaw, forehead, and nose side regions are least affected by these artifacts, possibly due to the relatively low frequency of the textures present and low perceptual weight.

6 APPLICATIONS

In this section, we outline two example algorithms for efficient rendering of face models leveraging facial saliency maps. Given the perceptual scaling value defined on the sparse landmarks, we perform biharmonic interpolation [Stein et al. 2018] to produce a continuous scalar map defined on the whole surface (see Figure 1).

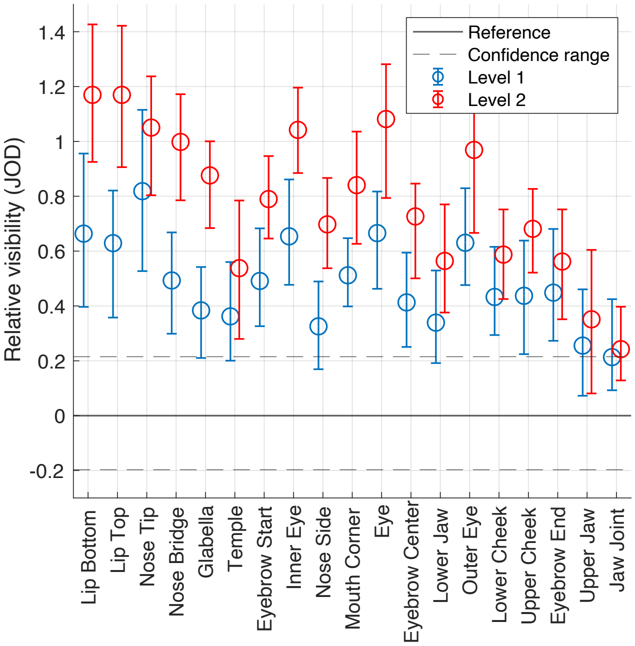


Fig. 5. This figure shows results for all studied locations on the face, aggregated across artifact types. Lower distortion intensity (level 1) is shown in blue, and higher (level 2) in red. Vertical lines represent 95% confidence intervals. Dashed horizontal lines represent the confidence interval of the reference, whose value is scaled to 0 by convention.

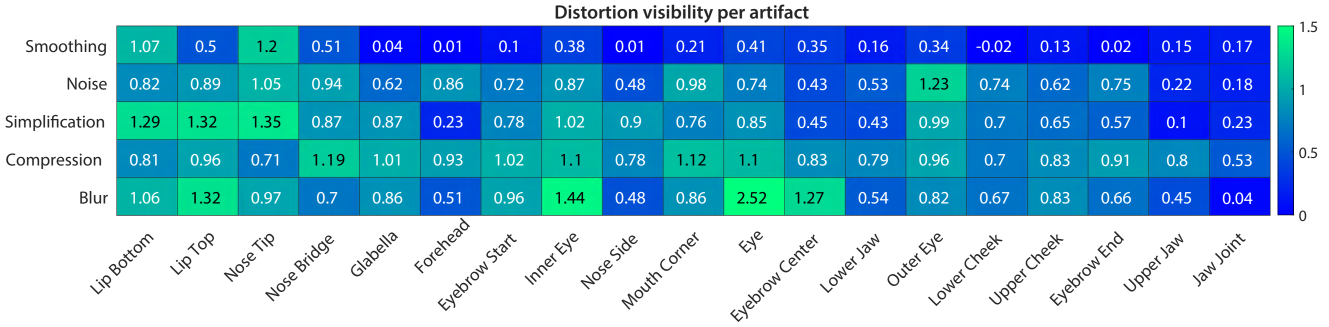
Note that applying FaceMap to meshes that are not a part of our dataset requires a topology transfer to a new template. We employed a semi-manual procedure to obtain the results in Section 6.1 and Section 6.2, assuming a target template mesh with shared connectivity (triangulation) and a fixed number of vertices. When different faces are semantically aligned, our saliency map can be defined per vertex. For a new template, (e.g. compare Figure 7 and Figure 18 for differences in face topology and UV), we use the Wrap4D software[4] to place the 32 landmarks used in the main study. This manual process took approximately 5 minutes, and was performed once per template. We then perform a closest point interpolation to transfer each of our distortion-driven saliency maps onto the new topology, and rasterize the value as an image in the texture space. Note that within each application different subjects typically share the same UV map, so the transferred map can be applied as-is.

