EUROGRAPHICS 2018 / D. Gutierrez and A. Sheffer
(Guest Editors)

Volume 37 (2018), Number 2

Visual Attention for Rendered 3D Shapes

Guillaume Lavoué[1], Frédéric Cordier[2], Hyewon Seo[3], and Mohamed-Chaker Larabi[4]

[1]CNRS, Univ. Lyon, LIRIS, France
[2]University of Haute-Alsace, LMIA, France
[3]CNRS, University of Strasbourg, ICube, France
[4]CNRS, Univ. Poitiers, XLIM, UMR 7252, Poitiers, France

Figure 1: *We conducted eye tracking experiments on rendered 3D objects. The human saliency information is mapped on the 3D meshes (in the form of fixation density maps) and serves to study which factors influence human attention and to evaluate state-of-the-art saliency algorithms. At the bottom we show the Pearson correlation (ρ) between saliency maps from humans and algorithms.*

Abstract

Understanding the attentional behavior of the human visual system when visualizing a rendered 3D shape is of great importance for many computer graphics applications. Eye tracking remains the only solution to explore this complex cognitive mechanism. Unfortunately, despite the large number of studies dedicated to images and videos, only a few eye tracking experiments have been conducted using 3D shapes. Thus, potential factors that may influence the human gaze in the specific setting of 3D rendering, are still to be understood.

In this work, we conduct two eye-tracking experiments involving 3D shapes, with both static and time-varying camera positions. We propose a method for mapping eye fixations (i.e., where humans gaze) onto the 3D shapes with the aim to produce a benchmark of 3D meshes with fixation density maps, which is publicly available. First, the collected data is used to study the influence of shape, camera position, material and illumination on visual attention. We find that material and lighting have a significant influence on attention, as well as the camera path in the case of dynamic scenes. Then, we compare the performance of four representative state-of-the-art mesh saliency models in predicting ground-truth fixations using two different metrics. We show that, even combined with a center-bias model, the performance of 3D saliency algorithms remains poor at predicting human fixations. To explain their weaknesses, we provide a qualitative analysis of the main factors that attract human attention. We finally provide a comparison of human-eye fixations and Schelling points and show that their correlation is weak.

CCS Concepts

•***Computing methodologies*** → ***Interest point and salient region detections; Perception; Mesh models;***

1. Introduction

Visual attention is an aspect of human perception that has been widely explored by the vision science community. Indeed, understanding where exactly human observers look in images is fundamental for many computer vision and computer graphics applications (e.g. foveated compression, indexing, cropping, selective rendering, game level design). Eye-tracking is the main way of studying and understanding this property. One of the seminal studies was performed by Yarbus in the 60s [Yar67] in which he showed Ilya Repin's painting to several observers and assigned to them different viewing tasks. Yarbus noted that the observation of stationary objects such as images translates into a sequence of saccades and fixations on interest (i.e. *salient*) points of the observed image. Since then, many computational models of saliency have been proposed to automatically predict the

salient regions of an image [IKN98, LLBT06, JEFT09]. These image saliency models have been validated by using fixation maps obtained from eye-tracking experiments [BTSI13]. Some authors have proposed mesh saliency models operating directly on 3D data [LVJ05, SF07, LST12, SLMR14, TCL*15]. These approaches rely on geometry information to predict the salient regions on the surface of a shape. However, while these approaches may be efficient, none of them have been evaluated with respect to a ground-truth of fixation maps obtained from an eye-tracking experiment.

Very recently Wang et al. [WLL*16] conducted an eye-tracking experiment to verify if the findings of Yarbus (i.e. observers shift their gaze to salient features), based on the observations of flat 2D stimuli, are still valid for 3D shapes. They asked human observers to inspect physical objects (printed 3D shapes) and mapped their fixations on the surface of these shapes. They found that, like for 2D images, there are visually salient features on 3D shapes that attract the observer's visual attention. They also used the fixation locations obtained to benchmark simple mesh saliency estimators (curvature and difference-of-Gaussian based regional saliency [LVJ05]) and showed that they fail to predict fixations. This recent work by Wang et al. is the first attempt to study visual attention on 3D shapes and to produce a dataset of 3D meshes with fixation locations.

In this work, we conduct new eye-tracking experiments involving 3D shapes, with both static and time-varying camera positions. Contrary to Wang et al. [WLL*16], we are interested in *rendered* 3D shapes since, except when printed, 3D assets are mostly viewed on a screen. After mapping the raw 2D fixations onto the 3D shapes, we use the fixation maps obtained to study the influence of shape, camera position, material and illumination on the 3D fixations. We also create a large dataset of 32 shapes with fixation maps, used afterwards to benchmark state-of-the-art saliency predictors and analyze factors attracting human attention.

Contributions. The contribution of this paper is four-fold: First, we introduce two benchmark datasets of 3D meshes with mapped human-eye fixations, consisting respectively of 54 images and 81 videos from 3 shapes, and 96 images from 32 shapes. Second, we provide a rigorous statistical analysis of the influence of 3D shape and rendering parameters on the 3D eye fixation locations. Third, we perform a quantitative comparison of four saliency models from the literature [LVJ05, LST12, SLMR14, TKD15], with an analysis of their successes and failures. Finally, we provide an analysis of factors influencing human attention and a comparison of fixation maps with Schelling points [CSPF12]. This dataset is available at `http://liris.cnrs.fr/glavoue/data/saliency/`.

2. Previous Work

As stated in the introduction, visual attention has been widely explored by the scientific community. In Computer Vision, many experiments have been conducted and many saliency models have been proposed (e.g. [JEFT09, LL15]). This section details existing work in Computer Graphics, with a focus on mesh saliency models and eye-tracking experiments.

2.1. Saliency models for 3D meshes

Early work on saliency detection for 3D objects considered 2D algorithms applied on rendered images. For instance Yee et al. [YPG01] used the model from Itti et al. [IKN98] to evaluate the saliency of a dynamic 3D scene. More recently researchers have proposed saliency models directly based on the 3D data (mostly geometry information). In a pioneering work, Lee et al. [LVJ05], inspired by the 2D algorithm from Itti et al. [IKN98], consider a difference-of-Gaussian operator based on the mean curvature map. This operator is applied at multiple scales, and single-scale results are then aggregated to obtain the final saliency. Shilane and Funkhouser [SF07] propose detection of *distinctive* regions of a shape by examining how useful they are for distinguishing this shape from others of different classes. Their algorithm thus requires a database of meshes partitioned into classes. Leifman et al. [LST12] consider the distinctiveness of each vertex (how it is different from the others) as well as extremities of protrusions and then apply a spatial regularization to this per-vertex information to obtain the final saliency. Song et al. [SLMR13, SLMR14] use spectral approaches which detect the irregularities in the spectrum (*i.e.* eigen-values of the Laplace operator). Finally several recent algorithms [WSZL13, TKD15, TCL*15] first apply an over-segmentation and then compute the saliency per patch (instead of per vertex). Wu et al. [WSZL13] exploit the global rarity of the patches, while [TCL*15] and [LTC*16] estimate their saliency based on their relevance to some of the most unsalient ones. Finally, Tasse et al. [TKD15] first compute saliency values per patch by considering their uniqueness and distribution and then smoothly propagate them to the vertices. Much literature is available on the subject of 3D mesh saliency. Readers can refer to [LLS*16] for a very recent and comprehensive survey.

Algorithms presented above concern *shape saliency*, i.e. saliency based on geometry information only, which is also our focus of interest. However, higher level saliency models have also been proposed for complex 3D scenes, mostly in the context of video games [BSW10, KDCM14, KDCM16]; these are based on the semantic context of objects, rather than on their pure geometry.

2.2. Eye-tracking experiments and benchmarks for 3D meshes

In the field of computer vision, saliency models are usually evaluated using ground-truth datasets generated from eye tracking experiments, in the form of either fixation locations, or fixation maps [BTSI13, BJO*16]. However, in Computer Graphics, and particularly for mesh saliency, there are very few ground-truths available. Several eye-tracking experiments have been conducted: Howlett et al. [HHO05] carried out such an experiment in order to integrate human fixation information in mesh simplification algorithms. They consider two sets of respectively 37 and 30 models, with rather low resolutions (between 5K and 8K faces). The fixations were simply aggregated per face. Kim et al. [KVJG10] conducted an experiment involving 5 high resolution objects. Their goal was to compare the saliency model from Lee et al. [LVJ05] with curvature for the prediction of fixation locations. This evaluation was, however, carried out in the 2D image space. Ground-truth datasets from these two studies [HHO05, KVJG10] are not publicly available. Mantiuk et al. [MBM13] conducted an experiment with animated 3D scenes; they show that raw eye-tracker fixations are inaccurate for tracking small moving objects and propose an improved tracking algorithm. Finally, as raised in the introduction, Wang et al. [WLL*16] recently conducted an eye tracking experi-

ment involving 15 *real* printed shapes and gathered a ground-truth of 3D meshes with mapped fixations. Our work is closely related to this. However, we consider *rendered* 3D shapes, making it possible to investigate a whole range of experimental conditions (regarding material, lighting, and camera movement) and their influence on human-eye fixations.

