Enhancing Immersive Experiences through 3D Point Cloud Analysis: A Novel Framework for Applying 2D Visual Saliency Models to 3D Point Clouds

Marouane Tliba, Xuemei Zhou, Irene Viola, Pablo Cesar, Aladine Chetouani, Giuseppe Valenzise, Frédéric Dufaux

Enhancing Immersive Experiences through 3D Point Cloud Analysis: A Novel Framework for Applying 2D Visual Saliency Models to 3D Point Clouds

Marouane Tliba
PRISME, University of Orleans
Centrum Wiskunde & Informatica

Xuemei Zhou
Centrum Wiskunde & Informatica
TU Delft

Irene Viola
Centrum Wiskunde & Informatica

Pablo Cesar
Centrum Wiskunde & Informatica
TU Delft

Aladine Chetouani
PRISME, University of Orleans

Giuseppe Valenzise
CNRS, CentraleSupélec, L2S

Frederic Dufaux
CNRS, CentraleSupélec, L2S

***Abstract*—In the new area of immersive multimedia environments, understanding and manipulating visual attention are crucial for enhancing user experience. This study introduces an innovative framework that extends traditional 2D saliency maps to the analysis of 3D point clouds, a step forward in adapting saliency prediction to more complex and immersive environments. Our framework centers on the orthographic projection of 3D point clouds onto 2D planes, enabling the application of established 2D saliency models to this novel context. We further delve into the evaluation of these models on a 3D point cloud eye-tracking dataset, exploring various projection settings and thresholding techniques to maintain the integrity of saliency information in the transition from 2D to 3D. This research not only bridges a gap in applying visual attention models to 3D data but also offers insights into the optimization of quality of experience in immersive multimedia systems.**

***Index Terms*—Visual Saliency, 3D point cloud, Quality of experience, 3D Projection**

I. Introduction

Immersive multimedia environments, underpinned by advancements in virtual and augmented reality technologies [1], [2], are revolutionizing user experiences across a spectrum of fields [3], [4]. Central to this transformation is the adoption of three-dimensional (3D) point clouds, which offer unparalleled precision and detail in capturing real-world scenes [5]. Unlike traditional 3D data formats, point clouds excel in representing intricate features and complexities of physical environments with high fidelity [6]. This capability has positioned point clouds as a foundational element in developing interactive 3D models for diverse applications, including virtual and augmented reality [7], [8], robotics [9], and computer-aided design [10]. A critical aspect of delivering superior immersive experiences involves accurately modeling the visual attentive behavior of users toward 3D point clouds within transmission systems [11]. Human observers inherently prioritize certain areas within their Field-of-View (FoV), concentrating on regions of interest while disregarding others [12], [13]. These selective processes so-called attention mechanisms enable individuals to efficiently interpret and comprehend complex scenes by allocating their limited perceptual and cognitive resources towards the most relevant segments of sensory input [14]. Drawing inspiration from this phenomenon, saliency prediction or modeling seeks to emulate human gaze fixation patterns across visual content [15]. Culminating in the generation of saliency maps, each element within these maps quantifies the likelihood of attracting user attention, providing a quantifiable measure of visual prominence [16]. Therefore, optimizing visual attention in 3D point clouds is key to improving Quality of Experience (QoE) [17], by focusing on regions of interest. This approach reduces latency and enhances user satisfaction by efficiently utilizing computational resources, thereby elevating the immersive experience. Despite significant progress in 2D Saliency or Visual attention modeling enabled by expansive datasets, the development of 3D saliency models is constrained by the limited size of available datasets [18]. Current research primarily focuses on understanding human behavior and interactivity in immersive environments [19], neglecting the creation of large-scale datasets for comprehensively modeling human attention in 3D spaces. This gap impedes direct modeling of visual saliency on 3D data. This research addresses a critical need in the domain of 3D point cloud visual saliency by presenting a comprehensive framework specifically engineered to broaden the utility of 2D visual saliency models for 3D point clouds. Our primary contributions are:

1) **Innovative Framework for 3D Point Cloud Projection and Reconstruction:** We present a novel framework designed to extend the application of 2D visual saliency models to 3D point clouds, facilitating the orthographic projection of 3D data onto 2D planes while preserving

essential information for accurate 3D reconstruction. This approach enables the effective application of 2D saliency models to 3D environments, supporting modifications of associated features in the 2D domain without compromising the original 3D spatial information. The framework's capability to maintain the integrity of geometrical data while allowing for feature adjustments highlights its potential for a broad spectrum of applications, from enhancing visual attention analysis to improving immersive environment design. Our framework is available on : github.com/mtliba/qomex24

2) **Comprehensive Benchmarking and Performance Evaluation:** Alongside the framework, we conduct an extensive benchmarking study to evaluate the effectiveness of 2D saliency models when applied to 3D point clouds. Utilizing a specialized 3D point cloud eye-tracking dataset, our study investigates various projection and thresholding settings to determine optimal strategies for preserving saliency information.