6.1 Re-meshing

To demonstrate how to leverage collected user input in a traditionally geometry focused task, we perform a saliency-guided targeted re-meshing via the paradigm by Alliez et al. [2002]. Given a scalar field for the ideal edge length defined on the UV domain, we produce an adaptive 2D planar mesh with the Mmg Platform[5] [Dobrzynski and Frey 2008]. Using the UV map optimized with uniform low distortion as described by Rabinovich et al. [2017], new sample points are mapped to 3D and produce a mesh with a different topology.

To define the scalar field value corresponding to the desired edge length at each location, we make use of the *simplification* saliency map. We adjust a global error parameter until the total triangle count is at the desired value, then produce our final meshes (Fig. 7).

[4] https://docs.r3ds.com/Wrap/4DProcessingPipeline/4DProcessingPipeline.html
[5] https://www.mmgtools.org/



Distortion visibility per artifact

	Lip Bottom	Lip Top	Nose Tip	Nose Bridge	Glabella	Forehead	Eyebrow Start	Inner Eye	Nose Side	Mouth Corner	Eye	Eyebrow Center	Lower Jaw	Outer Eye	Lower Cheek	Upper Cheek	Eyebrow End	Upper Jaw	Jaw Joint
Smoothing	1.07	0.5	1.2	0.51	0.04	0.01	0.1	0.38	0.01	0.21	0.41	0.35	0.16	0.34	-0.02	0.13	0.02	0.15	0.17
Noise	0.82	0.89	1.05	0.94	0.62	0.86	0.72	0.87	0.48	0.98	0.74	0.43	0.53	1.23	0.74	0.62	0.75	0.22	0.18
Simplification	1.29	1.32	1.35	0.87	0.87	0.23	0.78	1.02	0.9	0.76	0.85	0.45	0.43	0.99	0.7	0.65	0.57	0.1	0.23
Compression	0.81	0.96	0.71	1.19	1.01	0.93	1.02	1.1	0.78	1.12	1.1	0.83	0.79	0.96	0.7	0.83	0.91	0.8	0.53
Blur	1.06	1.32	0.97	0.7	0.86	0.51	0.96	1.44	0.48	0.86	2.52	1.27	0.54	0.82	0.67	0.83	0.66	0.45	0.04

Fig. 6. This figure represents the JOD values recovered for our subjective study for all artifacts and locations on the face. By following the table horizontally, we can observe that each artifact's visibility can vary widely across locations - e.g. the left side of the table, containing landmarks on the center line, shows more sensitivity than points on the sides, with the notable exception of the eyes and mouth. If we observe the table along vertical lines, we can see that for each location on the face, the impact of two artifacts can be very different - for instance, the eye is most strongly affected by blur, while the nose tip, is most affected by geometry simplification and smoothing (as shown in Figure 11).

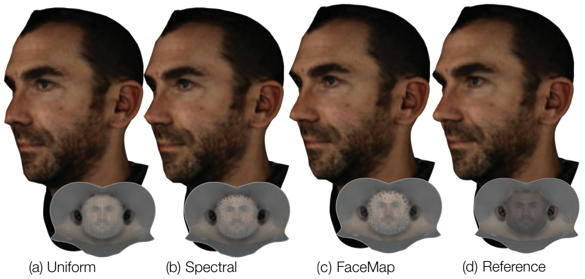


Fig. 7. A mesh simplification algorithm is guided by saliency maps to generate adaptive density meshes with 6% of the original vertices (Sec. 6.1). From left-to-right, meshes use the following strategy: a baseline with uniform sampling; an automatic saliency map generated by the method of Song et al. [2014]; FaceMap (ours); the full-density reference mesh. Note the differences in the nose and eye regions.

We conducted a validation study ($N = 9$ naïve participants) to test the effect of using FaceMap. 4 new models were obtained in the same way described in Sec. 3.1. They were compressed down to 30%, 20%, 10%, 8.3%, 6.6% and 5% of the initial 12 060 triangles (front face only) using the method described above, employing either a uniform grid, FaceMap, or an automatically generated saliency map using the method of Song et al. [2014]. After a training session, participants were tasked with rating each model on a scale of 1-10. N-way ANOVA showed FaceMap-based versions were significantly preferred to alternatives ($p \ll 0.01$, see Figure 8). Participant identity was also found to be a significant factor ($p \ll 0.01$), possibly due to different strategies chosen by users, but face model was not significant ($p = 0.44$). More details on this study and associated statistical analysis can be found in the supplementary sections (Section A.1 and Section D, respectively). Future work can aim at incorporating FaceMap priors into geometry-aware simplification methods [Liu et al. 2017], extending efficiency gains.