It is important to note that two benchmarks currently serve as ground-truth for the evaluation of mesh saliency algorithms [CSPF12, DCG12]. However, they do not result from eye-tracking experiments. Chen et al. [CSPF12], conducted a large crowdsourcing experiment where they showed rendered 3D models and asked people to "select points on the surface of a 3D object likely to be selected by other people". Similarly, Dutagaci et al. [DCG12] asked participants to "mark all the points they think are interesting or defining". While these data are interesting, they rather reflect *interest points* than *human fixations*. Hence we believe that they are not well suited to evaluating saliency algorithms. Moreover, by asking an observer to select points, his/her attention becomes task-driven (top-down), which is proven to be different from the stimulus-driven approach based on low-level features (bottom-up). One of the goals of our work is to propose an alternative benchmark, *truly* related to human fixations. Our benchmark will also make it possible to test whether these two different concepts (interest points and human fixations) are correlated.

3. Overview of the Eye-Tracking Experiments

We conduct two eye tracking experiments with several different purposes. The goal of the first one (see Section 4) is to evaluate how the 3D shape, its material, the lighting conditions and the camera movement influence human-eye fixations. The second experiment (see Section 5) aims at providing the scientific community with a benchmark for the objective evaluation of mesh saliency models. These two experiments share the rendering parameters, protocol and tools, which are detailed below.

3.1. Creation of stimuli

Given a 3D model, we created 2D visual stimuli by producing images and videos with HD resolution (1920×1080 pixels) using the 3D Studio Max software [Aut17]. These images and videos have all been rendered using the Phong shader under perspective projection. Each object was placed at the center of the scene with its local coordinate system aligned with that of the world. In case of dynamic stimuli (object rotating around its center), the videos were rendered with a camera whose view direction points toward the center of the scene. The used 3D shapes, materials, light orientations and camera positions depend on the experiments and will thus be described in sections 4.1 and 5.2.

3.2. Apparatus

We used the Tobii TX-120 standalone eye-tracking device. This device allows both eyes to be tracked simultaneously and reports an accuracy equal to 0.5° under ideal conditions. While recording eye-tracking, data are delivered every 8 ms (120 Hz). A higher sampling is unnecessary in our case since we do not need to analyze the results in a finer detail. Every recorded point is characterized by its screen coordinates (x, y) and can be classified as a fixation or a saccade. The *I-VT* fixation filter developed by the device manufacturer was used to handle this task. Stimuli were displayed on the Eizo ColorEdge CG303W 30" monitor with a refresh rate of 60 Hz. The display was viewed in a calibrated test room with controlled lighting at 64 lux on the surface of the display. The distance between observer-eyetracker-display was defined in such a way as to guarantee an accurate recording while also ensuring comfortable viewing for the observer. The TOBII Track Status tool was used to place the participant at the appropriate distance from the eye tracker. The actual distance between the observer and the display was approximately 90cm. This distance, as well as the head position, was checked at every stage of the experiment. Observers were instructed to move eyes instead of their head and to keep their head in approximately the same position as it was in when calibrating the eye tracker.

Details about the calibration and accuracy check. To ensure the quality of the recorded gaze data, a calibration step has been performed for each observer and before each session. This step consists of presenting a neutral image with 3×3 calibration points, successively appearing. The software of the eye-tracker computes both accuracy and precision and displays the results graphically. The observer can run the test only if his/her calibration is within a tolerable interval. In addition to this calibration, we also checked the accuracy and precision of the eye-tracker using the Tobii Accuracy Test Tool, regularly along the experiment. This tool uses a regular 3×3 point image and computes the mean offsets in millimeters and degrees of visual angle.

3.3. Procedure

Every observer participated individually in the eye-tracking experiment alone in the test room. The experiment started with the calibration step. The whole series of images/videos (depending on the experiment) was then presented without a break except for the mid gray plate (see Figure 2), which was intentionally inserted to reduce the memory effect between stimuli. The stimuli were presented randomly. Extra stimuli were added at the beginning of the test to allow the observer to adapt to the test and to focus appropriately. The results of these first stimuli were discarded at the time of the analysis. The experiments were conducted in a free-watching setting, and no specific task was given to observers other than to observe the content of the stimuli. As demonstrated by Yarbus [Yar67], giving observers a task would make the results usable only in this specific context. The procedure is depicted in Figure 2.

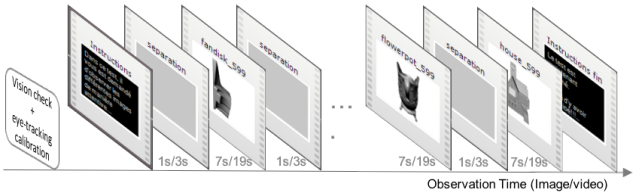

Figure 2: *Illustration of the procedure followed for our eye-tracking experiments. The duration shown is given in the format: image / video.*

3.4. Data processing

The parameters of the I-VT fixation filter applied are tuned in order to detect even short fixations while watching the stimuli. If several fixations occur within a spatial interval of less than 0.5° of visual angle, they are merged into a single fixation. In the filter parameters, we consider that the minimum duration of fixation is 60 ms. Eye-movements with a velocity higher than $30^\circ/s$ are assumed to be saccades. Such a tuning provides the opportunity to detect pursuit movements that are considered as a sequence of fixations. Analysis of eye-tracking records starts with the selection of only valid data, when both eyes are open.

3.5. From 2D Fixations to the 3D Fixation Density Map

The data from the eye-tracker is a sequence of fixations, where each fixation is defined by its spatial 2D position (x,y) (in the screen space) and its duration, *i.e.* the number of temporal samples (every 8 ms) composing it. In order to determine which surface point on the 3D mesh corresponds to the fixation pixel, we compute the ray emitted by the camera pinhole and passing through the pixel on the image (point p_i in Figure 3). We compute the intersection of this ray with the mesh model in the scene at the time of the fixation. The closest point of intersection is taken as the 3D fixation point (point p_m in Figure 3).

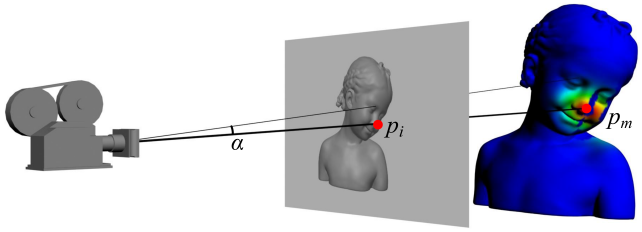

Figure 3: *Computation of the fixation density distribution on the 3D model, corresponding to a 2D fixation point p_i. α corresponds to 1° of visual angle.*

Sometimes, the fixation point is located near the contour of the mesh. In this case, depending on the precision of the eye-tracker, the projection ray may miss the intersection with the mesh, emitting away to the background. Since it clearly shows the observer's interest in the silhouette part of the object, we wish to take these fixations into account. To this end, we replace the ray by a cone and we project a Gaussian distribution on the 3D mesh (see Figure 3). The standard deviation of the Gaussian distribution is set to 49 pixels, corresponding to 1° of visual angle in our experimental conditions *i.e.* the radius of the fovea of the human visual system. By summing the contributions of all fixation points from all observers, we obtain the fixation density map (referred to as the fixation map hereafter in the paper) as illustrated in Figure 1, 2nd column.

3.6. Measuring similarity between maps

To analyze the agreement between observers (see Section 4) and to benchmark the saliency models (see Section 5), we need to be able to compute a similarity value between two fixation maps (derived from humans or saliency models). Many evaluation metrics have been proposed to compare fixation maps (or sets of fixation locations) [LB13], e.g. receiver-operator-characteristic (ROC) curve, Information Gain, Pearson correlation and so on. Bylinskii et al. [BJO*16] recently analyzed a set of 8 different metrics and compared their properties. According to their recommendations, we selected the Pearson's Linear Correlation Coefficient (ρ), since it provides a balanced handling of false positives and false negatives. For two maps x and y, it is defined as follows:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (1)$$

For benchmarking the saliency models, we also selected the area under the ROC curve (AUC). In this case, the fixation maps are thresholded to be converted into binary fixation maps (in practice we threshold to obtain 20% of visible vertices considered as fixations; more thresholds are illustrated in the supplementary material). The saliency map is then treated as a binary classifier of these fixations. The ROC curve represents the relationship between probability of false positives and probability of true positives and is obtained by varying the decision threshold on the saliency map. The area under the ROC curve (AUC) can then be used as a direct indicator of performance (1 corresponds to a perfect classification while 0.5 corresponds to a random one).

