TOWARDS VISUAL SALIENCY COMPUTATION ON 3D GRAPHICAL CONTENTS FOR INTERACTIVE VISUALIZATION

Mona Abid, Matthieu Perreira Da Silva, Patrick Le Callet

Nantes University, LS2N, CNRS, UMR 6004, Nantes, France.

ABSTRACT

Understanding human visual attention mechanisms and interaction in immersive scenes are of great importance in perception. In immersive context, users are able to interact with increasingly rich/ complex 3D contents during rendering. Therefore, to avoid latency or rendering issues, there is a critical need for simplifying and filtering the primitives and levels of detail of these high-quality 3D graphics (according to viewing conditions). In order to ensure a high user's quality of experience (QoE) during interactive visualization, these processing operations should take into account perceptual information. To do so, we suggest an approach that uses visual saliency information of the 3D scene to guide simplification and level of details selection. In this paper, we question the efficiency of our novel approach to compute visual saliency on 3D graphics. This approach takes into consideration the viewpoint from which the 3D content was seen/rendered when computing saliency (whichever the considered viewpoint is), by using saliency maps of view-based method computed offline. Such technique could help alleviate rendering constraints during interactive visualization.

Index Terms— Visual attention, 3D contents, interactive visualization, 2D-3D projection, saliency

1. INTRODUCTION

Three-dimensional (3D) graphics are commonplace in many applications such as digital entertainment, cultural heritage, architecture, and scientific simulation. These data are increasingly rich and detailed; as a complex 3D scene may contain millions of geometric primitives, enriched with various appearance attributes such as texture maps designed to produce a realistic material appearance.

The way of consuming and visualizing this 3D content is nowadays evolving from standard screens to Virtual and Mixed Reality (VR/MR) and possibly via the network. However, the visualization and interaction with such large and complex data remain an unresolved issue due to strong latency and rendering problems that could be encountered, especially when the 3D content is stored on remote servers; as it is streamed on the display device. Therefore, there is a critical need to compress and simplify these high-quality 3D contents while ensuring an optimal user's quality of experience (QoE). Taking advantage of the visual attention/ saliency information is a convenient way to drive these processing operations.

Visual saliency is an important feature of the human visual system, it describes visual attention distribution or eye movements for a given scene based on human perception [1]. Detecting such visually salient regions is fundamental to preserve them during simplification or compression. Thanks to its efficiency to ensure a good perceived quality of visual data, visual attention has been a significant component over the last two decades in computer vision community, leading to a wide range of visual attention models mainly dedicated to 2D images [2, 3] and videos [4, 5]. These saliency models were validated using fixation maps obtained from eye-tracking experiments

Moreover, fewer works dedicated to stereoscopic visual attention have been proposed, both for images and videos [6, 7, 8, 9]. These works take into account 3D through depth perception from stereoscopic disparity, but do not deal with 3D meshes.

However, the computer graphics community has also explored the modeling of visual attention on 3D objects,in particular the saliency of 3D meshes [10, 11, 12, 13] and only few papers handle point sets [14, 15, 16]. These approaches operate directly on 3D data and take into account the geometry of the scene but are view-independent. They are relevant for many application such as mesh simplification, viewpoint selection [10], directing users attention [17], etc. However, they do not take into account the way the scene is rendered (i.e. viewing conditions; point and field of view, viewing angle, object occlusion, light, etc). The study of [18] on printed 3D objects, affirms the need, at least, of the orientation of the object towards the observer to predict saliency.

To overcome this lack of appearance attributes and viewing conditions, we suggest a novel approach to compute visual saliency on 3D graphics. It relies on image-based rendering algorithms.

This work was funded by the French National Research Agency as part of ANR-PISCo project (ANR-17-CE33-0005).