6.2 Gaussian Splatting

3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] optimizes a scene captured from multi-view images using a set of 3D anisotropic

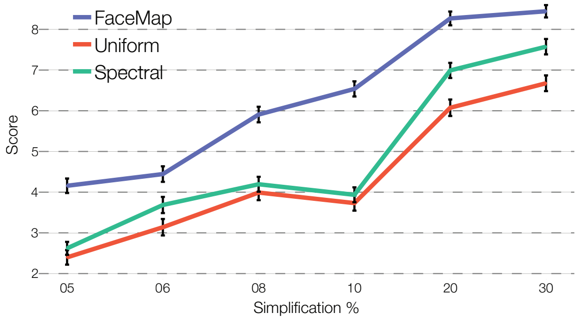


Fig. 8. This figure shows the results of our validation study for the saliency-guided mesh simplification application. The x-axis shows simplification as a percentage of the reference, and the y-axis shows the subjective scale employed in the study. Vertical bars show standard error. Note that FaceMap-based results are consistently preferred over the alternatives - for example, FaceMap meshes at 8% density obtained nearly the same score as uniform meshes at 20%.

Gaussians. This method presents a trade-off: increasing the number of 3D Gaussians improves quality, but requires more computational and memory resources [Lee et al. 2023]. Limiting the number of 3D Gaussians is crucial for e.g. compute constrained mobile devices.

We employ the multi-view face dataset [Wang et al. 2024], consisting of 160 views of each face, 3D reconstructed mesh, and uv-texture. We examine two 3DGS initialization schemes: uniformly sampled on the face as a baseline, and using a FaceMap-guided adaptive sampling. We then perform 3DGS optimization while maintaining a constant number of 3D Gaussians for 30,000 iterations. To validate this approach, we ran a user study with 11 additional naïve participants. The initialization schemes were compared in a 2AFC task at 5 different density levels. The results can be seen in Fig. 12. N-way ANOVA analysis showed FaceMap initialization was preferred significantly more often ($p \ll 0.01$). No significant effect of face model ($p = 0.25$) was present. Participant identity was again a significant factor ($p \ll 0.01$). No significant between-factor interactions were found. More details on this study and associated statistical analysis can be found in the supplementary sections (Section A.1 and

Section D, respectively). See supplementary Sec. A.2 for additional details.

7 LIMITATIONS AND FUTURE WORK

Identity. The sex or race of the model and viewer may affect the perception of facial saliency [McDonnell et al. 2021]. Although we did not find the face model to be a significant effect in our main or validations studies, future work could explore this possibility.

Expressions and Animation. Our studies were performed on faces with neutral expressions. It is possible that variations in a model's expression would produce an effect on the perceived importance of facial regions (such as a smile increasing the saliency of the mouth). Our study design can be leveraged to obtain directly comparable JOD-scale outputs for varying model expressions in future work.

3D Graphics. Our work examines the perception of geometry and texture artifacts on faces. The interplay between these factors is complex, and exploring it in detail is beyond the scope of this paper. However, we believe our data can be used in this type of effort, as we provide results on the visibility of distortions modifying one while keeping the second fixed. Analysis of our data could be beneficial for a complete model in the future.

Facial features. Our study did not investigate the effect of accessories or facials features, such as beards and glasses on FaceMap.

Template transfer. Using FaceMap on new templates may require a manual matching step as described in Section 6.

8 CONCLUSIONS

We performed the first distortion-driven study on the perceptual importance of regions of the human face. A number of distortions relevant to modern visual computing applications were studied, and regions were selected that sampled the face both spatially and semantically. Localized distortions were then applied, generating a dataset consisting of 960 unique models.