The significance of our work lies in its potential to advance the design and optimization of immersive multimedia systems, by providing a more nuanced understanding of modeling visual attention in 3D environments.

II. Related Work

This section reviews the advancements in 2D saliency modeling, discusses the challenges of 3D saliency modeling for point clouds, and highlights the efforts toward understanding visual attention in immersive environments.

A. 2D Saliency Modeling

The advent of deep learning, particularly Convolutional Neural Networks (CNNs) [20], has significantly propelled advancements in 2D saliency modeling [21]–[23]. These models leverage extensive datasets to achieve remarkable accuracy in predicting human attention in 2D images. Noteworthy among these are the SALICON dataset [21], which is one of the largest saliency datasets, and the MIT300 benchmark [22], which has been instrumental in evaluating the performance of saliency models, over well representative set of metrics. Such benchmarks have facilitated the development of more sophisticated saliency prediction models, and provide a solid platform for comparative analysis [22]. Despite these advancements, the direct application of the same theory behind these 2D models to 3D point cloud data has remained a challenge, primarily due to the intrinsic characteristics of 3D spatial data, and the lack of analogous comprehensive datasets for 3D point clouds.

B. Visual Attention in Immersive Environments

Research on visual attention in immersive environments, such as VR/AR, has begun to shed light on how users interact with and perceive 3D spaces [24], [25]. To this end, eye-tracking experiments are pivotal for understanding human visual behavior in 3D contexts, showing where the eyes are looking within the 3D scene. Notable studies include Sitzmann *et al.* [26], who analyzed gaze and head orientation in stereoscopic panoramas, revealing the inadequacy of existing saliency predictors in VR settings. Nguyen *et al.* [27]. introduced a significant saliency dataset for 360-degree videos, employing a methodology grounded in psychology and compatible with Head-Mounted Displays (HMDs). Lavoué *et al.*. presented a dataset capturing eye movements for 3D shapes under various conditions, while Ding *et al.* [28] offered a 3D colored mesh saliency dataset based on an eye-tracking experiment alongside a saliency detection framework focusing on color and geometric features. Alexiou *et al.* [29] conducted an eye-tracking experiment with static point clouds in a 3D scene allowing 6 Degrees-of-Freedom (DoF), introducing a method for exploiting gaze measurements to identify areas of fixation within a point cloud. Zhou *et al.* [30] proposed a dynamic point cloud dataset alongside an eye-tracking experiment and a quality score in 6 DoF. The experiment aimed to evaluate the perceptual quality of dynamic point clouds under various distortion levels and to analyze the variation in visual attention corresponding to the distortion level.

C. Challenges in Current Research

Conventional 2D saliency modeling cannot be directly applied to 3D point clouds. Furthermore, the lack of large-scale datasets for training and evaluating 3D saliency models hampers progress in this field. Hence, there is a pressing need for a comprehensive framework capable of accurately projecting, reconstructing, and empirically comparing 2D saliency maps onto 3D point clouds. This study introduces such a framework, aiming to bridge the gap between 2D saliency modeling and 3D point cloud analysis. By extending the application of 2D saliency models to 3D point clouds and devising a method for precise 3D reconstruction from 2D predictions, our research endeavors to enhance the prediction and analysis of visual attention in immersive environments. Our study also establishes an initial benchmark for applying 2D saliency models to 3D point clouds, accounting for certain critical aspects: viewpoints selection and saliency thresholding, thereby offering a foundation for future research in enhancing immersive visual experiences.

III. 3D Point Cloud Projection and Saliency Mapping Framework

In this part, we introduce a comprehensive framework designed for analyzing 3D point clouds through the lens of human visual perception. Our framework encompasses a sequence of operations starting from the preprocessing of the 3D point cloud, applying the projection of 3D points onto a 2D plane, followed by the application of a computational saliency model, and culminating in the reconstruction of a 3D saliency map. The overall workflow of our framework is described in figur (1).

A. Preprocessing

1) View Selection: To align the input with desired viewing perspectives, the original point cloud P_o undergoes rotations