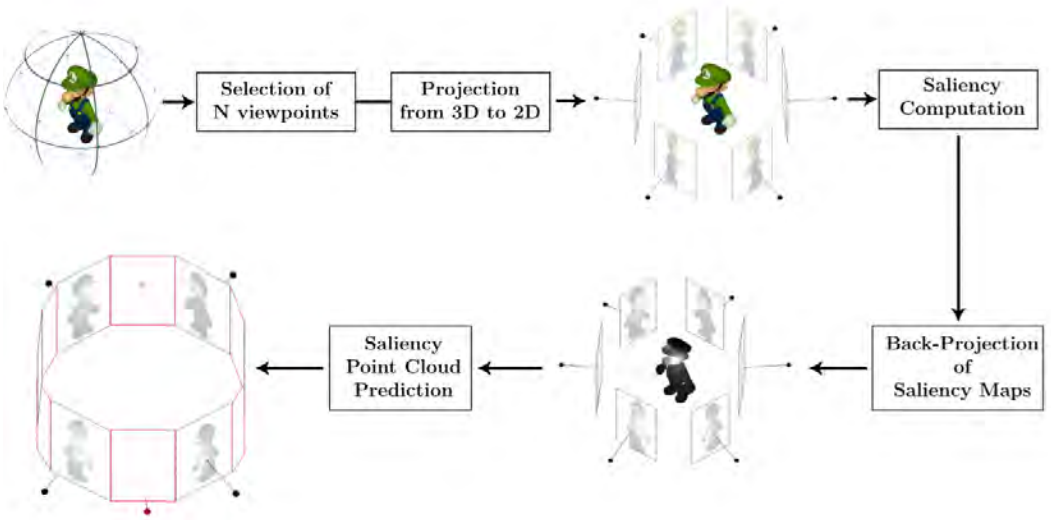

Fig. 1. Simplified illustration of the proposed framework.

2. THE PROPOSED METHOD - FRAMEWORK

In this paper, we focus on saliency detection for high-quality 3D graphics including different 3D data representations:
- 3D meshes which are represented as a set of vertices in 3D space and a connectivity list that describes how each vertex is connected to each other.
- Point clouds and colored vertices which are represented as a set of unorganized points in the 3D space.

The question we tried to answer is: Can we compute saliency of 3D graphical contents by taking into account the viewpoint from which the 3D content is rendered (i.e viewpoint-aware approach)?

2.1. Proposed framework

In this section, the key steps of the view-dependent scheme are presented.
Given a 3D object, we select N viewpoints and we render the 3D object under these different viewpoints leading to N views generation. These views have all been rendered under perspective projection using *Blender* rendering engine.
Let's consider a 3D-scene $S^{(3)} \in R^3$ consisting of the points $p = \{x, y, z\} \in S^{(3)}$. Rendering the scene projects all points of $S^{(3)}$ to the image plane: $F(S^{(3)}) = S^{(2)}, F(\{x, y, z\}) = \{\mathbf{x},\mathbf{y}\} \in R^2$. As we know the parameters of the used virtual camera for rendering, we are able to compute the perspective projection matrix as well as the model view matrix. Hence, it is possible to find the function **G** which back-projects 2D pixel coordinates to the 3D space $G(\{\mathbf{x},\mathbf{y}\}) = \{x, y, z\} \in S^{(3)}$ and results in partial 3D point clouds (i.e. set of point cloud which depends on the rendered view). For the 2D-3D projection, we apply the ray cast by considering the adequate ray origin and direction; If the ray hits the 3D graphical object the 2d viewport pixel and the corresponding 3d coordinates as well as the color are stored in a structured data format.
Note that the resolution of the partial point cloud is exclusively related to the resolution of the 2D rendered view.

