Most Relevant Viewpoint of an Object: A View-Dependent 3D Saliency Approach

Marie Pelissier-Combescure[a], Sylvie Chambon[b] and Géraldine Morin[c]
University of Toulouse, IRIT, Toulouse INP, France

Keywords: View-Dependent Saliency, Best View Selection, 3D Models/2D Images, User Study Evaluation.

Abstract: A viewpoint of a 3D object is the position from which we observe the object. A viewpoint always highlights some 3D parts and discards other parts of an object. Here, we define a good viewpoint as offering a relevant view of the object: a view that best showcases the object and that is the most representative of the object. Best view selection plays an essential role in many computer vision and virtual reality applications. In this paper, given a model and a particular viewpoint, we want to quantify its relevance -not aesthetics. We propose a geometric method for selecting the most relevant viewpoint for a 3D object by combining visibility and view-dependent saliency. Evaluating the quality of an estimated best viewpoint is a challenge. Thus, we propose an evaluation protocol that considers two different and complementary solutions: a user study with more than 200 participants to collect human preferences and an analysis of image dataset picturing objects of interest. This evaluation highlights the correlation between our method and human preferences. A quantitative comparison demonstrates the efficiency of our approach over reference methods.

1 INTRODUCTION

In video games, guiding players through a 3D environment via a 2D screen is a challenge. Techniques from movies, such as sound, lighting, and camera angles, are used to convey information without extra text or signs. The viewpoint of an object has a similar effect: a well-highlighted and easily recognizable angle makes an object more attractive. As (Fabricatore and *et al.* , 2002) notes, clear understanding of object semantics is crucial for good player interaction. This aligns with the main subject of this paper: best view selection.

(a) (b) (c)

Figure 1: Three examples of viewpoints for a cup: a relevant view (a), an accidental view (b) and an occluded view (c).

[a] https://orcid.org/0009-0004-2797-462X
[b] https://orcid.org/0000-0001-8104-1637
[c] https://orcid.org/0000-0003-0925-3277

In our work, accurately place and well-highlight 3D features in 2D interfaces is essential to get the best understanding of the object. To assess the relevance of a viewpoint, we measure the essential 3D information it contains. In Figure 1, there are multiple viewpoints of a cup. Figure 1a presents an relevant viewpoint where all representative parts are visible without any ambiguity. However, Figure 1b depicts an so-called accidental view (Blanz and Tarr, 1999), causing a perspective issue that reduces the global understanding. Moreover, obscured views, like Figure 1c, hides some characteristic parts, leading to misinterpretation; in this case, the cup could be mistaken for a glass

Traditionally, given a 3D object, in order to evaluate the quality of a point of view, two aspects can be considered: the amount of visible features (Plemenos and Sokolov, 2006) or its saliency (Leifman and *et al.* , 2016). Our first contribution consists in proposing an original measure of the viewpoint relevance that combines several efficient view-dependent 3D characteristics, see Figure 2 (inside the blue box).

To evaluate our method, we establish an original evaluation protocol: we compare our best viewpoint with viewpoints chosen explicitly or implicitly by users, as illustrated in Figure 2 (inside the purple box). More precisely, we have carried out a user study

Pelissier-Combescure, M., Chambon, S. and Morin, G.
Most Relevant Viewpoint of an Object: A View-Dependent 3D Saliency Approach.
DOI: 10.5220/0012379500003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 3: VISAPP*, pages 482-491
ISBN: 978-989-758-679-8; ISSN: 2184-4321

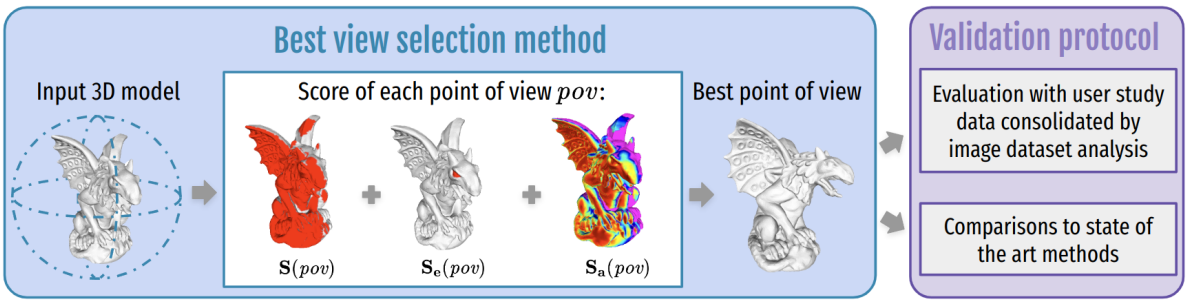


Figure 2: Three terms (surface visibility $\mathbf{S}$, eyes surface visibility $\mathbf{S_e}$, and saliency of visible vertices $\mathbf{S_a}$) are computed for each viewpoint (blue box) from a 3D model. The view with the highest score is selected as the best viewpoint. Evaluation (purple box) is based on user study.

on 3D objects to collect human preferences, and we also analysed 2D images relative to the same 3D objects to extract reference viewpoints. Finally, we also compare our method with state of the art best viewpoint selection approaches.

After presenting a review of methods for best view selection, in Section 2, we introduce our approach based on a view-dependent 3D saliency, in Section 3. In Section 4, for validation, we conduct a rigorous user study to capture human preferences and present an original metric to assess the similarity between a proposed viewpoint and this ground truth. Based on this evaluation, we analyse the proposed score and include an ablation study, in Section 5, and finally present comparisons in Section 6, before concluding.