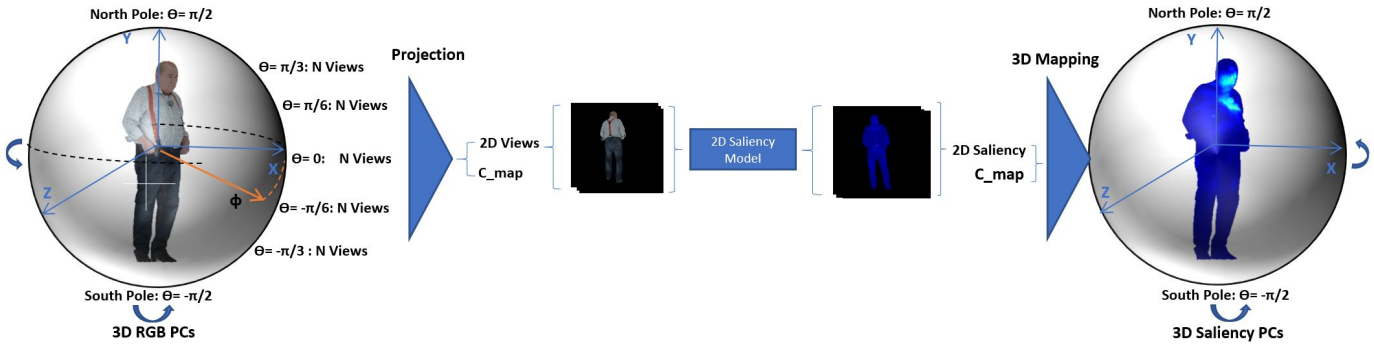

Fig. 1. A High-Level Overview of Our Proposed Framework: Illustrating the Core Workflow and Key Components

R around the principal axes (x, y, z), adjusting the orientation in 3D space. This process is mathematically represented as:

$$P = R_{(x,y,z)}(\theta, \Phi, \psi)P_o,$$

where P is the rotated point cloud, (θ, Φ, ψ) symbolizes the combined rotation angles for axes x, y, and z, respectively.

2) Point Cloud Centering and Normalization: This phase involves normalizing the point cloud to fit within a predefined view volume, facilitating a more coherent and effective visualization and analysis. The step is mathematically represented as:

$$P_{\text{norm}} = \frac{P - \mu(P)}{\max(P) - \min(P)}$$

where P_{norm} is the normalized point cloud, and $\mu(P)$ denotes the mean position of the point cloud.

B. 3D-to-2D Projection

1) Point-Wise Orthographic Projection: The first stage involves transforming 3D point cloud data into a 2D representation. Our approach builds upon the foundational principles of orthographic projection [31] but extends it significantly to cater to the demands of 3D point cloud analysis. Orthographic projection traditionally involves mapping the points from a 3D space onto a 2D plane without considering perspective effects, thus maintaining uniform scale irrespective of depth. To do so we applied a point-wise orthographic projection while keeping into account the perspective parameters.

Given the preprocessed point cloud P_{norm} with points $p_i = (x_i, y_i, z_i)$, accompanied with a set of camera parameters as follows: ($'left'$:l, $'right'$: r, $'top'$: t, $'bottom'$:b). These parameters ensure the accurate 2D mapping of 3D point within the same specified projection volume, where $'left'$, $'right'$, $'top'$, and $'bottom'$ outline the viewport's horizontal and vertical bounds. Finally, the projection onto a 2D plane is defined as $p_i = (x_i, y_i)$. This step ignores the z dimension, focusing on x and y coordinates.

2) scaling and translation: This step is applied to adjust the projected points to fit the image dimensions based on the camera parameters, using:

$$\text{scale} = \begin{bmatrix} \frac{\text{image\_height}}{r-l} \\ \frac{\text{image\_width}}{t-b} \end{bmatrix}, \quad \text{offset} = \begin{bmatrix} \text{image\_height}/2 \\ \text{image\_width}/2 \end{bmatrix}$$

$$p_{i,\text{scaled}} = p_i \times \text{scale} + \text{offset}$$

3) Z-Buffering: To achieve the realistic projection of 3D data onto a 2D plane, we employ z-buffering. This technique evaluates the visibility of each point by its depth, enabling the rendering of 2D images that precisely reflect the spatial relationships and occlusions present in the 3D scene. The core of this technique is the z-buffer (Z), which stores the depth information of the nearest point to the viewer at each pixel.

For each projected point $p_{i,\text{scaled}}$ with a depth z_i, the z-buffer at a specific pixel location (x_i, y_i) is updated as follows:

$$Z(x,y) = \begin{cases} z_i & \text{if } z_i < Z(x_i, y_i), \\ Z(x_i, y_i) & \text{otherwise.} \end{cases} \quad (1)$$

This ensures that, for any given pixel, the z-buffer retains the depth of the closest point, thus enabling the accurate rendering of overlapping objects.

To formalize this process, consider the set of all points $P_{\text{scaled}} = \{p_{1,\text{scaled}}, p_{2,\text{scaled}}, ..., p_{n,\text{scaled}}\}$. The subset of P_{scaled}, denoted as P', comprises points that satisfy the Z-buffer criterion. Mathematically, this can be expressed as:

$$P' = \{p'_{i,\text{scaled}} \in P | Z(p_{i,\text{scaled}}) \text{ is defined in the Z-buffer}\} \quad (2)$$

In this manner, P' includes only those points whose depth values (z_i) are recorded in the Z-buffer, ensuring that only the closest point to the viewer's perspective is visible at any pixel location, effectively managing the rendering of overlapping objects.