Afterwards, we apply a 2D view-based saliency model on the 2D projections of the 3D object. In this work, we considered Salicon model [19] as it showed the highest performances when computed on computer generated contents [20]. The 2D saliency map represents a probability map, and the saliency value at each location indicates the chances of how likely people paying attention there. Pixel-wise saliency values are stored in a partial point cloud structure by considering 2D-3D correspondence. This step enables saliency information to be stored efficiently since it could be accessed as easily as the color information.
To predict the saliency of an *intermediate view*, we use at least 2 neighbour partial point clouds. Let PC_i, PC_j be 2 neighbour partial point clouds (with 3D points coordinates and RGB color information) and PC_i^s, PC_i^s their corresponding saliency information saved in a point cloud structure (that we call point cloud saliency). In this paper, we consider as a first step the middle view as the *intermediate view* of the PC_i and PC_j. Let PC' be the partial point cloud corresponding to the intermediate view of which we want to predict saliency. We first compute the correspondence set $K_{(PC',PC_i)}$ between PC' and and PC_i as well as the correspondence set $K_{(PC',PC_j)}$ between PC' and PC_j based on color information [21]. For each point in the partial point cloud PC', we find the corresponding points in the partial point clouds PC_i and PC_j respectively based on distance related to the resolution of the point cloud. We define the saliency contribution of each view based on the correspondence sets $K_{(PC',PC_i)}$ and $K_{(PC',PC_j)}$. To obtain $PC^{s\prime}$; point cloud saliency of PC', we combine saliency contributions from PC_i^s and PC_j^s by applying a weighted average. We call this step saliency interpolation. The weight values were determined based on the viewing angle deviation between PC' and PC_i,PC_j respectively.
To validate our approach, we conduct a proof-of-concept by comparing the interpolated saliency point cloud PC'^s with saliency obtained from eye-tracking experiment. As demonstrated in the study by [11], it is extremely difficult to obtain a ground-truth for 3D contents because saliency is strongly

Fig. 2. Illustration of one view of the 3D graphic contents and its corresponding saliency collected from human fixations; also called saliency ground-truth.

related to the viewpoint from which the content is rendered (i.e. perceived by the observer). In fact, the database obtained by [11] is a pseudo-ground truth because observers were asked to select important points that could be chosen by others, during the experiment. This takes the form of a task provided to the observer influencing the detection of saliency because everyone has a different definition of what it means to be perceptually important. In this paper, we build a ground-truth that is viewpoint-aware by conducting a psycho-visual eye-tracking experiment in order to effectively validate our approach.

2.2. Eye-tracking experiment

One possible way of rendering 3D graphical content relies on image-based rendering. It is convenient in this context as it enables to control the viewing conditions (viewing angle, viewing distance, visual acuity, etc) and ensures the repeatability of the eye-tracking experiment and thus apply statistics on the collected data in order to build ground-truth (by aggregating gaze data).

2.2.1. Stimuli generation

In this work, we selected twenty-two high resolution 3D graphical objects that belong to very different semantic categories (objects, human figures, animals, art, characters) and have different shapes (occupancy of the object; horizontally, vertically, etc). Since each 3D object has different visual information based on the viewpoint from which it is rendered, we considered 4 views corresponding to 4 faces of a cube This yields to a total of $22 \times 4 = 88$ rendered images.

2.2.2. Apparatus and participants

We used the *EyeLink100 Plus* eye-tracking device in the remote mode (i.e head free-to-move). This device allows binocular tracking and reports a spatial accuracy between the visual angle range of 0.25 and 0.50 degrees.
The distance between the observer and the experimental display which is a computer monitor display with full HD resolution (1920 x 1080) was approximately 110 cm. Based on this setting, there were 64 pixels per degree of visual angle in each dimension, and the display resolution was about 30×17 visual degrees.
34 observers participated in the eye-tracking experiment. They all had a normal or corrected to normal vision. The total time of the experiment was 15 minutes including vision check, calibration, and 88 images visualization for 3 seconds each. The experiment was split into 3 sessions to check if the calibration is always valid. Four observers were removed from the experiment due to the presence of too much invalid data (e.g. eyes not looking at the screen or blinking too often).