A psychophysical study ($N = 72$) was conducted, obtaining over 18, 000 subjective comparisons. This data was processed and converted to a unified subjective quality scale in JODs, and analyzed statistically, producing quantifiable insights into the relative importance of face regions for different distortion types. Finally, we explored applications of our facial saliency map for geometry remeshing and Gaussian Splatting. Subjective studies were used to validate that FaceMap improves performance when compared against a uniform baseline and automatic saliency estimators.

ACKNOWLEDGMENTS

We thank Alex Gherman, Carla Romero, Ali Yousefi, and John Hill for their help in organizing and running the subjective studies. We are grateful to Laura Trutoiu and Lina Chan for help identifying and sourcing the face scans used in this work, and the subjects for allowing us to use their likeness. We thank John Doublestein for the help on face templates. We thank Ladislav Kavan, Rakesh Ranjan, and Christian Haene for the valuable discussions.

REFERENCES

Pierre Alliez, Mark Meyer, and Mathieu Desbrun. 2002. Interactive geometry remeshing. *ACM Transactions on Graphics (TOG)* 21, 3 (2002), 347–354.

Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. 2019. Salient object detection: A survey. *Computational Visual Media* 5 (2019), 117–150.

Mario Botsch and Leif Kobbelt. 2004. An intuitive framework for real-time freeform modeling. *ACM Transactions on Graphics (TOG)* 23, 3 (2004), 630–634.

Gary Bradski. 2000. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25, 11 (2000), 120–123.

Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2013. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2013), 413–425.

Moran Cerf, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch. 2007. Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems* 20 (2007).

Cécile Dobrzynski and Pascal Frey. 2008. Anisotropic Delaunay mesh adaptation for unsteady simulations. In *Proceedings of the 17th International Meshing Roundtable*. Springer, 177–194.

Ran Gal and Daniel Cohen-Or. 2006. Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics (TOG)* 25, 1 (2006), 130–150.

Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 11 (1998), 1254–1259.

Alec Jacobson, Daniele Panozzo, C Schüller, Olga Diamanti, Qingnan Zhou, N Pietroni, et al. 2013. libigl: A simple C++ geometry processing library. *Google Scholar* (2013).

Alec Jacobson, Elif Tosun, Olga Sorkine, and Denis Zorin. 2010. Mixed finite elements for variational surface modeling. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 1565–1574.

Robert Jewsbury, Abhir Bhalerao, and Nasir M Rajpoot. 2021. A QuadTree Image Representation for Computational Pathology. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 648–656.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (2023).

Youngmin Kim, Amitabh Varshney, David W Jacobs, and François Guimbretiere. 2010. Mesh saliency and human eye fixations. *ACM Transactions on Applied Perception (TAP)* 7, 2 (2010), 1–13.

Guillaume Lavoué. 2011. A multiscale metric for 3D mesh visual quality assessment. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1427–1437.

Guillaume Lavoué, Frédéric Cordier, Hyewon Seo, and Mohamed-Chaker Larabi. 2018. Visual attention for rendered 3D shapes. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 191–203.

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.

Chang Ha Lee, X Hao, and A Varshney. 2005. Mesh Saliency. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 659–666.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. 2023. Compact 3D Gaussian Representation for Radiance Field. *arXiv preprint arXiv:2311.13681* (2023).

George Leifman, Elizabeth Shtrom, and Ayellet Tal. 2016. Surface regions of interest for viewpoint selection. *IEEE transactions on pattern analysis and machine intelligence* 38, 12 (2016), 2544–2556.

Songrun Liu, Zachary Ferguson, Alec Jacobson, and Yotam I Gingold. 2017. Seamless: seam erasure and seam-aware decoupling of shape from mesh resolution. *ACM Trans. Graph.* 36, 6 (2017), 216–1.

Pavan C Madhusudana, Xiangxu Yu, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. 2021. Subjective and objective quality assessment of high frame rate videos. *IEEE Access* 9 (2021), 108069–108082.

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 Transactions on Graphics (TOG)* 40, 4 (2021).

Rachel McDonnell, Katja Zibrek, Emma Carrigan, and Rozenn Dahyot. 2021. Model for predicting perception of facial action unit activation using virtual humans. *Computers & Graphics* 100 (2021), 81–92.

Aliaksei Mikhailiuk, Clifford Wilmot, Maria Perez-Ortiz, Dingcheng Yue, and Rafal Mantiuk. 2021. Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization. In *2020 IEEE International Conference on Pattern Recognition (ICPR)*.