4. Experiment 1: How Shape and Rendering Parameters Influence Human Gaze

The goal of this first experiment is to evaluate the impact of the 3D shape and rendering parameters (camera movement, lighting and material) on human-eye fixations. For this purpose, we test whether different human observers tend to generate similar fixations for similar or different instances of shape, camera movement, lighting and material. We describe below a protocol for this task (see Section 4.3), inspired by [WLL*16]. The protocol is described for assessing the impact of the 3D shape ; we follow exactly the same protocol to assess the impact of camera movement, material and lighting.

4.1. Stimuli

For this first experiment, we selected three 3D objects of high resolution: *Igea* (101K vertices), *Dinosaur* (42K vertices) and *Blade* (200K vertices). They belong to very different semantic categories (a human face, an organic creature and a mechanical part) and have very different shapes. The idea behind having such a very small, yet representative, set of models is to be able to apply a large variety of rendering conditions (as described below) while keeping the eye-tracking experiment tractable.

These objects were rendered using three materials, obtained by varying the specular and glossiness coefficients ($\in [0,100]$) of the Phong shader from 3D Studio Max. Matte, mid gloss and glossy materials were rendered using different specular and glossiness coefficients set to $(0,10)$, $(20,30)$ and $(45,45)$ respectively.

Three lighting conditions were considered: (1) one spot light attached to the camera (i.e., front lighting), (2) one spot light placed above and slightly to the left of the camera (which is the assumed light direction by human vision for shape perception [SP98]) and

(3) three spot lights placed around the object (their location was chosen so as to minimize the amount of shadows). Examples of stimuli with different lighting/material conditions are provided in the supplementary material.
The 3D objects were rendered as static scenes under two camera positions, chosen manually as the most representative, and also as dynamic scenes, using three camera paths: an horizontal rotation around the object, a vertical rotation, and a random path covering the object. We thus obtained a total of $3 \times 3 \times 3 \times 2 = 54$ rendered images and $3 \times 3 \times 3 \times 3 = 81$ rendered videos. Video duration was set to 20 seconds (one complete rotation) while still images were displayed 7 seconds each. We split this first experiment into two sessions, one with the static setting and the other with the dynamic setting.

4.2. Participants

16 observers participated to the static session, and 13 to the dynamic session. Aged between 20 and 40, they were naive about the goals of the experiments. All observers had a normal or corrected to normal vision, verified by FrACT (Freiburg Visual Acuity Test). Participants were also screened using the Ishihara compatible color vision test for color blindness. The total time was 16 minutes for the static session (vision check + training/explanations + 54 images $\times$ 7 sec. + transitions), and 45 minutes for the dynamic session (split into two sub-sessions).
Two observers were removed from the experiment due to the presence of too much invalid data (e.g., eyes not looking at the screen).

4.3. Analysis protocol

To assess whether observers agree across shapes, we proceed in a similar way to Wang et al. [WLL*16]: the idea is to compute the similarity $S(Q_n^i, Q_m^i)$ (noted as $S_{n,m}^{i,i}$) between fixation maps Q_n^i and Q_m^i of two observers n and m for the same shape i (and the same camera movement, lighting and material) and then compare it with the similarity *across shapes* $S(Q_n^i, Q_m^{j \neq i})$ (noted as $S_{n,m}^{i,j}$). If $S_{n,m}^{i,i} > S_{n,m}^{i,j}$ for most pairs of observers then a certain agreement may be observed, i.e. different human observers tend to generate more similar fixations for the same shape than for different ones. More specifically, for each possible pair of shapes (i, j) (3 pairs in total), we compare $S_{n,m}^{i,i}$ and $S_{n,m}^{i,j}$ for each combination of camera movement, lighting and material (e.g., 18 in total for the static scenes) and each possible pair of 14 observers ($\binom{2}{14}$=91). We thus obtain, for each pair of shapes (i, j), two sets of 18×91=1638 similarity values: the first (referred to as an *identical* setting) corresponding to the agreement between observers computed on the same shape i ($S_{n,m}^{i,i}$), and the other (referred to as an *across* setting) corresponding to the agreement computed across shapes i and j ($S_{n,m}^{i,j}$). We then run one-tailed t-tests to assess the superiority of the *identical* setting. The null hypothesis is that the sets of similarity values from the two settings come from normal distributions with equal means, while the alternative hypothesis is that the *identical* mean is higher than the *across* one.
We also provide the number of trials where *identical* wins and loses against *across*. Results are presented in table 1, in the left column. We actually applied exactly the same protocol to study the impact of other rendering parameters (i.e. camera movement, material and lighting) on the fixation maps. For example, to study the impact of shape material, we compare for each possible pair of materials (k, l), $S_{n,m}^{k,k}$and $S_{n,m}^{k,l}$ for each combination of camera movement, lighting and shape and each possible pair of observers.
The metric used for computing the similarity $S()$ between fixation maps, is the Pearson's linear correlation coefficient. To fasten the analysis and allows similarity computation across shapes, 3D objects are re-sampled into isotropic meshes of 1K vertices.

4.4. Results

Static scenes
Table 1 presents the results of the impact for each parameter for static 3D scenes. Obviously, the shape itself has a high impact on fixation maps, meaning that observers tend to generate more similar fixations for the same shape than for different ones. These results confirm what was observed in [WLL*16] for printed shapes.
Lighting seems to have an impact as well, while some conditions lead to similar results. Front-camera lighting provides results significantly different from the other conditions. This can be accounted for by the fact that the front lighting direction, as already observed for quality assessment experiments [RH01], tends to mask the geometric details of the shape. Figure 4 illustrates this effect: with front lighting, the geometric features are less visible on the rendered image and thus observers concentrate their fixation on the head of the dinosaur due to its semantic importance. On the contrary, the top-left lighting condition emphasizes the geometric details, thus drawing observers' attention to geometrically salient parts (e.g. the shoulder).
Finally, the material, in certain conditions, may also have a significant impact on fixation location. Significant differences are observed, in particular, between glossy and matte materials. Figure 5 illustrates this influence: with the matte material, observers tend to gaze on regions where a contrast exists due to shadows, as well as around the center of the shape. However, for the glossy material, their fixations are concentrated around specular reflections.

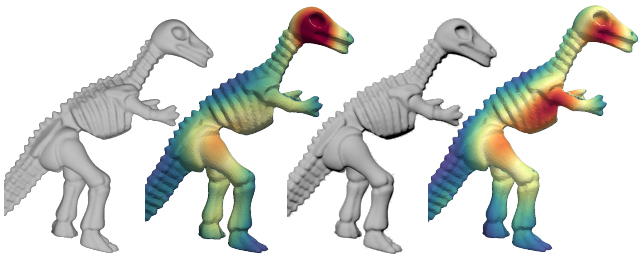

Figure 4: *Rendering and fixation maps for different illuminations (matte material). The scale of the color map is the same for both fixation maps. As front lighting decreases contrast, observers concentrate their gaze on the head of the dinosaur, while the top-left lighting condition emphasizes the geometric details which then draw observers' attention.*

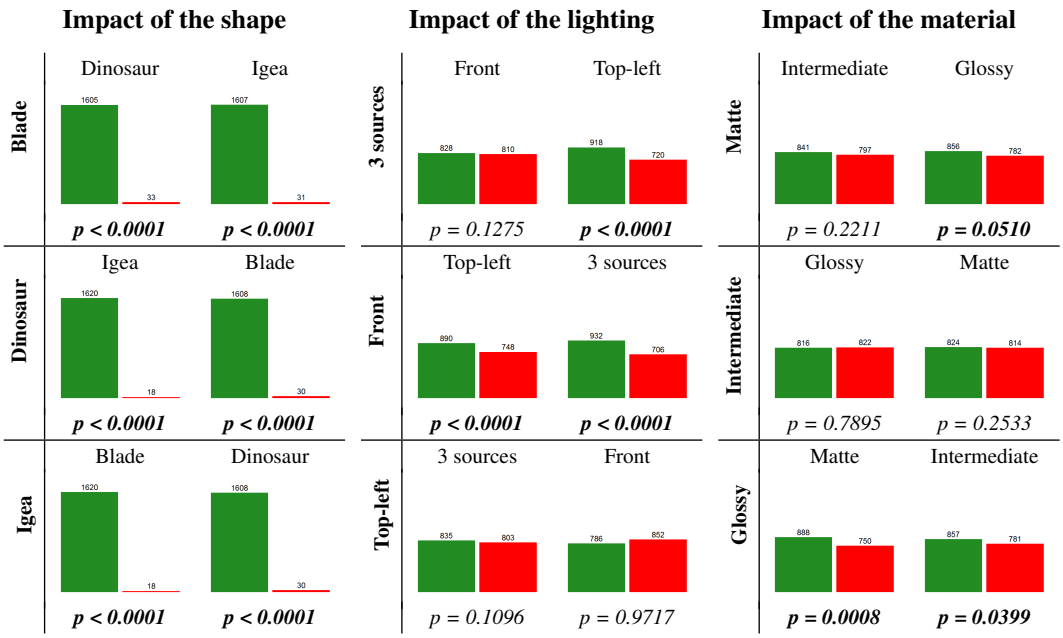

Table 1: *Influence of each parameter (shape, lighting and material) on the fixation maps of static scenes. Each parameter value is compared to the others (asymmetric comparison): we count the number of trials where the agreement of observers for the* identical *setting wins (green) or loses (red) regarding the agreement of observers for the* across *setting. We also provide the p-value for the rejection of the null hypothesis "There is no significant difference between these two parameter values".*