2 SELECTING THE BEST VIEWPOINT

The purpose of best view selection is to optimize the visual experience by identifying the most suitable viewpoint for a given, application, or user interaction, like for example when visualising a 3D object within a 3D virtual environment (Habibi and *et al.* , 2015) or for choosing among multiple videos capturing an object of interest during the same event (Rudoy and Zelnik-Manor, 2012). Different methods have been proposed for taking advantage of the geometric model attributes, which can be classified into two categories.

Face-Based Methods. (Vázquez and *et al.* , 2001) introduce the concept of *viewpoint entropy* which uses the projected face areas to determine the amount of information in a given view. Unfortunately, this method is dependent on the initial scene subdivision. To address this limitation, (Sokolov and Plemenos, 2005) propose to rather use a measure based on curvature.

Vertex-Based Methods. Methods can be split into two stages: the first step involves calculating the 3D intrinsic saliency of every vertex, followed by selection of the best view.

For the best view selection, (Nouri and *et al.* , 2015) associate to each viewpoint, a score that corresponds to the sum of the intrinsic saliencies of the visible vertices. (Leifman and *et al.* , 2016) and (Feixas and *et al.* , 2009) incorporate a weighted sum based on the angle between the surface normal at the vertex and the viewing direction.

For the use of the saliency, some best view selection methods attempt to combine different features of models. (Secord and *et al.* , 2011) proposes a linear combination of criteria, related to different aspects of a view such as surface or silhouettes that closely aligns with human preferences according to view *goodness*. (An and *et al.* , 2016) propose a method that computes an independent hybrid saliency for each vertex based on geometric and color information of the input mesh. Then, (Wu and *et al.* , 2013) present an approach that take into account both local contrast and global rarity.

All these methods aim to mimic human aesthetic preferences. In a complementary manner, in our work, we want to determine the most relevant view of an object: the most representative, the view that presents the most informative parts of the object in order to understand how it works or to recognize it unambiguously. In other words, the selected views should be similar to those that a human might have chosen, relative to its needs for a given task, and, so, coherent with our goal: measuring relevance, not aesthetic. To ensure this condition, we use the results of our user study to evaluate our method and others. Now, we will describe the details of our proposed geometric method for best view selection.

3 PROPOSED BEST VIEW SELECTION

Our proposed Point Of View *pov* score, that automatically selects the best point of view, is based on 3D saliency relative to a chosen view. A first step to consider viewpoint dependency is to determine the visibility of a vertex (Lee and *et al.* , 2005) or a face (Plemenos and *et al.* , 2004). We further weight vertex saliency by the layout of a visible vertices relative to the viewpoint.

3.1 Viewpoint Score

We introduce three different terms: visible surface size for the whole model and for the eyes (when applicable) and intrinsic saliency of the model. Given a 3D model M and a particular *pov* :

- **Surface Visibility $\mathbf{S}(pov)$:** quantifies the amount of surface visibility according to a point of view, i.e. the ratio between the visible 3D surface and the total 3D surface of the model. (Secord and *et al.* , 2011) has shown that the single-attribute method that best matches human preferences is the one based on surface visibility, so we keep that term to respect this observation.
- **Eyes Surface Visibility $\mathbf{S_e}(pov)$:** indicates the amount of surface visibility of the eyes, i.e. the ratio between the visible 3D surface and the total 3D surface of the eyes. This choice is motivated by the conclusion of the work of (Secord and *et al.* , 2011) that points out the importance of the visibility of the eyes. In our work, we manually annotated the faces of the 3D models corresponding to the eyes.
- **Saliency of Visible Vertices $\mathbf{S_a}(pov)$:** represents the amount of 3D saliency relative to the viewpoint. Like (Leifman and *et al.* , 2016), we sum the saliencies of visible vertices, weighted by a function f dependent on the angle α_v between visible vertices and the camera.

$$\mathbf{S_a}(pov) = \sum_{v \in V} S_i(v).f(\alpha_v), \quad (1)$$

with $S_i(v)$ the saliency of a visible vertex v and V the set of visible vertices from *pov*. We consider different possibilities for $f(\alpha_v)$, see section 3.2. This saliency term is normalized by dividing by the maximum value obtained over all the saliencies computed for all viewpoints.

Finally, the *pov* score is determined by the sum of these three terms and the best viewpoint $\widetilde{pov}^M$ of the mesh M is the one maximising this score:

$$\widetilde{pov}^M = \underset{pov \in P}{\arg\max} \left(\mathbf{S}(pov) + \mathbf{S_e}(pov) + \mathbf{S_a}(pov)\right) \quad (2)$$

where P is the set of the available viewpoints. Details of the computation of visible vertices and angles from a viewpoint are detailed in the appendix.

3.2 Saliency of Visible Vertices

In the proposed formula, the term assigned to the saliency $\mathbf{S_a}$ depends on the choice of two parameters: an intrinsic saliency method S_i and an angle function f. We detail these two aspects.

Vertex 3D Saliency. 3D saliency are used to highlight regions of interest and to guide visual attention to the most important features. Some methods determine the two principal curvatures of model vertices to highlight these areas, for example: (Taubin, 1995), (Meyer and *et al.* , 2003), and (Rusinkiewicz, 2004).

Others may be multi-resolution by varying neighbour sizes and merging values from different scales, as (Lee and *et al.* , 2005), (Miao and Feng, 2010), (Wang and *et al.* , 2015), (Jeong and Sim, 2017) and (Meynet and *et al.* , 2019).