Yana Nehmé, Johanna Delanoy, Florent Dupont, Jean-Philippe Farrugia, Patrick Le Callet, and Guillaume Lavoué. 2023. Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric. *ACM Transactions on Graphics* 42, 3 (2023), 1–20.

Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. 2012. Saliency filters: Contrast based filtering for salient region detection. In *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 733–740.

Maria Perez-Ortiz and Rafal K Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686* (2017).

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.

Ken Perlin. 2002. Improving noise. *ACM Transactions on Graphics (TOG)* 21, 3 (2002), 681–682.

Michael Rabinovich, Roi Poranne, Daniele Panozzo, and Olga Sorkine-Hornung. 2017. Scalable locally injective mappings. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1.

Robert Schaback and Holger Wendland. 2001. Characterization and construction of radial basis functions. *Multivariate approximation and applications* (2001), 1–24.

Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia* 126, 5 (2018), 1763–1768.

Jonathan Richard Shewchuk. 2005. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In *Applied Computational Geometry Towards Geometric Engineering: FCRC'96 Workshop, WACG'96 Philadelphia, PA, May 27–28, 1996 Selected Papers*. Springer, 203–222.

Philip Shilane and Thomas Funkhouser. 2007. Distinctive regions of 3D surfaces. *ACM Transactions on Graphics (TOG)* 26, 2 (2007), 7–es.

Eli Shusterman and Meir Feder. 1994. Image compression via improved quadtree decomposition algorithms. *IEEE Transactions on Image Processing* 3, 2 (1994), 207–215.

Ran Song, Yonghuai Liu, Ralph R Martin, and Paul L Rosin. 2014. Mesh saliency via spectral processing. *ACM Transactions On Graphics (TOG)* 33, 1 (2014), 1–17.

Ran Song, Wei Zhang, Yitian Zhao, Yonghuai Liu, and Paul L Rosin. 2021. Mesh saliency: An independent perceptual measure or a derivative of image saliency?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8853–8862.

Oded Stein, Eitan Grinspun, Max Wardetzky, and Alec Jacobson. 2018. Natural boundary conditions for smoothing in geometry processing. *ACM Transactions on Graphics (TOG)* 37, 2 (2018), 1–13.

Libor Váša and Jan Rus. 2012. Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1715–1724.

Gregory K Wallace. 1992. The JPEG still picture compression standard. *IEEE transactions on consumer electronics* 38, 1 (1992), xviii–xxxiv.

Xi Wang, Sebastian Koch, Kenneth Holmqvist, and Marc Alexa. 2018. Tracking the gaze on objects in 3D: How do people really look at the bunny? *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–18.

Xi Wang, David Lindlbauer, Christian Lessig, Marianne Maertens, and Marc Alexa. 2016. Measuring the visual salience of 3d printed objects. *IEEE computer graphics and applications* 36, 4 (2016), 46–55.

Ziyan Wang, Giljoo Nam, Aljaz Bozic, Chen Cao, Jason Saragih, Michael Zollhoefer, and Jessica Hodgins. 2024. A Local Appearance Model for Volumetric Capture of Diverse Hairstyle. In *International Conference on 3D Vision, 3DV 2024, Davos, Switzerland, March 18-21, 2024*. IEEE.

Andrew P Witkin. 1987. Scale-space filtering. In *Readings in Computer Vision*. Elsevier, 329–332.

Krzysztof Wolski, Laura Trutoiu, Zhao Dong, Zhengyang Shen, Kevin Mackenzie, and Alexandre Chapiro. 2022. Geo-Metric: A Perceptual Dataset of Distortions on Faces. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–13.

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.

Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, Rafal Mantiuk, and Frédéric Dufaux. 2018. The relation between MOS and pairwise comparisons and the importance of cross-content comparisons. In *Human Vision and Electronic Imaging Conference, IS&T International Symposium on Electronic Imaging (EI 2018)*.

Emin Zerman, Cagri Ozcinar, Pan Gao, and Aljosa Smolic. 2020. Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

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.

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 arXiv:2111.01082* (2021).