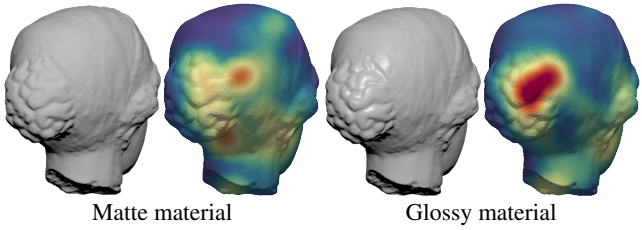

Figure 5: *Rendering and fixation maps for different materials (top-left lighting). For the glossy material, observers' attention is attracted by specular reflections.*

Dynamic scenes

Table 2 presents the results on dynamic 3D scenes. Similarly to static scenes, the shape itself has a high impact on fixation maps. Interestingly, we find that the camera path greatly influences eye fixations, while lighting and material seem to have less impact than for static scenes. This suggests that, in a dynamic 3D scene, camera movement is the prominent factor guiding user attention and that this movement even decreases the influence of lighting and material.

4.5. Differences between static and dynamic scenes

We observe that for the same 3D shape, fixations resulting from a dynamic scene are significantly different from those resulting from a static scene. For the latter, salient areas are mostly induced by geometry, emphasized to a greater or lesser extent by lighting and material. However, as stated above, in a dynamic scene, animation has a prominent role in the determination of salient parts. As illustrated in Figure 6, for the same object and the same combination of lighting and material, different camera movements may produce very different eye fixation maps. In this example, fixations are concentrated on Igea eyes and chin (prominent geometric features) for the static view, while they are spanned horizontally and vertically in the case of camera motion. This effect is due to the fact that people tend to keep looking near the center of the object/screen when there is a camera movement.

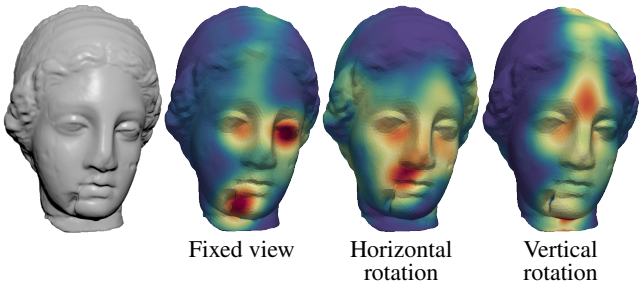

Figure 6: *Rendering and fixation maps for different camera paths (top-left lighting and glossy material). The camera movement strongly influences human attention.*

Another phenomenon accentuates the influence of camera movement on human fixations: as an object rotates about its center, some

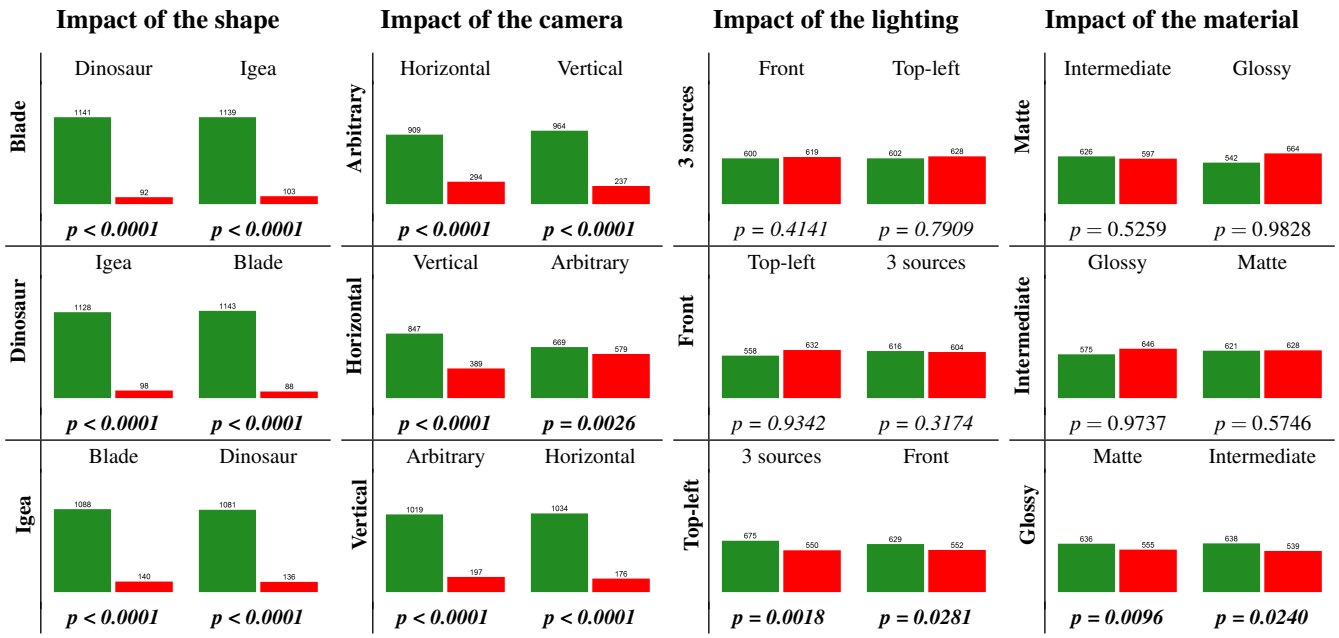

Table 2: *Influence of each parameter (shape, camera, lighting and material) on the fixation maps of dynamic scenes. Each parameter value is compared to the others (asymmetric comparison): we count the number of trials where the agreement of observers for the* identical *setting wins (green) or loses (red) regarding the agreement of observers for the* across *setting. We also provide the p-value for the rejection of the null hypothesis "There is no significant difference between these two parameter values".*

parts of its surface become visible to the observers and some other parts become occluded. To assess if observers tend to look more at the newly appearing parts of the rotating object, we analyze the distribution of the fixation map values with respect to the time during which the vertices are visible to the observers. For this purpose, we have computed, for each object and each camera movement, an histogram were the horizontal axis is the number of frames during which the vertices are visible to the observers. The vertical axis is the summation of fixation map values of the vertices. This histogram (blue curves in Figure 7) is compared to an histogram corresponding to a hypothetical observer looking at all the visible vertices equally for the entire duration of the video (red curves in Figure 7). These two histograms have been normalized such that their area is equal to 1. Figure 7 illustrates these histograms for the Blade and the Dinosaur objects (associated with the arbitrary camera movement). The complete set of histograms is available in the supplemental material. For the Blade model, the blue curve is above the red curve for vertices that are visible for less than 100 frames (about 1.33 second with a frame rate of 30 FPS) and the blue curve is below the red curve for vertices that are visible for more than 200 frames (about 6.66 seconds). This indicates that observers tend to look more at the newly appearing vertices rather than vertices that are visible since a large number of frames. In other words, geometric parts of the 3D model which appear suddenly during the video tend to attract human attention. This phenomenon is not observed for the Dinosaur model, for which eye fixations are mostly attracted by the head, associated with a strong semantic prior. Moreover this model is made of thin elongated parts which do not really hide each other from view.