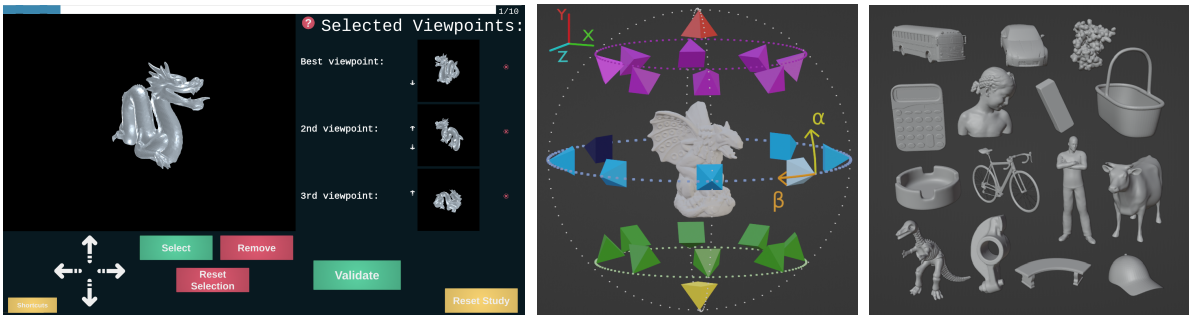
In this paper, we focus on three different salient features already identified in the literature, namely: (Song and *et al.* , 2014), (Tasse and *et al.* , 2015) and (Limper and *et al.* , 2016). More precisely, (Tasse and *et al.* , 2015) propose an approach that leverages cluster uniqueness and spatial distribution to formulate the final saliency detection results. In contrast to previous methods, (Song and *et al.* , 2014) do not use local geometric cues: they analyse the spectral attributes of the log-Laplacian spectrum to detect the irregularities of a 3D model. Then, (Limper and *et al.* , 2016) provides a multi-scale saliency value at each vertex based on Shannon entropy by using mean curvature.

Finally, for the vertex 3D saliency S_i, we have tested five different methods based on technique from classical mean curvature to entropy: (Lee and *et al.* , 2005), (Song and *et al.* , 2014), (Tasse and *et al.* , 2015), (Leifman and *et al.* , 2016) and (Limper and *et al.* , 2016)[1].

Visibility Weighting. For the angle function f, five different expressions depending on the angle α_v of view between the vertices and the camera are tested:

1. To favor the vertices facing the camera (with $\alpha_v = 0$) we used: $\cos(\alpha_v)$ and $\sqrt{\cos(\alpha_v)}$.
2. Inversely, to highlight the vertices on the sihouette we tested : $1-\cos(\alpha_v)$ and $1-\sqrt{\cos(\alpha_v)}$.

[1]Thanks to the dataset of (Lavoué and *et al.* , 2018), 26 models are supplied with saliency values for four methods and (Limper and *et al.* , 2016) provide their own code.



(a) Graphical interface of the user study (b) Set of camera positions (c) Examples of 3D models used

Figure 3: Illustrations of the technical aspects of the user study interface.

3. Finally, to take into account both the vertices facing the camera and those on the sihouette, we have tried: $0.5 + \frac{1-\sqrt{\cos(\alpha_v)}}{2}$.

In conclusion, we have twenty five different variants of the proposed score. In order to evaluate the interest of each proposition, we have investigated a user study, completed by analysis of real images. The proposed validation protocol is developed in the next section by distinguishing how we collect datasets and how we evaluate the quality of a selected viewpoint.

4 EVALUATION PROTOCOL

After reviewing existing evaluation methods, we present our user study, consolidated by the use of an image database well-known in the literature, before introducing a new quantitative evaluation metric.

4.1 Existing Validation Protocols

Validation of the best view selection methods is crucial to demonstrate the relevance of the selected views for an object or scene. Some methods rely on a qualitative evaluation shared by observers. These approaches do not involve external data or comparisons with other geometric methods (Vázquez and *et al.* , 2001), (Feixas and *et al.* , 2009), (Rudoy and Zelnik-Manor, 2012) and (Habibi and *et al.* , 2015). Others methods, such as (Plemenos and Sokolov, 2006), (Lee and *et al.* , 2005) and (Nouri and *et al.* , 2015), compare their results with those obtained from other methods, but the evaluation remains qualitative.

In our paper, we aim to provide quantitative evidence that the selected views correspond to human preferences. To achieve this, we conducted our own user study, collecting human preferences on a set of 3D models, see Section 4.2. Previous users study have been done using 3D data (Blanz and Tarr, 1999), or using comparison of two images (Secord and *et al.* , 2011), or more (Leifman and *et al.* , 2016) but unfortunately their results are not available anymore.

This kind of user evaluation is complex and expensive. In order to augment it, we propose in a second stage to also rely on the extraction of the best viewpoints from 2D images collected about the same 3D objects studied, see Section 4.3. Adding these images allows us to complete our evaluation by considering 3D textured objects in a particular context.

4.2 Proposed User Study

The 3D model database used in this work consists of 44 regular models. More precisely, we use 26 models from (Lavoué and *et al.* , 2018), and the other 18 are freely available on the web. A subset is illustrated in Figure 3c.

After recalling the main definitions about what is a viewpoint and what we define as a good viewpoint, the main question of the study is asked to the user: *Which viewpoint do you prefer to both showcase and recognize this object?* Then, each user has to study ten different 3D models, and to choose and order three best viewpoints to independently and positively answer the main question of the study. The ten models are randomly selected from the dataset and displayed from a starting random viewpoint. The part of the interface related to this step of the study is shown in Figure 3a. We register one JSON file[2] per user.