4) Correspondence Mapping: A distinctive feature of our framework is the maintenance of a correspondence map (C), which records the original 3D coordinates index of each point projected onto the 2D plane. This mapping facilitates the bidirectional transfer of information between the 2D projection and the 3D data, enabling the accurate reconstruction of modified associated features to the 3D spatial data from its 2D representation. Furthermore, our framework enhances reconstruction speed by integrating a $Mask$ map for identifying active regions and an $Index$ map for tracking efficiently correspondence point from (C), enabling efficient 2D-to-3D mapping.

C. 2D Saliency Prediction and 3D mapping

Upon obtaining the 2D projection, a saliency prediction model is applied. This model, denoted as S, processes the 2D image I and outputs a saliency map $S(I)$, highlighting regions that are likely to attract human attention. The saliency model is based on visual attention mechanisms and can be represented as:

$$S(I) = f_{\text{Saliency}}(I)$$

where f_{Saliency} represents the computational process of the saliency detection algorithm.

1) 3D Saliency Map Reconstruction: The final phase involves mapping the 2D saliency information back onto the 3D point cloud. This is achieved using the correspondence map C obtained during the projection. The reconstructed 3D saliency map M for point p_i is derived as follows:

$$M(p_i') = S(I)(C(p_i'))$$

This equation ensures that the saliency value assigned to each 3D point corresponds to the saliency value of its 2D projection.

D. Extended Applications

While the primary focus of our framework has been on saliency projection in 3D point clouds, its underlying methodologies present opportunities for extension to a variety of other applications, particularly those requiring detailed point-level analysis and generation. These include but are not limited to, part segmentation [32], style transfer [33] in 3D point clouds. Furthermore, the correlation between 2D projections and their corresponding 3D points enriches multi-modal deep learning strategies eg: (Quality Assessment Deep-based Metrics) [17], bridging 3D point cloud analysis with 2D visual data processing.

IV. Study on the Evaluation of 2D Saliency Models for 3D Point Clouds

In this section, we delve into the comprehensive benchmarking and performance evaluation that accompanies our framework. Our investigation is aimed at assessing the effectiveness of 2D saliency models when applied to the context of 3D point clouds.

A. Dataset

The dataset we used originates from the EPFL study [29], comprising 12-point clouds carefully selected to represent a wide range of visual characteristics, split evenly between objects and human figures. It was assembled from eye-tracking data from 21 participants within a VR environment. The visual attention data was thoroughly processed to produce high-quality fixation density maps, reflecting the average intersection of user gaze and head movements with the rendered points in a 3D VR viewport. This dataset is openly accessible for further research and benchmarking in the domain.

B. 2D-Saliency Models

For our study, we selected SalGAN [34], SATSal [23], and DeepGaze IIE [35] due to their exceptional performance in 2D saliency prediction, each underpinned by distinct theoretical foundations. SalGAN is noted for its use of adversarial training, SATSal for its innovative use of a multilevel self-attention mechanism, and DeepGaze IIE advances the capabilities of deep learning in saliency modeling by integrating contextual information and human visual bias. Notably, these models have demonstrated top-tier results on the MIT300 leaderboard [22], showcasing their effectiveness.

C. Evaluation Metrics

For evaluating the generated 3D saliency maps obtained from the application of the aforementioned 2D saliency models within our framework, we selected metrics including Similarity Metric (SIM) [36], Correlation Coefficient (CC) [36], and Kullback-Leibler Divergence (KLD) [36] due to their relevance in measuring prediction accuracy and distribution similarity. Unlike traditional 2D saliency analyses, metrics like NSS [37] and AUC [36] were not employed, as the ground truth data derived from VR experiments provide density maps rather than discrete fixation points. This approach reflects to how gaze data has been collected in VR conditions, where the nature of visual attention convergence within the 3D rendered scenes does not necessitate the intersection of eye-movement orientation ray with a one point. But rather the distribution of neighboring points is considered. This necessitating metrics that effectively capture the essence of spatial saliency distribution as estimated by user engagement with the 3D point sets.

D. View Selection Strategy

Central to our investigation is the exploration of how the quantity and distribution of selected views impact the fidelity of saliency maps when applying 2D saliency models to 3D point clouds. To this end, we devised a systematic view sampling strategy, distributing viewpoints across varied number of orbits around the point cloud, starting from a single one to five distinct orbits. This configuration includes one orbit along the equator and two additional orbits on either side, each separated by a 30-degree angular distance. From each orbit, views are sampled at intervals of 90 degrees for a set of four views and 45 degrees for a set of eight views, this results either to have a total of four or eight views per each orbit. This approach ensures a dense coverage and significant overlap between views, mirroring potential viewer positions within an immersive experience.