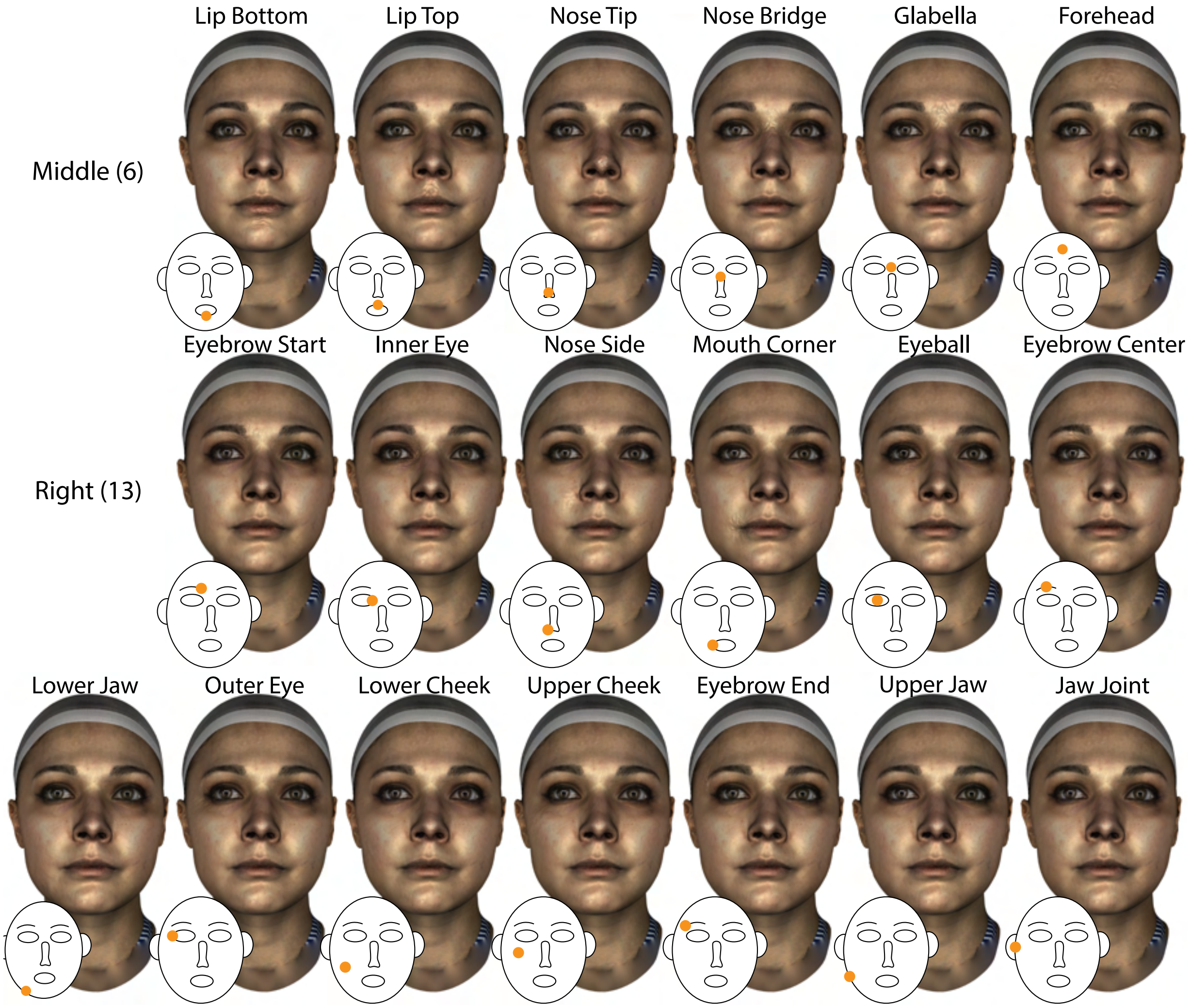


Fig. 9. Models distorted with geometry noise at each landmark location in our dataset (symmetrical regions on the left side not shown). Landmarks were chosen to obtain good coverage of the face, as well as to cover semantically important locations. To view, we recommend zooming in until each face approximately covers the height of the screen, as presented during our study. Note that distortions on the side of the face may not be visible in this front-facing render.