The Appendix section details the recruiting process carried out with the crowd-sourcing service offered by the Prolific[3] site. We want to obtain characteristic views of the objects, while minimizing the impact of aesthetic biases inherent in individual users. We achieve this by providing clear and precise instructions that emphasise relevance over aesthetics.

[2]Gitlab link : https://gitlab.irit.fr/bvs-study/data.git
[3]https://www.prolific.co/

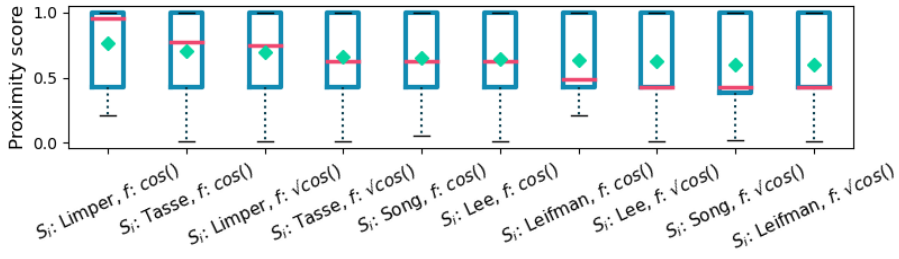


Figure 4: Statistical data of proximity score *PS* obtained by the ten best versions of our formula 2, specifying the intrinsic saliency method S_i and the angle function f used in the $\mathbf{S_a}$ term. Proximity scores are computed on the 26 models of (Lavoué and *et al.* , 2018).

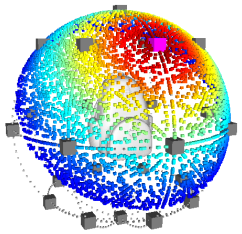


Figure 5: Representation of the 3D normal distribution centered in magenta camera and associated with the term **C** of the proximity score formula (3).

Models are presented without textures, placed in a neutral environment (a black screen), and initially displayed in random positions.

In order to harmonise the treatment of the models, they are positioned with the Y axis running from bottom to top: gravitational upwards (when it is meaningful). To minimise redundancy and to maximise the possible choices, we chose a compromise where the user can switch between 26 different camera positions distributed over a sphere centered on the mesh, as illustrated in Figure 3b.

Each camera has a label which depends on the value of two angles: (α, β), as illustrated in Figure 3b. For example, in the Figure 3b, the label for the red camera is *Top*, that of the yellow camera is *Bottom* and the label for the light blue camera is *Face*. Cameras clusters can be made according to the values of the angle α_j: the pink cameras form part of the *Top-Middle* set of cameras, then the blue cameras form the *Middle* set, and, finally the green cameras form the *Bottom-Middle* set. The results of the user study are represented by an histogram per 3D model indicating the popularity of the 26 viewpoints by a weight[4], see examples in Figure 7 and Figure 8.

4.3 Consolidation with Images

We have also collected images from the *Objectnet3D* (Xiang and *et al.* , 2016) database. Among our models, 22 correspond to a category in Objectnet3D where each image is associated a viewpoint. We have projected the different viewpoints of a given object category onto the study sphere. Then, as in the post-processing of the user study, (described in the appendix section), we have computed the distribution of the image views on the 26 study viewpoints. In this way, we determine which viewpoints are the most used when users take a photo, results are explained in Section 6.2.

[4]The appendix section details the computation of 26 viewpoints positions and weights for each 3D model.

4.4 New Evaluation Metric

To evaluate coherence with respect to user viewpoint selection, we defined a metric to determine a proximity score *PS* that represents the closeness between a given point of view and the one chosen by users. Given a mesh M, a determined best viewpoint $\widetilde{pov}_X^M$ by a geometric method X and the best viewpoint $\widetilde{pov}_U^M$ chosen by users, this metric is defined by:

$$PS_X^M = \max(\mathbf{C}(\widetilde{pov}_X^M, \widetilde{pov}_U^M), \mathbf{W}(\widetilde{pov}_X^M)) \quad (3)$$

More precisely:

- $\mathbf{C}(\widetilde{pov}_X^M, \widetilde{pov}_U^M)$ is the 3D normal distribution (Chave, 2015), centered in $\widetilde{pov}_U^M$, with support $\sigma = 1.3$ and translated to range $[0,1]$. This term measures the amount of information common to two viewpoints i.e. quantifies the information they share. The impact of the numeric value chosen for σ is illustrated in the Figure 5.
- $\mathbf{W}(\widetilde{pov}_X^M)$ is the weight associated with the $\widetilde{pov}_X^M$ in the histogram of model i, divided by the maximum weight of the histogram to range $[0,1]$. This term takes into account the views of all users, not just a single view.

In the case where $\widetilde{pov}_X^M$ does not correspond exactly to $\widetilde{pov}_U^M$, there are two situations in which $\widetilde{pov}_X^M$ is considered relevant to human opinion. First, it is possible that $\widetilde{pov}_X^M$ is a viewpoint that is not very popular with users but is semantically very close to $\widetilde{pov}_U^M$: two juxtaposed viewpoints on the sphere can share a large number of visible features. Thus

a viewpoint contained within the close neighborhood of the user viewpoint will be considered as acceptable proportionally to its distance from the exact position. Second, $\widetilde{pov}_X^M$ can be a viewpoint almost as popular as $\widetilde{pov}_U^M$, which corresponds to a meaningful viewpoint. After several empirical tests with different combinations, the use of the maximum allows us to consider these two situations. This evaluation metric is used in the rest of the paper both to choose the best configuration of the proposed method and to compare to state of the art approaches.