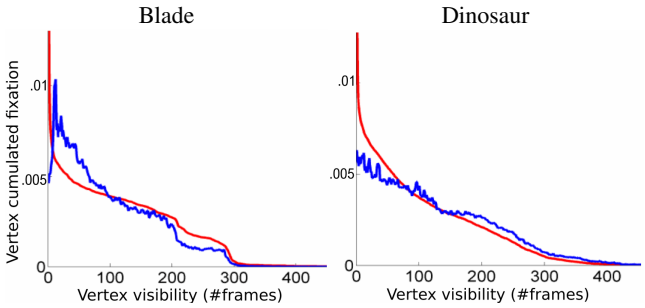

Figure 7: *Illustration of the influence of disocclusions (i.e., sudden appearance of hidden geometric parts) on human fixations in dynamic scenes, for the Blade and Dinosaur models (arbitrary camera movement). The blue curves represent the distributions of fixation map values with respect to the time during which the vertices are visible to the observers. The red curve is the histogram corresponding to a hypothetical observer looking at all the visible vertices equally.*

5. Experiment 2: Benchmarking Saliency Models

5.1. Saliency algorithm selection

Numerous saliency models have been introduced by the Computer Graphics community (see Section 2). We select the earliest and most popular one: the center-surround model from Lee et al. [LVJ05]. We also select several of the most recent and/or most cited algorithms: the spectral approach from Song et al. [SLMR14],

the algorithm from Leifman et al. [LST12] based on region distinctness and, finally, the point set saliency estimator from Tasse et al. [TKD15]. For accuracy of the results, we asked the authors of each method to compute the saliency on our models, except for Lee et al., for which we used the original author's implementation.

Center bias. With existing eye-tracking datasets for natural images [LLBT06,JEFT09,JDT12,BJO*16], researchers have observed that eye fixations tend to be biased towards the center of images. To check this hypothesis for rendered 3D models, we have introduced a center prior model. This model, illustrated in Figure 8, is a two-dimensional isotropic Gaussian function centered on the center of the bounding box of the rendered object and projected onto its 3D shape. As in [JDT12], we introduce *weighted* versions of the saliency models by fitting a linear model, as follows:

$$newMap = w \times centerMap + (1-w) \times saliencyMap \quad (2)$$

Blur. Analogous to center bias, several authors [JDT12] have noticed that blurrier saliency maps tend to perform better at predicting fixations than saliency maps with sharp edges. That seems straightforward since fixation maps are smooth by nature since they are derived from Gaussian filtering. Moreover, blurring saliency maps eliminates scale dependency caused by the fact that fixation map scale depends on the spatial resolution (in pixels per degree of visual angle) of the observed images. Therefore, for each saliency map obtained by automatic algorithms we produce 3 blurred versions, as illustrated in Figure 9.

Visibility. Note that to accurately measure the performance of these saliency models, we need to take into account the *visibility* of the shape under the viewpoints considered. For static scenes, this is a binary field (visible or not visible) over the 3D object. In the following performance calculation, our saliency models are multiplied by this visibility binary field.

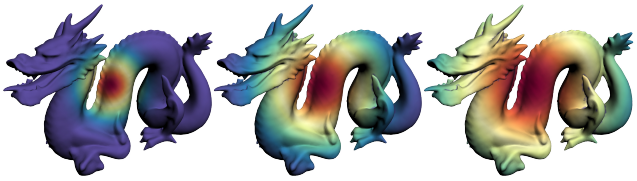

Figure 8: *Gaussian center model with increasing standard deviations. From left to right: 100, 200 and 300 pixels.*

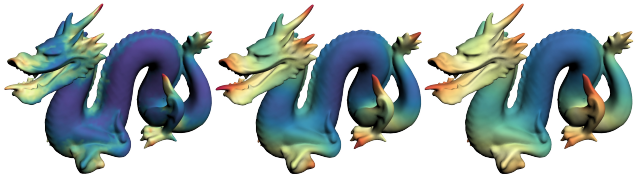

Figure 9: *Different blurred versions of the saliency map from Leifman et al. [LST12]. From left to right: no blurring, 40 iterations and 120 iterations.*

5.2. Stimuli

5.2.1. Static or dynamic ground-truth?

Before creating a benchmark for the evaluation of saliency predictors, we have to choose whether we select static or dynamic scenes. As shown in Section 4, these settings may produce very different results, and observers tend to agree less on dynamic scenes. Another problem of dynamic scenes is that their fixations are really hard to predict by automatic estimators as they are not directly related to 3D geometry, but rather to the changes in shadowing/reflection that occur during the camera movements. These suggest that predicting saliency of a 3D shape under a moving camera would require a specific dynamic saliency model. Since no existing models from the literature consider this aspect, we select static scenes for our benchmark.

5.2.2. 3D object selection and rendering parameters

We selected 32 models from different public databases: Aim@Shape †, TOSCA ‡, SHREC 2007 §, Georgia Tech model archive ¶ as well as from some private collections. These models have been carefully selected to ensure a large variety of shapes, vertex numbers and semantic importance. They have been properly remeshed whenever necessary using Vorpaline ‖ in order to obtain high resolution isotropic meshes, which constitute high quality inputs for both rendering and saliency algorithms. Compared to existing benchmarks for interest points [CSPF12, DCG12], most of our 3D models are of a higher resolution with more geometric details. They belong to four classes, as detailed in Table 3: *humans*, *animals and creatures*, *familiar objects* and *mechanical parts*. These classes have been chosen because they convey varying semantic priors (i.e. very high for Humans and very low for Mechanical parts). *Protein* has been placed in the *mechanical parts* class because of its low semantic prior. In the same spirit, *Hand* appears in the *familiar objects* class because it does not present the same semantic attraction level as human face or body.

As mentioned above, we have rendered *static* scenes for each of these models. To ensure we cover most of the shape, we consider three manually chosen viewpoints. To guarantee a tractable experiment and given the large number of 3D models, we used a single material and lighting condition. We selected the intermediate material (i.e. not too glossy nor too matte) and the top-left illumination since this is the lighting direction assumed by the visual system when viewing the image of a shaded 3D surface [SP98, OBA08]. We thus obtained a total of $32 \times 3 = 96$ rendered images.

To make performance computation tractable and eliminate unnecessary precision, both human fixation values and outputs of saliency algorithms are mapped on down-sampled versions of the 3D models. These versions are obtained by isotropic remeshing with 20K vertices ensuring a distance of roughly 0.1 degrees of visual angle between vertices, which we consider to be a sufficient

† http://visionair.ge.imati.cnr.it/ontologies/shapes/
‡ http://tosca.cs.technion.ac.il/book/resources_data.html
§ http://watertight.ge.imati.cnr.it/
¶ http://www.cc.gatech.edu/projects/large_models/
‖ http://alice.loria.fr/index.php/erc-vorpaline.html

Table 3: *Our dataset. Object sizes (in thousands of vertices) are given in parenthesis.*

Animals	Humans	Mechanical	Familiar objects
Horse (113)	Igea (101)	Casting (5)	Vase (15)
Dinosaur (42)	Planck (204)	Meca (15)	Car (17)
Bunny (35)	Bimba (75)	Blade (200)	House (196)
Dragon (50)	Torso (142)	Rocker (40)	Hand (37)
Cow (46)	James (51)	Carter (25)	A380 (173)
Octopus (17)	Jessi (70)	Fandisk (6)	Flowerpot (83)
Gargoyle (150)	Mickael3 (53)	Turbine (100)	Chair (11)
Camel (19)	Mickael8 (53)	Protein (50)	Harley (276)

precision. A380, Harley, and House data contain many tiny, concave parts and have thus been remeshed with respectively 48K, 55K and 30K vertices to obtain approximately the same resolution. Possible interior (i.e. not visible) parts have been removed from all these down-sampled versions. Note that these down-sampled versions serve only for the performance computation. The human fixation values and outputs of saliency algorithms are computed from the original high-resolution models.

5.3. Participants

20 observers participated to this experiment. Just like the first one, they were aged between 20 and 40, were naive to the goals of the experiment and had a normal or corrected to normal vision. The total time was 20 minutes (vision check + training/explanations + 96 images x 7 sec. + transitions). One observer was rejected due to the presence of too much invalid data.

5.4. Optimizing blurriness and centeredness of saliency models

As explained in Section 5.1, it has been observed in computer vision experiments that performance of saliency models is enhanced by blurring saliency maps and by combining them with a center bias model. To evaluate performance of the tested models independently from these blurriness and centeredness factors, we optimize their parameters for each 3D object and each saliency model, as follows:

- For the blurring level, for each saliency algorithm we create 4 versions of the saliency map (0, 10, 40 and 120 smoothing iterations) and select the top-performing one for each view of each 3D object.
- For the Gaussian center prior, we sample several σ values from 100 to 400 pixels and select the best performing value for each view of each 3D object.
- For each view of each 3D object, the linear model between the blurred saliency map and the center bias (see Equation 2) is fitted on the remaining objects from the same class.