3. RESULTS

In this section, we present the results of saliency obtained from the proposed approach (cf. section 2).
To evaluate the reliability of the proposed approach we consider the following steps:
1. We compare the interpolated saliency information of the *intermediate view* with the corresponding ground-truth obtained from the aggregation of human fixations.
2. We apply Salicon model [19] on the same projected *intermediate view* of the 3D content then we project the saliency on the point cloud structure to have the same data format (i.e. partial point cloud).
3. We finally compare the obtained scores as summarized in Table I.
The so-called ground truth (GT) obtained from the eye-tracking experiment is illustrated in figure 2.

Metric	KLD	CC	Sim
Our approach	0.46 ± 0.33	0.68 ± 0.07	0.51 ± 0.2
2D Salicon model	0.43 ± 0.25	0.67 ± 0.07	0.50 ± 0.2
Correlation coefficient	0.76	0.70	0.71

Table 1. Metrics evaluation for our presented approach and the 2D Salicon model

We chose to represent the most informative view of the 3D object in this illustration. As there is no metric, in literature, that allows the comparison of two point clouds by taking the color information into consideration, we considered metrics [22] that are used on 2D saliency maps such as Kullback-Divergence (KLD), Pearson's correlation coefficient (CC) and similarity (Sim). KLD computes the divergence between two distributions (the lower the better) whereas CC and Sim compute the correspondence between two distributions (the higher the better). Since saliency values are significant only on the informative part of the partial point cloud, we apply different metrics only on that mask wrapping the visual content. For the 3 metrics, we computed the mean value of the 88 rendered views and the standard deviation.
We summarize the preliminary results in table 1. Based on the presented results, the mean metric's values are very similar. As contents are quite different, a complementary statistical analysis is needed to ensure that the mean value is reliable to draw solid conclusions. To do so, we computed the correlation coefficient between the metric scores of both methods (i.e. our approach and the 2D Salicon model). The obtained correlation coefficients are presented in table 1. Assuming a normal distribution, we also compute the p-value which tests the hypothesis that there is no relation between observed phenomena (i.e. null hypothesis). P-values range from 0 to 1, where values close to 0 correspond to a significant correlation and a low probability of observing the null hypothesis. We obtained ***p-values* = 0** for the 3 metrics which indicated that the correlation is significant. The proposed approach is therefore validated.

4. DISCUSSION

Once statistical analysis conducted, the results suggest that the proposed approach appears to be an effective and convenient way to compute saliency on 3D graphical objects since it takes into consideration the viewpoint from which the 3D content was rendered. Such technique seems promising to fill the gap between computer vision and computer graphics communities. However, some limitations are worth noting. Although the interpolation takes into account the angular deviation, if the used neighbor views contribute with irrelevant saliency information, the interpolation result will also be irrelevant. In fact, interpolation results depend on the used 2D saliency model to generate different saliency maps view offline. For this reason, we should evaluate different saliency models and consider the most-performing one once applied on computer-generated contents. We could also, if enough data are available, fine-tune the 2D computational model on ad-hoc contents. In figure 3, we show an example that illustrates irrelevant saliency information. Since the 2D saliency model was trained on natural images, people in the photos look to the camera, therefore the prediction resulting from our approach (cf. fig. 3 first row, column (c)) is coherent with the ground truth (cf. fig. 3 first row, column (b)); as the character face is visible (cf. fig.3 first row, column (a)). Whereas, the second row in figure 3 shows a mismatch between the predicted result and the ground truth. The reason behind this irrelevance is that saliency information obtained from neighbor views indicate that the feet are salient (which is not coherent with the ground-truth).

Fig. 3. Illustration of 6 partial point clouds, each line represents: (a) the rendered view, (b) the ground-truth saliency information and (c) the resulted saliency based on interpolation technique.

5. CONCLUSION

In this paper, we suggested a novel approach to compute visual saliency on colored 3D graphics in the context of interactive visualization. As saliency is related to the perceived visual information, our hybrid approach takes into consideration the viewpoint from which the 3D content was rendered. It aims to predict the unseen views of a 3D object by effectively interpolating pre-computed saliency information. To evaluate the proposed approach, we conducted an eye-tracking experiment and compared the obtained gaze data with saliency information resulting from our approach. Our results suggest that this approach appears to be effective and promising to alleviate rendering constraints during interactive visualization and therefore optimize QoE based rendering of 3D graphical contents.