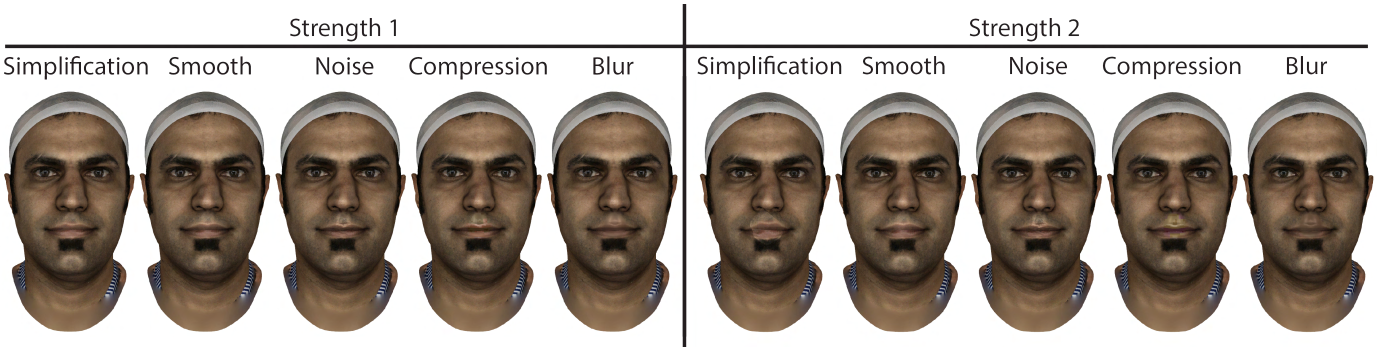


Fig. 10. This figure shows all distortion types at a single location (upper lip) at both levels of strength. To view, we recommend zooming in until each face approximately covers the height of the screen (as presented during our study).