5.5. Results and observations

The 96 fixation maps obtained (32 objects $\times$ 3 views) provide an opportunity to investigate the performance of saliency models in predicting human fixations. We also study the kind of features attracting visual attention and finally analyze the difference between eye fixations and Schelling/interest points. Examples of fixation maps are illustrated in Figures 10 and 15 (together with saliency maps). Fixation and saliency maps for all objects are available in the supplementary material.

Figure 10: *2D heat maps and 3D mesh fixation maps for several objects. From left to right and top to bottom: A380, Turbine, RockerArm, Car, Igea, Cow, Meca, James, Camel, Michael3, Dinosaur, Bimba, Carter and Vase.*

Overall saliency model performance. Figure 11 shows the overall performance of humans and saliency models (including the centered model) using two metrics: Area under ROC curve (AUC) and Pearson's Linear Correlation Coefficient. To calculate human performance, we use the fixation map obtained from ten random observers to predict the fixation map obtained by the ten others. Both performance metrics show similar tendencies: (1) The centered models provide better results than every other saliency model, meaning that, as with natural images, observers tend to look at the center of objects. (2) Even combined with a centered model, saliency algorithms remain poor at predicting fixation locations, compared to human performance; best methods do not exceed 0.50 linear correlation and 0.76 AUC. (3) Results show very high standard deviations, suggesting that performance of saliency models (including the center one) greatly depends on the 3D objects displayed.
To assess the superiority of saliency models among others, we conducted pairwise Student's t-tests on AUC values; the resulting similarity groups at the 95% significance level are reported in Figure 12; p-values and results for the linear correlation metric are detailed in the supplementary material.

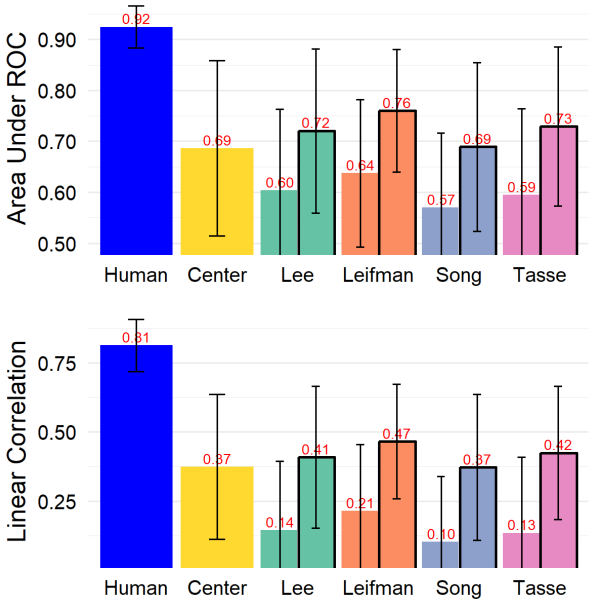

Figure 11: *Performance, in terms of area under ROC curve (top) and linear correlation (bottom), of humans, center model and saliency models from Lee [LVJ05], Leifman [LST12], Song [SLMR14] and Tasse [TKD15]. Bars with black borders mean that saliency models are combined with the center model according to eq. 2*

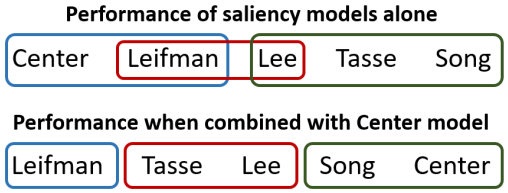

Figure 12: *The 95% similarity groups, as revealed by one-tailed paired t-tests (conducted on AUC values). Items in the same set are statistically indistinguishable. Saliency models are ranked from left (best) to right (worst).*

Analysis of the center bias. Figures 13 and 14 detail performance results for each class and each 3D object, respectively. More results are given in the supplementary material. We observe that the center bias is more important for *mechanical* and *familiar* objects. Such behavior is related to the semantic prior of the 3D objects. For mechanical parts, observers do not recognize anything as perceptually salient and thus tend to gaze at the center of the rendered image. The same effect is observed with the most simple objects from the *familiar objects* class (e.g., chair, flowerpot and vase) or, on the contrary, with highly complex models (e.g., protein). For *humans* and *animals*, the observers' gaze is attracted by semantic features (faces, eyes) and thus moves away from the center. The section below provides more details about features that attract human attention.

Where do people look on a rendered 3D shape? Besides the center of objects, we identified the features that attract people's attention when looking at a rendered 3D shape. They are described below.

- *Faces (human and animal)*. Looking at fixation maps from all animal and human models, it seems obvious that the *face* strongly attracts eye fixations (see Michael, James, Camel, Dinosaur, Cow in Fig. 10 and Gargoyle and Horse in Fig. 15). Surprisingly this semantic attraction to the *head* is also observable with the Airplane model (Fig. 10), for which the head attracts fixations whereas it does not present any particular geometric feature. Note that, for certain human faces, fixations seem to be more concentrated on cheeks than on eyes. This surprising result would need a further specific investigation.
- *Large scale geometric features*. As can be seen on the Blade (Fig. 15) and, RockerArm and Vase model (Fig. 10), large scale geometric features like protrusions, have an effect on the gaze attraction. It can also be observed with the hair buns of Bimba and Igea.
- *Small scale geometric features*. Tiny or sharp geometric details greatly attract attention as well, particularly when they are located on smooth areas. This effect clearly appears in Fig. 10 on the noisy cracks of Turbine, on the tiny geometric artifact at the bottom of Vase and on the geometric features (e.g., mirror, antenna) of the Car model.
- *Holes and other topological features*. As can be observed on Carter and Meca (Fig. 10), topological holes also attract visual attention.
- *Highlights and other lighting effects*. Finally, as observed in Experiment 1, specular reflections are also attractive. For instance, for Cow (Fig. 10) and Horse (Fig. 15), besides the head, people clearly directed their gaze to the specular highlights of the body.

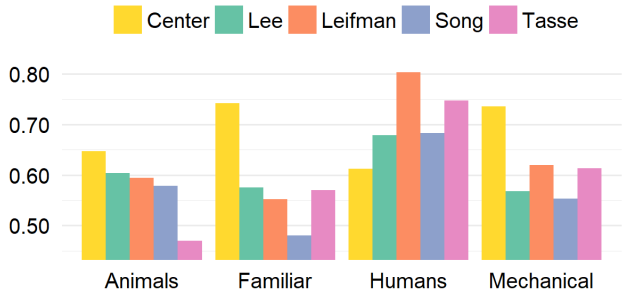

Figure 13: *AUC values per class for all tested saliency models (without the combination with the center model).*

Analysis of successes and failures of saliency models. As detailed in the paragraph above, several semantic and geometric features have proved they attract attention. However, looking at the fixation maps, it seems very complicated to predict, for each object, which feature will have more influence than others and thus will attract the main attention of the observer. The complexity of this cognitive mechanism accounts for the poor results of tested saliency models. In the Animal class, saliency models detect protrusions such as

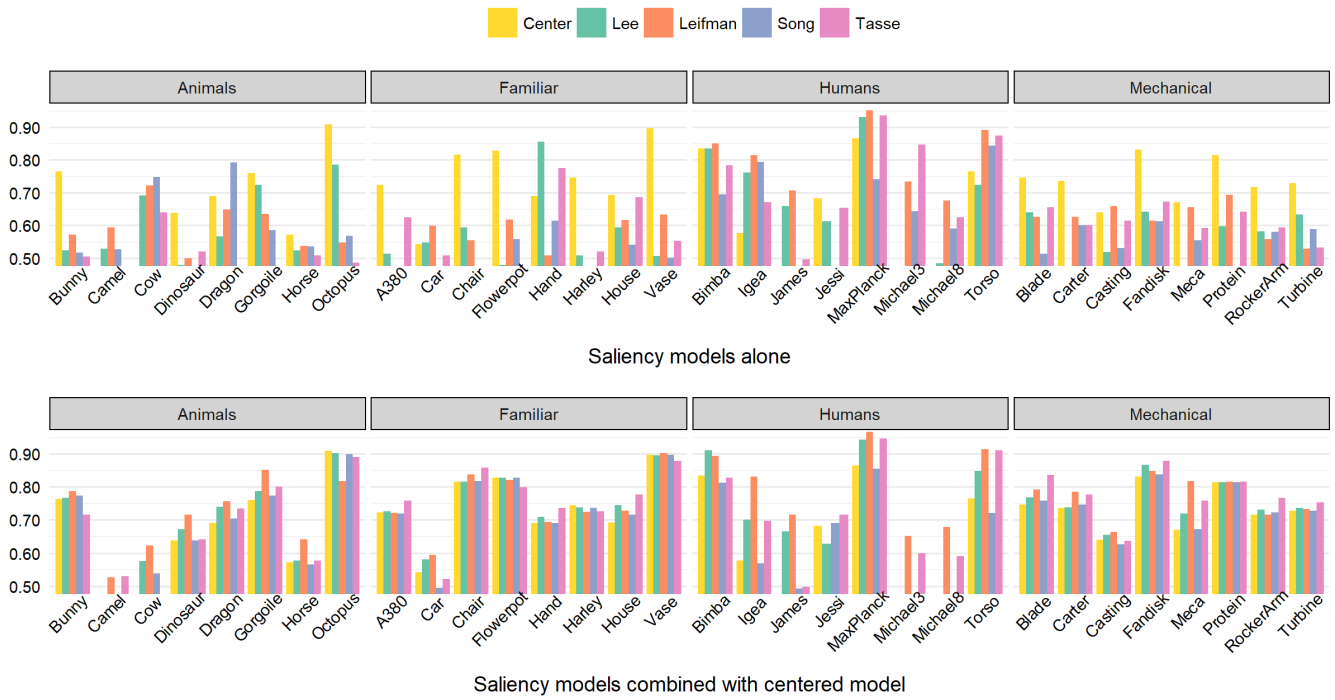

Figure 14: *AUC values per 3D object for all tested saliency models (alone and combined with the center model).*