5 EVALUATION FOR PROPOSED METHOD

5.1 Saliency of Visible Vertices

In order to choose among the vertex saliency measures S_i and the angle function f discussed in Section 3.2, we compute the proximity score for each available version of our formula, for all models. Our formula has to provide best viewpoints that make sense to humans and are aligned with human reasoning. According to the Figure 4, which summarized the results in the form of a box-and-whisker diagrams, the proposed point of view score best performs using (Limper and *et al.* , 2016)[5] for computing S_i and $\cos(\alpha_v)$ for the angle function f. The angle functions in the top 10 are those that highlight the vertices facing the camera. We could expect the silhouette to give important information, but favouring silhouette probably also favours accidental views (Figure 1b).

5.2 Ablation Study

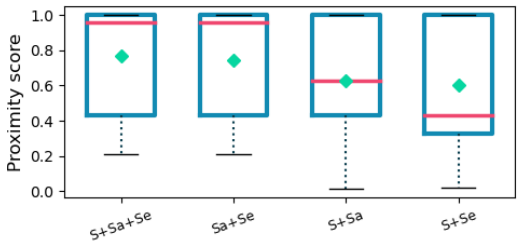
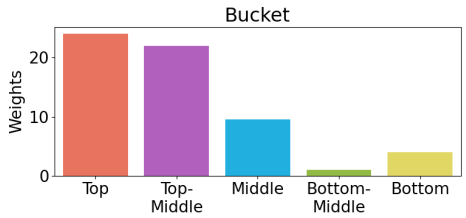


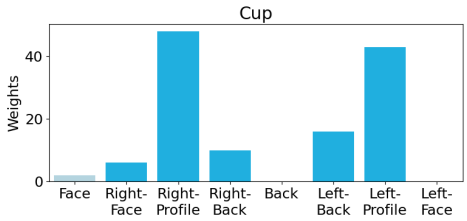
Figure 6: The ablation study highlights that the three terms **S**, $\mathbf{S_e}$ and $\mathbf{S_a}$ are contributing and, in particular, **S** improves the average by ~ 0.02.

To show the importance of each term in our formula 2, we perform an ablation study. The contribution of each term is displayed in Figure 6. The left box

[5](Limper and *et al.* , 2016) proposed a 3D saliency not a best view selection method.



(a) Example of a container object: comparing the latitude of the selected views, we see that most views are from above.



(b) Object with accidental and occlusion views: comparing the longitude of the selected views on the equator, we see that accidental and occlusion views are avoided.

Figure 7: Histogram of weights from different viewpoints (user study). In (a), cameras are arranged in clusters like the camera color code shown in Figure 3b whereas in (b), the histogram displays only the 8 blue cameras positioned on the equator of the sphere.

(three terms sum $\mathbf{S}+\mathbf{S_a}+\mathbf{S_e}$) has an average of 0.765 whereas the second one ($\mathbf{S_a}+\mathbf{S_e}$) has an average of 0.746. So, the surface visibility **S** does contribute but $\mathbf{S_a}$ and $\mathbf{S_e}$ have a stronger impact. Even if $\mathbf{S_a}$ is more important in average, note that only the models with a face (with eyes) have their proximity score changed by $\mathbf{S_e}$. With $\mathbf{S_e}$, the median rose from 0.627 to 0.955.

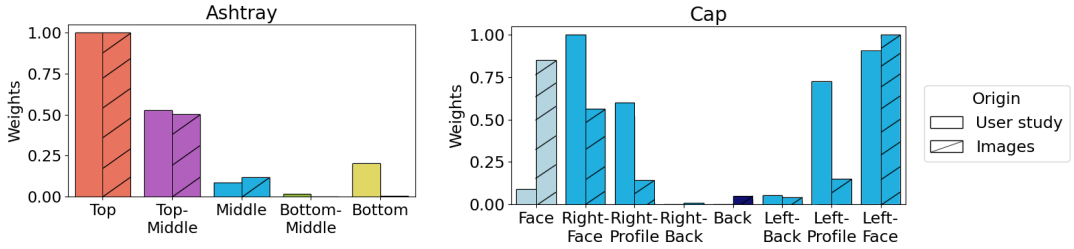
6 RESULTS AND COMPARISONS

To be able to compare with the human preferences collected during the user study, we only consider the 26 viewpoints previously mentioned for each of the 44 meshes studied.

6.1 Analysis of User Study

Now that we know which viewpoint users prefer, we can examine whether or not our observations are in agreement with those of (Blanz and Tarr, 1999) and (Secord and *et al.* , 2011):

- According to (Blanz and Tarr, 1999) : "*Most participants preferred off-axis views to straight front- or side-views*". Based on our observations, it depends



(a) (b)

Figure 8: Comparisons of human preferences between two pairs of non-textured 3D objects and 2D image of theses objects, like in Figure 9.

Figure 9: Non-textured 3D object (left) and an image (right) of a cap from the viewpoint *Face*.

on the nature of the object. Indeed, for models representing humans, the preferred viewpoints are those from the front, whereas for animals or creatures, off-axis and side views are strongly selected. Finally, for objects, this strongly depends on their functionality, contrary to (Blanz and Tarr, 1999) analysis. For example, top views are largely selected for objects that are containers. Figure 7a illustrates the distribution of viewpoints for the Bucket model (the colors match those of Figure 3b). The *Top* and *Top-middle* viewpoints have the greatest impact.

• Like in the observations of (Blanz and Tarr, 1999), accidental views are avoided. These are the types of view where there is a perspective problem: for example, in Figure 1b, the handle of the cup corresponds to a thick line. Figure 7b shows that this view for the cup, labelled *Back*, has never been chosen by users.