6. REFERENCES

[1] Christof Koch and Shimon Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of intelligence*, pp. 115–141. Springer, 1987.

[2] A. Borji, "Saliency prediction in the deep learning era: Successes and limitations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[3] Tilke Judd, Krista A. Ehinger, Frédo Durand, and Antonio Torralba, "Learning to predict where humans look," *2009 IEEE 12th International Conference on Computer Vision*, pp. 2106–2113, 2009.

[4] Wenguan Wang, Jianbing Shen, Jianwen Xie, Ming-Ming Cheng, Haibin Ling, and Ali Borji, "Revisiting video saliency prediction in the deep learning era," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[5] Xiaodi Hou and Liqing Zhang, "Dynamic visual attention: Searching for coding length increments," in *Advances in neural information processing systems*, 2009, pp. 681–688.

[6] Junle Wang, Matthieu Perreira Da Silva, Patrick Le Callet, and Vincent Ricordel, "Computational model of stereoscopic 3d visual saliency," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2151–2165, 2013.

[7] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin, "Saliency detection for stereoscopic images," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2625–2636, 2014.

[8] Haksub Kim, Sanghoon Lee, and Alan Conrad Bovik, "Saliency prediction on stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1476–1490, 2014.

[9] Yuming Fang, Chi Zhang, Jing Li, Jianjun Lei, Matthieu Perreira Da Silva, and Patrick Le Callet, "Visual attention modeling for stereoscopic video: a benchmark and computational model," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4684–4696, 2017.

[10] Chang Ha Lee, Amitabh Varshney, and David W Jacobs, "Mesh saliency," in *ACM SIGGRAPH 2005 Papers*, pp. 659–666. 2005.

[11] Xiaobai Chen, Abulhair Saparov, Bill Pang, and Thomas Funkhouser, "Schelling points on 3d surface meshes," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, pp. 1–12, 2012.

[12] Ran Song, Yonghuai Liu, Ralph R Martin, and Paul L Rosin, "Mesh saliency via spectral processing," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 1, pp. 1–17, 2014.

[13] George Leifman, Elizabeth Shtrom, and Ayellet Tal, "Surface regions of interest for viewpoint selection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 414–421.

[14] Xiaoying Ding, Weisi Lin, Zhenzhong Chen, and Xinfeng Zhang, "Point cloud saliency detection by local and global feature fusion," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5379–5393, 2019.

[15] Elizabeth Shtrom, George Leifman, and Ayellet Tal, "Saliency detection in large point sets," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 3591–3598.

[16] Oytun Akman and Pieter Jonker, "Computing saliency map from spatial information in point cloud data," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2010, pp. 290–299.

[17] Gunhee Kim, Daniel Huber, and Martial Hebert, "Segmentation of salient regions in outdoor scenes using imagery and 3-d data," in *2008 IEEE Workshop on Applications of Computer Vision*. IEEE, 2008, pp. 1–8.

[18] Xi Wang, David Lindlbauer, Christian Lessig, Marianne Maertens, and Marc Alexa, "Measuring the visual salience of 3d printed objects," *IEEE computer graphics and applications*, vol. 36, no. 4, pp. 46–55, 2016.

[19] Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao, "Salicon: Saliency in context," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1072–1080.

[20] M. Abid, M. P. Da Silva, and P. L. Callet, "On the usage of visual saliency models for computer generated objects," in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, Sep. 2019, pp. 1–5.

[21] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun, "Colored point cloud registration revisited," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 143–152.

[22] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand, "What do different evaluation metrics tell us about saliency models?," *arXiv preprint arXiv:1604.03605*, 2016.