legs, ears and tails, while fixations concentrate on faces (see Gargoyle, Horse and Dragon in Figures 15 and 1). In the Familiar and Mechanical classes, fixations are mostly around the center, while saliency models tend to detect sharp features (see Blade in Figure 15). However, combinations of saliency models with the centered Gaussian may yield results quite close to fixation maps (e.g. see the good performance of Tasse and Leifman for Carter and Meca, and of Tasse for House in Figure 14, in the bottom row). The relative success of saliency models for the Human class is due to the positive correlation between the semantic features (eyes, nose) and the geometric features which are detected by these models (see MaxPlanck in Figure 15).

Difference with Schelling/interest points. As discussed in Section 2, two existing benchmarks [CSPF12, DCG12] have already been used for evaluating mesh saliency algorithms, such as in the recent study by Tasse et al. [TKD16]. In the experiment conducted by Chen et al. [CSPF12] people were asked to "select points on the surface of a 3D object likely to be selected by other people". In a similar way, Dutagaci et al. [DCG12] asked people to "mark all the points they think are interesting or defining". While these data are interesting, it is questionable whether these Schelling/interest points are really correlated with human fixations. We have selected ten objects from our benchmark, which have already been used in the Schelling points benchmark from Chen et al. [CSPF12] and have computed the similarity between fixation maps and Schelling distributions. Figure 16 details the results in terms of AUC (linear correlation results are presented in the supplementary material) while Figure 17 visually compares these two human-generated scalar fields. The results show that Schelling points and human fixations are not correlated except for Human faces where they both concentrate on semantic features such as nose, eyes and mouth. Schelling points tend to concentrate on protrusions (as can be seen in Figure 17), whereas fixations stem from a more complex cognitive process.

6. Conclusion and perspectives

In this paper, we present two eye-tracking experiments conducted on rendered 3D shapes, providing fixation density maps on 3D meshes. We show that, for static scenes, both material and lighting have a significant influence on human visual attention. This influence decreases for dynamic scenes, in which the camera path becomes the prominent factor driving attention, together with the 3D shape itself. We provide extensive comparisons of several saliency models from the state-of-the-art, for prediction of human fixations, and demonstrate the significant importance of the center bias. We also investigate the main factors that attract human attention in a 3D setting. We believe that these results and new insights into human viewing behavior, as well as our publicly available dataset, will greatly stimulate research on saliency models for 3D graphics.
Limitations and future work directions. Our study is one of the first that aims at understanding human attention for rendered 3D shapes, and thus have some limitations that open avenues for future work. First, we considered non-textured, Phong-shaded 3D models, which is not the most realistic shading scenario. A future study could involve more physically correct materials as well as realistic environment maps for scene illumination; such study could

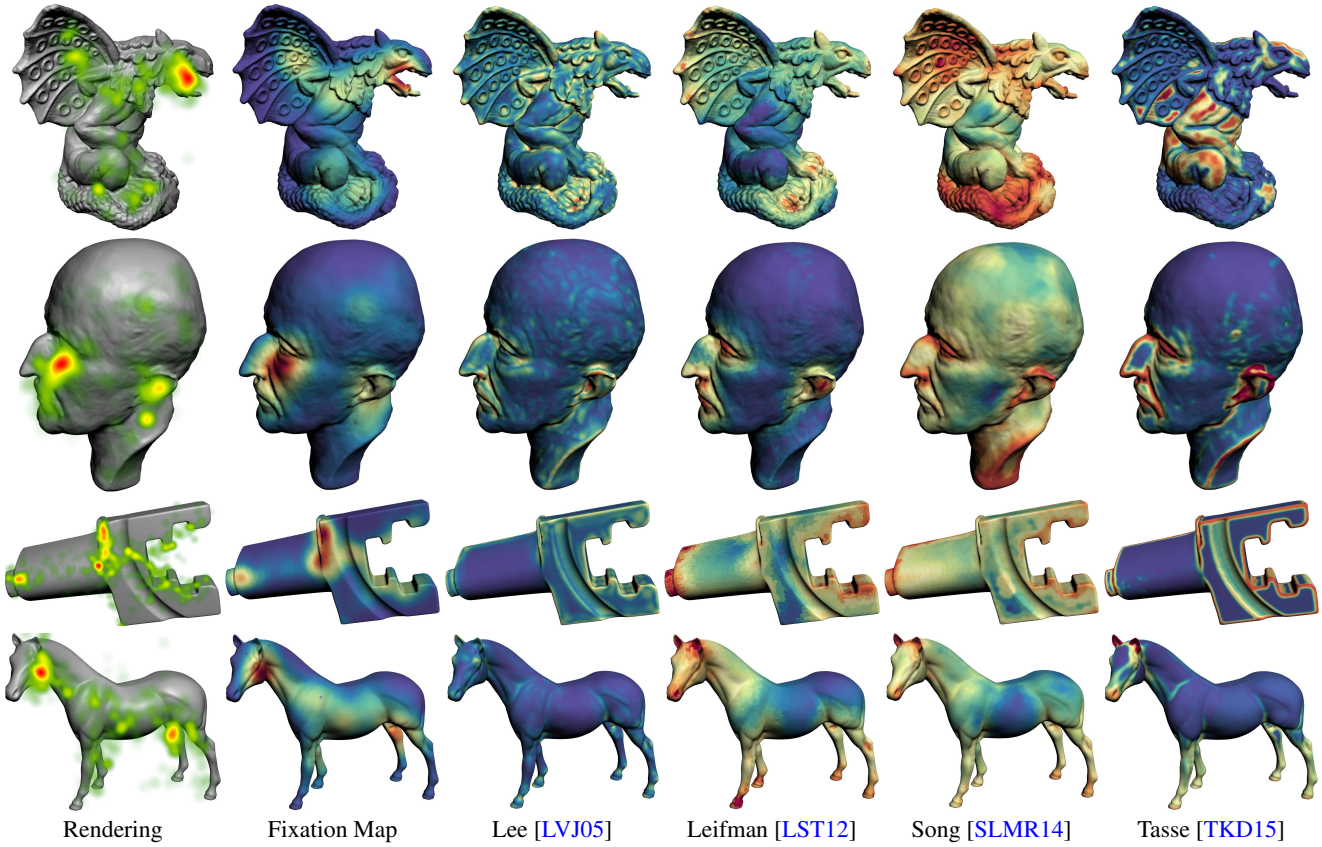

Figure 15: *Rendering (with 2D heat maps), fixation maps and tested saliency models for several 3D objects. From top to bottom: Gargoyle, Max-Planck, Blade and Horse*

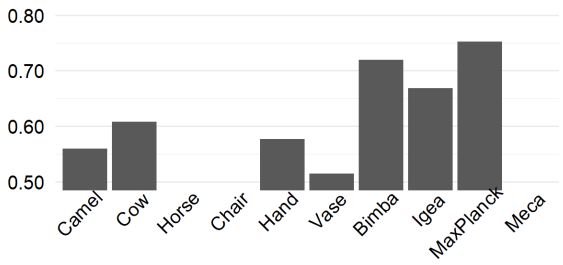

Figure 16: *AUC values between human fixations and Schelling distributions [CSPF12].*

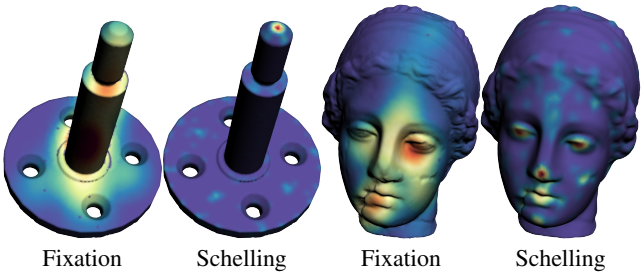

Figure 17: *Visual comparison between human fixations and Schelling distributions [CSPF12]. Schelling distributions tend to concentrate on protrusions.*

allow to confirm and enrich recent results on material perception (e.g., [KFB10]) that suggest a strong interplay between geometry, material and illumination in the perception of 3D shape. Second, we chose a free-watching scenario without any specific task assigned to the observers. While this protocol is valid and widely used for producing fixation maps in computer vision, it produced a strong center bias in our experiment (especially in the dynamic setting). It could be of interest to conduct an experiment where different tasks are given to observers (e.g., describe the object to another person) to evaluate their influence. Third, we did not evaluate the performance of image-based saliency algorithms [BTSI13] in predicting our 3D saliency maps. Their evaluation could lead to interesting and surprising results.