E. Rationale and Objectives

The rationale behind our multi-orbit, multi-view sampling strategy is twofold. First, it aims to simulate a wide range of possible viewer orientations, thereby capturing the saliency information from various perspectives. Second, by providing a high degree of view overlap, our method enhances the comprehensiveness of the resultant saliency maps. This overlapping

is critical for ensuring that any potential viewer perspective during the immersive experience is adequately represented.

F. Saliency Prediction and 3D Reconstruction

By applying 2D saliency models to strategically selected views, we assess their capacity to identify salient features across diverse perspectives. Our framework then synthesizes these 2D predictions into unified 3D saliency maps. A key step in this process involves normalizing each reconstructed saliency view independently before implementing thresholding levels T0, T1, and T2. These thresholds are critical for determining the extent to which saliency information from 2D predictions is retained in the 3D reconstruction. Specifically, T0 includes all saliency data from 2D views, whereas T1 and T2 progressively filter this information, retaining only the top 10% and 5% of salient data, respectively. For regions where saliency predictions overlap, we calculate the mean saliency value, ensuring a balanced representation. The entire reconstructed 3D saliency map undergoes a final normalization to accurately represent visual attention as it would occur in an immersive setting. This approach aims to craft a 3D saliency map that not only aggregates but also preserves the integrity of saliency cues across various viewer orientations, enhancing the understanding of how salient features are distributed within the point cloud.

TABLE I
MODELS COMPARISON - ONE ORBIT (4 VIEWS AND 8 VIEWS)

Model	4 Views			8 Views		
	SSIM ↑	CC ↑	KLD↓	SSIM↑	CC↑	KLD ↓
Satsal-T0	0.160	0.062	18.287	0.215	0.076	16.596
Salgan-T0	0.164	0.025	18.247	0.223	0.038	16.536
DeepGaze-T0	0.173	0.023	18.211	0.235	0.031	16.495
Satsal-T1	0.116	0.061	19.824	0.171	0.079	18.113
Salgan-T1	0.118	0.040	19.901	0.170	0.054	18.151
DeepGaze-T1	0.132	0.083	19.725	0.192	0.103	17.952
Satsal-T2	0.077	0.069	21.093	0.121	0.089	19.810
Salgan-T2	0.074	0.043	21.163	0.111	0.060	19.999
DeepGaze-T2	0.088	0.107	20.892	0.130	0.131	19.679

TABLE II
MODELS COMPARISON - THREE ORBIT (12 VIEWS AND 24 VIEWS)

Model	12 Views			24 Views		
	SSIM↑	CC↑	KLD↓	SSIM ↑	CC↑	KLD↓
Satsal-T0	0.286	0.095	14.071	0.357	0.118	11.125
Salgan-T0	0.289	0.027	14.026	0.360	0.039	11.081
Deepgaze-T0	0.313	0.030	13.960	0.388	0.041	11.003
Satsal-T1	0.229	0.100	16.011	0.307	0.128	12.908
Salgan-T1	0.218	0.048	16.338	0.288	0.064	13.288
DeepGaze-T1	0.253	0.121	16.019	0.334	0.158	12.841
Satsal-T2	0.164	0.107	18.396	0.237	0.134	15.729
Salgan-T2	0.143	0.0521	18.917	0.202	0.068	16.596
DeepGaze-T2	0.176	0.146	18.347	0.246	0.186	15.881

G. Quantitative, and Qualitative Analysis

Our experiments systematically explored the performance of the aforementioned saliency models under various saliency thresholding levels (T0, T1, and T2) across different view configurations. The outcomes, delineated in Table I, encompass the analysis of models applied to 4 and 8 views derived from a singular equatorial orbit where $\theta = 0$, offering a baseline for comparison. we observe an increase in model performance metrics (SSIM, CC, KLD) as the number of views increases. This increment suggests a positive correlation between the comprehensiveness of view sampling and the accuracy of saliency predictions.

TABLE III
MODELS COMPARISON - FIVE ORBIT (20 VIEWS AND 40 VIEWS)

Model	20 Views			40 Views		
	SSIM↑	CC ↑	KLD↓	SSIM↑	CC↑	KLD↓
Satsal-T0	0.322	0.088	12.480	0.388	0.111	9.562
Salgan-T0	0.326	0.023	12.438	0.389	0.032	9.526
DeepGaze-T0	0.353	0.023	12.364	0.418	0.032	9.447
Satsal-T1	0.276	0.096	14.046	0.348	0.122	11.312
Salgan-T1	0.262	0.044	14.177	0.327	0.057	11.311
DeepGaze-T1	0.299	0.044	14.177	0.373	0.057	10.881
Satsal-T2	0.208	0.105	16.735	0.281	0.138	13.801
Salgan-T2	0.180	0.050	17.419	0.242	0.066	14.738
DeepGaze-T2	0.221	0.138	16.804	0.294	0.171	13.801

Further extending our analysis, Table II also details the results obtained from a broader sampling, incorporating 12 and 24 views across three orbits with $\theta = 0$, $\theta = -\pi/6$, and $\theta = \pi/6$. Notably, models exhibit enhanced performance, over all metrics, highlighting the benefit of capturing varied range of viewer perspectives. for a richer saliency mapping.