• Viewpoints where the model hides some of its parts, as in the Figure 1c with the cup hiding his handle, are also avoided. This observation is similar to that made by (Blanz and Tarr, 1999). In the case of the cup, this *Face* viewpoint is very rarely chosen, as shown in the histogram of Figure 7b.

• We integrate unfamiliar models into our study: mechanical parts and a protein. As in (Blanz and Tarr, 1999), these novel objects have no preferred view: all views have been selected for the protein. But accidental views are always avoided.

• For symmetrical objects, the left-hand view was as often selected as the right-hand view, see histogram in Figure 7b.

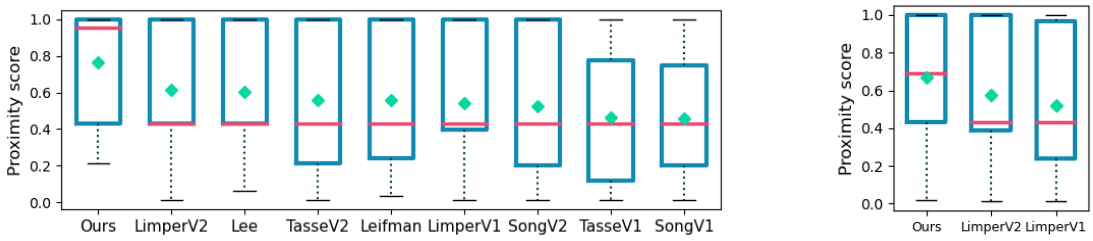
• In (Secord and *et al.* , 2011), the importance of the attribute that quantifies the presence of eyes in a viewpoint is high. This means that humans prefer to see models in the eyes, whenever possible. This hypothesis is confirmed by the fact that users chose frontal views when observing a human and slightly profile views for animals and creatures.

6.2 Comparisons with Real-Life Conditions

We analyse the best viewpoints extracted from the 2D images collection and viewpoints presented in Section 4.3. In most cases, the most frequently used viewpoints for taking pictures are similar to those selected in the user study. For example, Figure 8a shows the distribution of viewpoints used in the two situations: in our experimental conditions and in real-life conditions. In the case of the ashtray, the *Top* and *Top-Middle* views are largely selected. The main factor influencing the choice of viewpoint is the functionality of the object. However, the ashtray is often a neutral object in terms of texture. If we take the example of the cap, we can see that the most popular views are the *Right-Face* and *Left-Face* views, see Figure 8b. However, the *Face* viewpoint is much more chosen in real life than in the study. Indeed, logos are often visible on the front of caps, as in Figure 9. There is a desire to showcase the logo. In this case, texture needs to be considered and not only geometry.

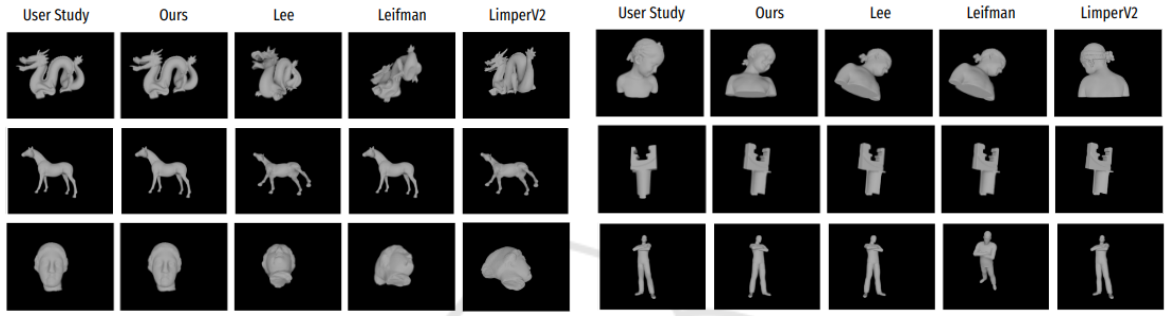
6.3 Comparisons with Studied Methods

To compare the efficiency of our formula 2, we considered two state-of-the-art geometric methods: (Lee and *et al.* , 2005) and (Leifman and *et al.* , 2016). Then, as previously mentioned, the dataset (Lavoué and *et al.* , 2018) provided the vertex saliency values for 26 models from two different methods: (Tasse and *et al.* , 2015) and (Song and *et al.* , 2014). We decided to use them in the two best view selection formulas proposed by (Lee and *et al.* , 2005) and (Leifman and *et al.* , 2016), since the subject of their work



(a) (b)

Figure 10: Statistical data of proximity scores *PS* from the best (left) to the worst (right): (a) computed on 25 models proposed in (Lavoué and *et al.* , 2018) and (b) all the 43 models (*prot* model has been withdrawn).



(a) (b)

Figure 11: Sub-set of best selected viewpoints by users (first column of (a) and (b)), by the proposed method (second column) and three other state-of-the-art geometric methods (from third to fifth columns). With some models, we have selected the same view as the users (a), while on others the viewpoints are different (b) but there are coherent and proposed a point of view that seems more informative.

is not the best view selection. These new methods are called respectively : TasseV1 and SongV1 (TasseV2 and SongV2) with the formula of (Lee and *et al.* , 2005) (formula of (Leifman and *et al.* , 2016)). Finally, to have a common method for all 44 models, we use entropy-based saliency computation method of (Limper and *et al.* , 2016). Once this saliency had been calculated for each vertex of models, we use the best view selection formula of (Lee and *et al.* , 2005) and (Leifman and *et al.* , 2016) to obtain a new geometric method, respectively named LimperV1 and LimperV2.