Note that several works have been carried out related to high-level, task-oriented gaze prediction in the context of video games [BSW10, KDCM14, KDCM16]. Most of these studies are based on high-level semantic properties and tend to ignore the appearance or shape of 3D objects in the scene. In our opinion, a good saliency

predictor for such complex 3D scene should rely on an appropriate combination of a bottom-up appearance-based model and a top-down task-oriented model.

Acknowledgements

We are deeply grateful to Flora Tasse, George Leifman, Ran Song and Chang Ha Lee for kindly running their saliency algorithms on our dataset (or providing their source code). We also thank Kevin Gaudet for his help to compute the fixation density distributions and the observers who participated in the experimentation. We finally thank all the anonymous reviewers whose feedback led to significant improvements. This work was partly supported by Auvergne-Rhone-Alpes region under the COOPERA grant "ComplexLoD", by the CPER NUMERIC within the Cinema action of e-Creation and by French National Research Agency as part of ANR-FILTER2 project (ANR-16-CE39-0013) and ANR-PISCo project (ANR-17-CE33-0005-01).

References

[Aut17] AUTODESK: *3ds Max*. https://www.autodesk.com/products/3ds-max/overview, 2017. 3

[BJO*16] BYLINSKII Z., JUDD T., OLIVA A., TORRALBA A., DURAND F.: *What do different evaluation metrics tell us about saliency models?* Tech. rep., 2016. arXiv:1604.03605. 2, 4, 8

[BSW10] BERNHARD M., STAVRAKIS E., WIMMER M.: An empirical pipeline to derive gaze prediction heuristics for 3D action games. *ACM Transactions on Applied Perception 8*, 1 (2010), 1–30. 2, 12

[BTSI13] BORJI A., TAVAKOLI H. R., SIHITE D. N., ITTI L.: Analysis of scores, datasets, and models in visual saliency prediction. *IEEE International Conference on Computer Vision* (2013), 921–928. 2, 12

[CSPF12] CHEN X., SAPAROV A., PANG B., FUNKHOUSER T.: Schelling points on 3D surface meshes. *ACM Transactions on Graphics 31*, 4 (jul 2012), 1–12. 2, 3, 8, 11, 12

[DCG12] DUTAGACI H., CHEUNG C., GODIL A.: Evaluation of 3D interest point detection techniques via human-generated ground truth. *The Visual Computer 28*, 9 (jun 2012), 901–917. 3, 8, 11

[HHO05] HOWLETT S., HAMILL J., O'SULLIVAN C.: Predicting and evaluating saliency for simplified polygonal models. *ACM Transactions on Applied Perception 2*, 3 (2005), 286–308. 2

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 11 (1998), 1254–1259. 2

[JDT12] JUDD T., DURAND F., TORRALBA A.: A Benchmark of Computational Models of Saliency to Predict Human Fixations A. *MIT-CSAIL-TR-2012-001* (2012). 8

[JEFT09] JUDD T., EHINGER K., FREDO DURAND, TORRALBA A.: Learning to Predict Where Humans Look. In *International Conference on Computer Vision* (2009). 2, 8

[KDCM14] KOULIERIS, DRETTAKIS G., CUNNINGHAM D., MANIA K.: An Automated High-Level Saliency Predictor for Smart Game Balancing. *ACM Transactions on Applied Perception 11*, 4 (2014), 1–21. 2, 12

[KDCM16] KOULIERIS G. A., DRETTAKIS G., CUNNINGHAM D., MANIA K.: Gaze prediction using machine learning for dynamic stereo manipulation in games. *2016 IEEE Virtual Reality (VR)* (2016), 113–120. 2, 12

[KFB10] KRIVÁNEK J., FERWERDA J. A., BALA K.: Effects of global illumination approximations on material appearance. *ACM Transactions on Graphics 29*, 4 (2010), 1. 12

[KVJG10] KIM Y., VARSHNEY A., JACOBS D., GUIMBRETIÈRE F.: Mesh saliency and human eye fixations. *ACM Transactions on Applied Perception (TAP) 7*, 2 (2010), 1–13. 2

[LB13] LE MEUR O., BACCINO T.: Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods 45*, 1 (2013), 251–66. 4

[LL15] LE MEUR O., LIU Z.: Saccadic model of eye movements for free-viewing condition. *Vision Research 116* (2015), 152–164. 2

[LLBT06] LE MEUR O., LE CALLET P., BARBA D., THOREAU D.: A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 5 (2006), 802–817. 2, 8

[LLS*16] LIU X., LIU L., SONG W., LIU Y., MA L.: Shape context based mesh saliency detection and its applications: A survey. *Computers and Graphics 57* (2016), 12–30. 2

[LST12] LEIFMAN G., SHTROM E., TAL A.: Surface regions of interest for viewpoint selection. *Computer Vision and Pattern Recognition* (2012). 1, 2, 8, 10, 12

[LTC*16] LIU X., TAO P., CAO J., CHEN H., ZOU C.: Mesh saliency detection via double absorbing markov chain in feature space. *Visual Computer 32*, 9 (Sept. 2016), 1121–1132. 2

[LVJ05] LEE C., VARSHNEY A., JACOBS D.: Mesh saliency. In *ACM Siggraph* (2005), pp. 659–666. 1, 2, 7, 10, 12

[MBM13] MANTIUK R., BAZYLUK B., MANTIUK R. K.: Gaze-driven Object Tracking for Real Time Rendering. *Computer Graphics Forum 32*, 2pt2 (2013), 163–173. 2

[OBA08] O'SHEA J., BANKS M., AGRAWALA M.: The assumed light direction for perceiving shape from shading. *ACM Symposium on Applied perception in graphics and visualization* (2008). 8

[RH01] ROGOWITZ B. E., HOLLY E. RUSHMEIER: Are image quality metrics adequate to evaluate the quality of geometric objects? *Proceedings of SPIE* (2001), 340–348. 5

[SF07] SHILANE P., FUNKHOUSER T.: Distinctive regions of 3D surfaces. *ACM Transactions on Graphics 26*, 2 (jun 2007). 2

[SLMR13] SONG R., LIU Y., MARTIN R., ROSIN P.: 3D point of interest detection via spectral irregularity diffusion. *The Visual Computer* (2013), 1–10. 2

[SLMR14] SONG R., LIU Y., MARTIN R. R., ROSIN P. L.: Mesh Saliency via Spectral Processing. *ACM Transactions on Graphics 33*, 1 (2014). 1, 2, 7, 10, 12

[SP98] SUN J., PERONA P.: Where is the sun? *Nature neuroscience* (1998), 183–184. 4, 8

[TCL*15] TAO P., CAO J., LI S., LIU X., LIU L.: Mesh saliency via ranking unsalient patches in a descriptor space. *Computers & Graphics 46* (2015), 264–274. 2

[TKD15] TASSE F. P., KOSINKA J., DODGSON N.: Cluster-Based Point Set Saliency. *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), 163–171. 1, 2, 8, 10, 12

[TKD16] TASSE F. P., KOSINKA J., DODGSON N. A.: Quantitative Analysis of Saliency Models. In *Siggraph Asia Technical Briefs* (2016). 11

[WLL*16] WANG X., LINDLBAUER D., LESSIG C., MAERTENS M., ALEXA M.: Measuring Visual Salience of 3D Printed Objects. *IEEE Computer graphics and application 36*, 4 (2016). 2, 4, 5

[WSZL13] WU J., SHEN X., ZHU W., LIU L.: Mesh saliency with global rarity. *Graphical Models 75*, 5 (2013), 255–264. 2

[Yar67] YARBUS A. L.: *Eye Movements and Vision*. Plenum. New York., 1967. 1, 3

[YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics 20*, 1 (2001), 39–65. 2