To deepen our insights, Table III presents findings from an even more extensive view sampling approach, involving 20 and 40 views across five orbits. This includes the previously mentioned orbits and introduces two additional orbits at $\theta = -\pi/3$ and $\theta = \pi/3$. This setting presents the most significant performance improvements. This comprehensive view approach optimizes the results of the 3D reconstructed saliency from 2D across all models but also demonstrates the critical role of extensive view overlap in mirroring potential viewer perspective. Our analysis, employing thresholds T0, T1, and T2, highlighted the balance between the breadth of view sampling and the quantity of saliency information. The T0 threshold consistently outperformed, yet the performance gap between from T0 to T1 and T2 was minimal on each independent setting. Crucially, extensive view sampling even with (T2) proved more vital than the complete quantity of saliency data, while with just T2 (5%) of saliency data in multiple-orbit, multiple-view scenarios outperformed denser saliency mappings in (T0 and T1) in settings with fewer views and orbits. This reveals that diversity in views is key to effective 3D saliency reconstruction, rather than relying on larger quantities of saliency information from 2D models with limited viewing perspective.

Figure 2 illustrates the qualitative performance of predicted 3D saliency maps for three methods within our framework using the 40 views setting, the results closely matching the ground truth Fixation Density PCs. DeepGaze IIE's predictions are notably overestimated than those of SATSal and SalGAN

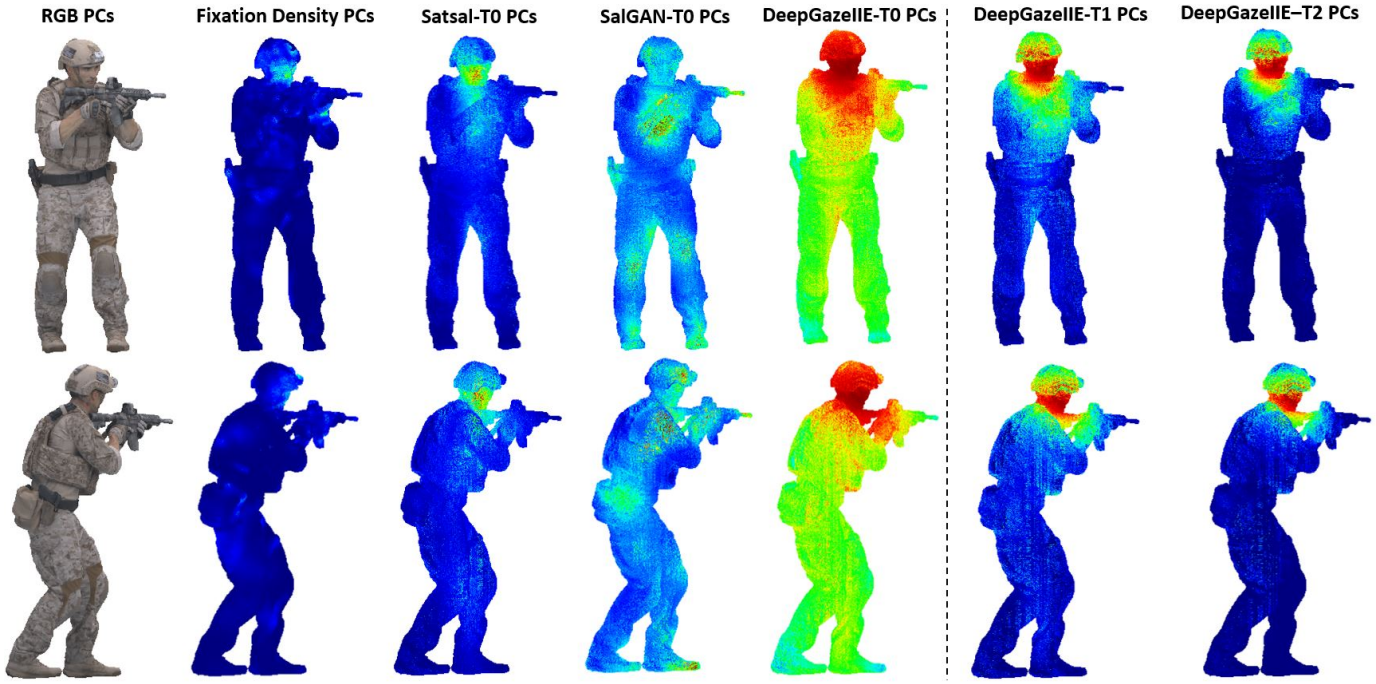

Fig. 2. Qualitative Assessment of 3D Saliency Mapping: Extending 2D Saliency Predictions to 3D Point Clouds Using Our Framework with a 40-View Setting. On the left, the SatSal, SalGAN, and DeepGaze IIE models are illustrated with full saliency integration (T0, 100%). On the right, we display DeepGaze IIE outcomes applying reduced saliency thresholds T1 (10%) and T2 (5%). Fixation Density PCs refers to the ground truth obtained from VR experiments [29]

within T0. Yet, the thresholding refines these predictions, effectively highlighting the method capacity in capturing salient features more accurately.