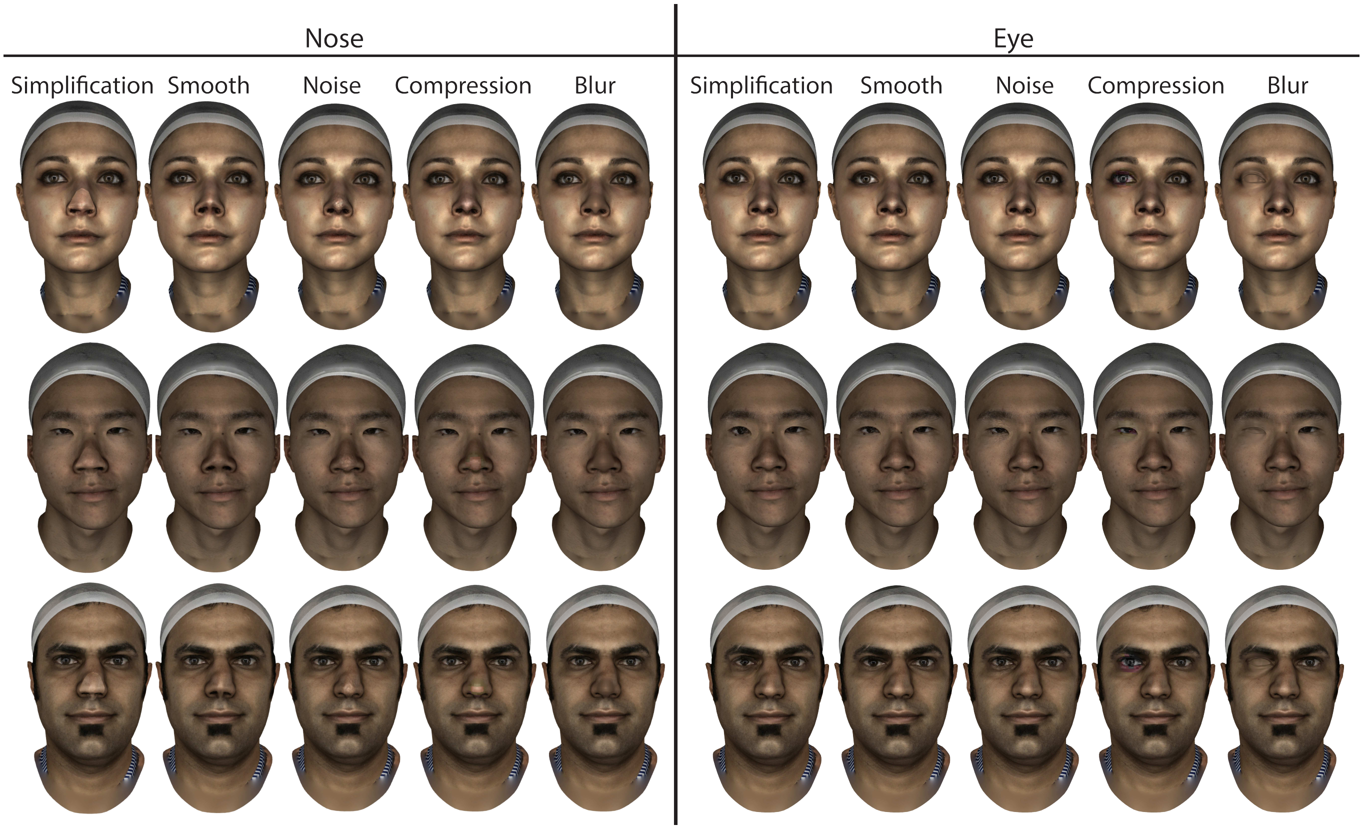


Fig. 11. This figure illustrates all distortions (geometry: simplification, smoothing, noise; and texture: compression, blur) applied to two locations (nose tip and right eye) for each of the base heads in the dataset. Note that as a geometrically salient feature, the nose is especially affected by simplification and smoothing. Conversely, the eye is especially affected by blur and compression.

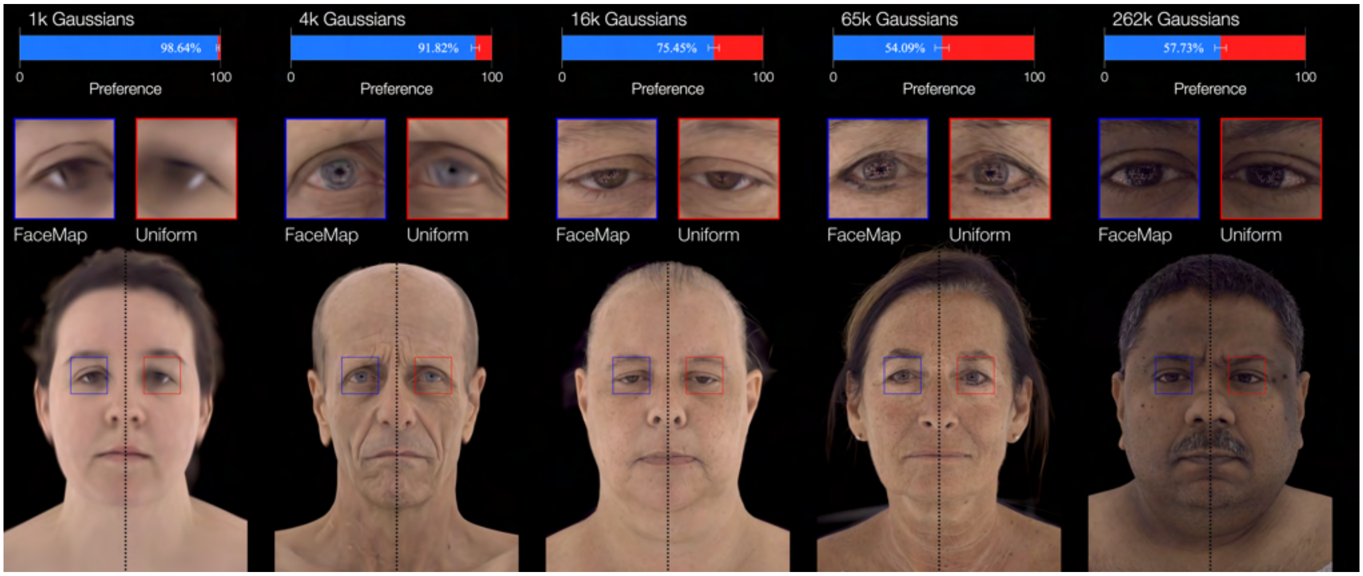


Fig. 12. Faces are rendered using Gaussian Splatting, as described in Sec. 6.2. From left to right, models are presented with an increasing total number of Gaussians. For each model, left and right sides show FaceMap and uniform initialization, respectively. Bars above each model show the overall preference in our validation study over all examined faces between the two initializations, and insets show detail in the eye region. Error bars show standard error.