Each model has a best viewpoint for each of the nine studied methods. The proximity scores *PS* of each model for each studied methods are shown in Figure 10. In the first instance, after studying the results obtained by the *prot* model (shown in the upper right-hand of Figure 3c), no view was more relevant than an other: this model was removed from our study. In Figure 10a, we compare the best version of our formula with eight state-of-the-art methods. As can be seen, the diagram of our method is the best: 50% of models obtain a score above 0.955, while the other methods have a median of 0.43 at most. On average, our method achieves a proximity score of 0.76, while the others do not exceed 0.61. Same results are obtained when considering the set of all models studied in Figure 10b.

Moreover, some visual results are presented in Figure 11. We can see that some of our viewpoints, in (a), are exactly the same, unlike those of the other two geometric methods. Thanks to our purely geometric method, we were able to find a viewpoint chosen with subjective bias. In (b), we wanted to show the cases where our viewpoint and that of the users differ. For example, in the first line with *Bimba* model, our viewpoint may be considered less aesthetically pleasing, but this is intentional, as it highlights the facial features, particularly the eyes. Secondly, the point of view chosen by users for the mechanical piece in the second line is an accidental one, so the view determined by the three geometric methods removes this perspective ambiguity. Finally, the selected point of view by the users for the man with arms crossed is also aesthetic and specific to human culture whereas our point of view better highlights the details of the position: the crossed arms and the face.

7 CONCLUSION

We have presented a new measure to select best viewpoint of a 3D object. The advantage of this contribution relies on the fact that we consider both aspects (saliency and visibility). Being view-dependent leads to a more realistic saliency score. Then we have conducted an original and intensive evaluation to better study the interest of the proposition. Compared to reference approaches, our method selects the viewpoints that are the most similar to user selection. Our visual analysis also highlights that when our approach differs from the user study, it still proposes and interesting view (maybe free from human biases).

In future work, we plan to work with textured models in order to find the most relevant views for a textured object, and close the gap with the viewpoints defined by images.

REFERENCES

An, G. and *et al.* (2016). Mesh simplification using hybrid saliency. In *IEEE International Conference on Cyberworlds*.

Blanz, V. and Tarr, M. (1999). What object attributes determine canonical views? *Perception*.

Chave, A. (2015). A note about gaussian statistics on a sphere. *Geophysical Journal International*.

Fabricatore, C. and *et al.* (2002). Playability in action videogames: A qualitative design model. *Human-Computer Interaction*.

Feixas, M. and *et al.* (2009). A unified information-theoretic framework for viewpoint selection and mesh saliency. *ACM Trans. Appl. Percept.*

Habibi, Z. and *et al.* (2015). Good feature for framing: Saliency-based gaussian mixture. In *IEEE Int. Conf. Intellig. Rob. Sys.*

Jeong, S. and Sim, J. (2017). Saliency detection for 3d surface geometry using semi-regular meshes. *Trans. Multimedia*.

Lavoué, G. and *et al.* (2018). Visual attention for rendered 3d shapes. In *Computer Graphics Forum*.

Lee, C. and *et al.* (2005). Mesh saliency. In *ACM SIGGRAPH*.

Leifman, G. and *et al.* (2016). Surface regions of interest for viewpoint selection. *PAMI*.

Limper, M. and *et al.* (2016). Mesh saliency analysis via local curvature entropy. In *Eurographics*.

Meyer, M. and *et al.* (2003). Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics*.

Meynet, G. and *et al.* (2019). Pc-msdm: A quality metric for 3d point clouds. In *Conference on Quality of Multimedia Experience*.

Miao, Y. and Feng, J. (2010). Perceptual-saliency extremum lines for 3d shape illustration. *The Visual Computer*.

Nehmé, Y. and *et al.* (2023). Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric. *ACM TOG*.

Nouri, A. and *et al.* (2015). Multi-scale mesh saliency with local adaptive patches for viewpoint selection. *Signal Processing: Image Communication*.

Plemenos, D. and Sokolov, D. (2006). Intelligent scene display and exploration. In *International Conference GraphiCon*.

Plemenos, D. and *et al.* (2004). On viewpoint complexity of 3d scenes. In *International Conference GraphiCon*.

Rudoy, D. and Zelnik-Manor, L. (2012). Viewpoint selection for human actions. *IJCV*.

Rusinkiewicz, S. (2004). Estimating curvatures and their derivatives on triangle meshes. In *IEEE Symposium on 3D Data Processing, Visualization and Transmission*.

Secord, A. and *et al.* (2011). Perceptual models of viewpoint preference. *ACM TOG*.

Sokolov, D. and Plemenos, D. (2005). Viewpoint quality and scene understanding. In *Eurographics Symposium on Virtual Reality*.

Song, R. and *et al.* (2014). Mesh saliency via spectral processing. *ACM TOG*.

Tasse, F. and *et al.* (2015). Cluster-based point set saliency. In *ICCV*.

Taubin, G. (1995). Estimating the tensor of curvature of a surface from a polyhedral approximation. In *ICCV*.

Vázquez, P. and *et al.* (2001). Viewpoint selection using viewpoint entropy. In *Vis., Model., and Visualizat.*

Wang, S. and *et al.* (2015). Multi-scale mesh saliency based on low-rank and sparse analysis in shape feature space. *Computer Aided Geometric Design*.

Wu, J. and *et al.* (2013). Mesh saliency with global rarity. *Graphical Models*.

Xiang, Y. and *et al.* (2016). Objectnet3d: A large scale database for 3d object recognition. In *ECCV*.