Our study within the use of our proposed framework, encapsulating both quantitative and qualitative analyses, demonstrates the potential to create representative saliency maps in 3D point clouds, providing a baseline for future improvements of 2D saliency models in 3D environments. Despite a noticeable drop in quantitative performance relative to 2D benchmarks [21], [22], this is attributed to the complexity of point clouds, as the case of high dimensional 360 images [38]. Our approach marks a significant step towards adapting 2D saliency predictions for 3D applications. This initial endeavor highlights the framework's effectiveness in navigating the challenges of high-dimensional saliency modeling.

V. Discussion

As a discussion, we reflect on the broader implications of our study for immersive experiences. Our analysis confirms the critical role of the quantity and arrangement of views in identifying salient areas for 3D data. By exploring a range of viewpoints and their influence on 2D saliency predictions, we not only validate the capability of 2D models prediction to be adapted to 3D contexts but also enable the synthesis of robust 3D saliency maps. Moreover, our framework opens avenues for developing deep learning methods that leverage projected saliency data, advocating for advanced techniques like domain adaptation [39], [40] to bridge the gap between 2D and 3D saliency predictions further. These efforts aim to refine saliency prediction tools for immersive environments, ensuring that salient features stand out, enhancing the user's visual experience.

VI. Conclusion

Our research presents a pioneering framework designed to extend the application of 2D saliency prediction models to the realm of 3D point clouds, aiming to enhance the quality of experience in immersive environments. Through detailed analysis across diverse settings, we've showcased the framework's proficiency in producing meaningful saliency maps and emphasized the crucial role of strategic view selection for accurate salient feature identification. This work lays the groundwork for future advancements in 3D saliency modeling, promising to significantly improve user engagement and immersion by ensuring that salient elements within a scene are consistently and accurately highlighted.

References

[1] J. L. Rubio-Tamayo, M. Gertrudix Barrio, and F. García García, "Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation," *Multimodal technologies and interaction*, vol. 1, no. 4, p. 21, 2017.

[2] H. Wu, T. Cai, D. Luo, Y. Liu, and Z. Zhang, "Immersive virtual reality news: A study of user experience and media effects," *International Journal of Human-Computer Studies*, vol. 147, p. 102576, 2021.

[3] L. A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–6.

[4] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "Objective and subjective qoe evaluation for adaptive point cloud streaming," in *2020 twelfth international conference on quality of multimedia experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[5] Y. Cui, W. Chang, T. Nöll, and D. Stricker, "Kinectavatar: fully automatic body capture using a single kinect," in *Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part II 11*. Springer, 2013, pp. 133–147.

[6] V. Barrile, G. Candela, and A. Fotia, "Point cloud segmentation using image processing techniques for structural analysis," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 187–193, 2019.

[7] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, 2017.

[8] E. Alexiou, N. Yang, and T. Ebrahimi, "Pointxr: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[9] F. Pomerleau, F. Colas, R. Siegwart *et al.*, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends® in Robotics*, vol. 4, no. 1, pp. 1–104, 2015.

[10] A. Durupt, S. Remy, G. Ducellier, and B. Eynard, "From a 3d point cloud to an engineering cad model: a knowledge-product-based approach for reverse engineering," *Virtual and Physical Prototyping*, vol. 3, no. 2, pp. 51–59, 2008.

[11] C. Cao, M. Preda, and T. Zaharia, "3d point cloud compression: A survey," in *The 24th International Conference on 3D Web Technology*, 2019, pp. 1–9.

[12] Y. Dahou, M. Tliba, K. McGuinness, and N. O'Connor, "Atsal: An attention based architecture for saliency prediction in 360° videos," in *Pattern Recognition. ICPR International Workshops and Challenges*. Cham: Springer International Publishing, 2021, pp. 305–320.

[13] M. A. Kerkouri, M. Tliba, A. Chetouani, and M. Sayeh, "Salypath360: Saliency and scanpath prediction framework for omnidirectional images," *Proc. IST Int'l. Symp. on Electronic Imaging: Human Vision and Electronic Imaging*, pp. 168–1 – 168–7, 2022.

[14] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image processing*, vol. 23, no. 10, pp. 4270–4281, 2014.

[15] N. Krucien, M. Ryan, and F. Hermens, "Visual attention in multi-attributes choices: What can eye-tracking tell us?" *Journal of Economic Behavior & Organization*, vol. 135, pp. 251–267, 2017.

[16] X. Zhou, Y. Zhang, N. Li, X. Wang, Y. Zhou, and Y.-S. Ho, "Projection invariant feature and visual saliency-based stereoscopic omnidirectional image quality assessment," *IEEE Transactions on Broadcasting*, vol. 67, no. 2, pp. 512–523, 2021.

[17] M. T. et al., "Pcqa-graphpoint: Efficient deep-based graph metric for point cloud quality assessment," in *ICASSP 2023 - 2023 IEEE (ICASSP)*, 2023, pp. 1–5.

[18] Y. Zhang, S. Kwong, and S. Wang, "Machine learning based video coding optimizations: A survey," *Information Sciences*, vol. 506, pp. 395–423, 2020.

[19] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2012.

[20] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*. Ieee, 2017, pp. 1–6.

[21] M. Jiang, S. Huang, J. Duan, and Q. Zhao, "Salicon: Saliency in context." in *CVPR*. IEEE Computer Society, 2015, pp. 1072–1080. [Online]. Available: http://dblp.uni-trier.de/db/conf/cvpr/cvpr2015.html#JiangHDZ15

[22] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in *MIT Technical Report*, 2012.

[23] M. Tliba *et al.*, "Satsal: A multi-level self-attention based architecture for visual saliency prediction," vol. 10, 2022, pp. 20 701–20 713.

[24] E. Alexiou, Y. Nehmé, E. Zerman, I. Viola, G. Lavoué, A. Ak, A. Smolic, P. Le Callet, and P. Cesar, "Subjective and objective quality assessment for volumetric video," in *Immersive Video Technologies*. Elsevier, 2023, pp. 501–552.

[25] C. Meske, T. Hermanns, M. Jelonek, and A. Doganguen, "Enabling human interaction in virtual reality: An explorative overview of opportunities and limitations of current vr technology," in *HCI International 2022 – Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence*, J. Y. C. Chen, G. Fragomeni, H. Degen, and S. Ntoa, Eds. Cham: Springer Nature Switzerland, 2022, pp. 114–131.

[26] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "Saliency in vr: How do people explore virtual environments?" *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1633–1642, 2018.

[27] A. Nguyen and Z. Yan, "A saliency dataset for 360-degree videos," in *Proceedings of the 10th ACM Multimedia Systems Conference*, 2019, pp. 279–284.

[28] X. Ding, Z. Chen, W. Lin, and Z. Chen, "Towards 3d colored mesh saliency: Database and benchmarks," *IEEE Transactions on Multimedia*, pp. 1–12, 2023.

[29] E. Alexiou, P. Xu, and T. Ebrahimi, "Towards modelling of visual saliency in point clouds for immersive applications," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 4325–4329.

[30] X. Zhou, I. Viola, E. Alexiou, J. Jansen, and P. Cesar, "Qava-dpc: Eye-tracking based quality assessment and visual attention dataset for dynamic point cloud in 6 dof," in *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2023, pp. 69–78.

[31] T. S. Kwanghee Ko, "Orthogonal projection of points in cad/cam applications: an overview," *Journal of Computational Design and Engineering*, 2014.

[32] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, "Segmentation of point clouds via joint semantic and geometric features for 3d modeling of the built environment," *Automation in Construction*, vol. 125, p. 103584, 2021.

[33] D. Jin, Z. Jin, Z. Hu, O. Vechtomova, and R. Mihalcea, "Deep learning for text style transfer: A survey," *Computational Linguistics*, pp. 155–205. [Online]. Available: https://aclanthology.org/2022.cl-1.6

[34] J. Pan, C. C. Ferrer, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, and X. Giro-i Nieto, "Salgan: Visual saliency prediction with generative adversarial networks," *arXiv preprint arXiv:1701.01081*, 2017.

[35] A. Linardos, M. Kümmerer, O. Press, and M. Bethge, "Deepgaze iie: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12 899–12 908, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:235195970

[36] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" *IEEE Transactions on Pattern Analysis amp; Machine Intelligence*, vol. 41, no. 03, pp. 740–757, mar 2019.

[37] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision research*, vol. 45, no. 18, pp. 2397–2416, 2005.

[38] E. David, J. Gutiérrez, M. L.-H. Vo, A. Coutrot, M. Perreira Da Silva, and P. Le Callet, "The salient360! toolbox: Processing, visualising and comparing gaze data in 3d," in *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, ser. ETRA '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3588015.3588406

[39] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[40] M. Tliba, A. Sekhri, M. A. Kerkouri, and A. Chetouani, "Deep-based quality assessment of medical images through domain adaptation," *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 3692–3696, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:252992797