APPENDIX

Computing Visible Faces and Vertices. To know which vertices are visible, first we have determined which faces are visibles. Given a Point of View *pov*, we determine which faces are facing the camera using back face culling. More precisely, a face F is oriented towards the camera if the cosine of the angle α_F between its outgoing normal $\overrightarrow{n_F}$ and the camera vector $\overrightarrow{pov - c_F}$, with c_F the center of F, is greater than an epsilon $\varepsilon = 10^{-5}$. Some of these faces can be occluded. To filter them out, we use the depth information contained in the depth maps available for each viewpoint. A face is considered visible if the depth associated with the 2D projection of its center (we take the barycenter) is the same as that contained in the depth map. Often, the 2D coordinates of the centers are not integer. In the following we will

consider: $i_b = \lfloor y_{cF} \rfloor$, $j_b = \lfloor x_{cF} \rfloor$, $i_a = i_b + 1$ and $j_a = j_b + 1$.

To determine the exact depth that lies in the depth map at the center's position, we perform three interpolations. Specifically, given a face F, $c_F = (x_{c_F}, y_{c_F})$ are the 2D coordinates of its center and z_{c_F} its depth after the 2D projection. Then the depth z_D associated with (x_{c_F}, y_{c_F}) in the depth map D is calculated as bi-linear interpolation:

- z_b is determined with a interpolation between (i_b, j_b) and (i_a, j_b)
- z_a is determined with a interpolation between (i_b, j_a) and (i_a, j_a),
- z_D is determined with a interpolation between z_b and z_a.

If $|z_D - z_{c_F}| \leq 10^{-2}$, then F is visible.

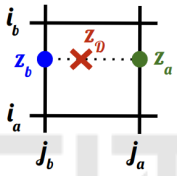


Figure 12: Depth z_D associated with the non-integer 2D coordinates of $c_F = (x_{c_F}, y_{c_F})$

Then, we identify which vertices are visibles. Given a Point of View *pov*, we consider a vertex to be facing the camera if at least one of its adjacent faces is visible. This allows us to perform an initial filtering and to speed up the search process. Then, as for faces, some vertices may be occluded. We first compute for each vertex v, its normal $\overrightarrow{n_v}$ as the unit vector with the same direction as the sum of normals of the face containing the vertex v. To remove vertices, we then calculate for each of the previously selected v vertices, the angle α_v between their outgoing normals $\overrightarrow{n_v}$ and the camera vector $\overrightarrow{pov - v}$. Only those vertices v whose $cos(\alpha_v) > 0$ are kept.

User Study Details - Camera Positions. Camera positions $C_{i,j}$ on the sphere centered on a 3D model are defined by two angles (α_j, β_i) as illustrated in Figure 3b: $\forall i \in [0,7]$, $\forall j \in [0,4]$

$$C_{i,j} = \begin{cases} x_{ij} = R\cos(\alpha_j)\cos(\beta_i) \\ y_{ij} = R\sin(\alpha_j) \\ z_{ij} = R\cos(\alpha_j)\sin(\beta_i) \end{cases} \quad (4)$$

with $R = 2.2$, $\beta_i = i.\pi/4$ and $\alpha_j = (2-j).\pi/4$.

When $|\alpha_j| = \pi/2 \Rightarrow \cos(\alpha_j) = 0$, we obtain the same camera position eight times. Only one of them is kept because we always have exactly the same view.

User Study Details - Participants. Following the observations of (Nehmé and *et al.* , 2023), to recruit reliable users, we used the crowd-sourcing service offered by the Prolific[6] site. With this platform, 203 users completed our study: 121 males and 82 females. They are aged between 19 and 71. To filter out the less serious users, we have evaluated three different parameters per user: the total time to complete the study, the time to treat one mesh and the average navigation complexity (the average percentage of viewpoints visited per mesh). For each parameter p, we have computed statistical data : the lower quartile Q_1^p and the interquartile range IQR^p. To identify outliers, we use the IQR method to set up an aera outside of Q_1. If a user has one of these three parameters smaller than $Q_1^p - (1.5 \times IQR^p)$, then the user is removed. According to these conditions, only one female user has been retired.

User Study Details - Post processing. The goal is to determine which viewpoints have been most selected by users. As mentioned previously, in the user study, there are 26 study viewpoints, noted POV_k, with $k \in [1,26]$. Given a model, a user u must select 3 ordered viewpoints : $pov_{u,v}$, with $v \in [1,3]$. Each selected viewpoint $pov_{u,v}$ has an impact on the study viewpoints POV_k. In other words, each of $pov_{u,v}$ assigns a weight $p^k_{u,v}$ to each POV_k. These weights are calculated with a 3D normal distribution (Chave, 2015), centered in $pov_{u,v}$ and with support $\sigma = 0.58$. As the parametrization of camera positions is not regular in the study, a $pov_{u,v}$ can have an impact on several POV_k. Moreover, most of these weights $p^k_{u,v}$ are zero, because $pov_{u,v}$ have a local impact. Next, a normalization is applied to ensure that each choice $pov_{u,v}$ has the same importance. To this purpose, the sum of the weights from $pov_{u,v}$ is normalized to 1. Finally, to calculate the score for each POV_k, we simply sum the weights $p^k_{u,v}$ weighted by a factor $(4-v)$. Thus, the weights of viewpoints chosen first ($v = 1$) will be favoured over the weights of viewpoints chosen last ($v = 3$). The formula is :

$$\forall k \in [1,26],\ Weight_k = \sum_{u=1}^{N} \sum_{v=1}^{3} (4-v).p^k_{u,v} \quad (5)$$

with N the number of users.

[6]https://www.prolific.co/