Immersive-Labeler: Immersive Annotation of Large-Scale 3D Point Clouds in Virtual Reality

Achref Doula
doula@tk.tu-darmstadt.de
Technical University of Darmstadt
Darmstadt, Germany

Tobias Güdelhöfer
tobias@guedelhoefer.de
Technical University of Darmstadt
Darmstadt, Germany

Andrii Matviienko
matviienko@tk.tu-darmstadt.de
Technical University of Darmstadt
Darmstadt, Germany

Max Mühlhäuser
max@informatik.tu-darmstadt.de
Technical University of Darmstadt
Darmstadt, Germany

Alejandro Sanchez Guinea
sanchez@tk.tu-darmstadt.de
Technical University of Darmstadt
Darmstadt, Germany

ABSTRACT

We present *Immersive-Labeler*, an environment for the annotation of large-scale 3D point cloud scenes of urban environments. Our concept is based on the full immersion of the user in a VR-based environment that represents the 3D point cloud scene while offering adapted visual aids and intuitive interaction and navigation modalities. Through a user-centric design, we aim to improve the annotation experience and thus reduce its costs. For the preliminary evaluation of our environment, we conduct a user study (N=20) to quantify the effect of higher levels of immersion in combination with the visual aids we implemented on the annotation process. Our findings reveal that higher levels of immersion combined with object-based visual aids lead to a faster and more engaging annotation process.

ACM Reference Format:
Achref Doula, Tobias Güdelhöfer, Andrii Mativiienko, Max Mühlhäuser, and Alejandro Sanchez Guinea. 2022. Immersive-Labeler: Immersive Annotation of Large-Scale 3D Point Clouds in Virtual Reality. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '22 Posters), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532719.3543249

1 INTRODUCTION

The training of deep learning models requires the data to be carefully labeled by humans through a laborious process, referred to as annotation. The annotation process becomes particularly complex and time-consuming when the dataset is composed of 3D point cloud scans that represent complex scenes, like urban areas.Numerous previous works proposed software tools to facilitate the annotation process. However, the majority uses 2D visualization supports, such as monitors, which compromises the quality of perception and interaction with 3D data [Zimmer et al. 2019]. Motivated by the recent advances in virtual reality and its ability to improve the perception of 3D scenes, few previous works proposed to perform annotation in virtual reality (VR) [Wirth et al. 2019]. However, their design does not account for the interaction effects between the high level of immersion and the implemented interaction modalities with the scene, which caused the users to report perception-related problems during the annotation process [Wirth et al. 2019]. We argue that increasing the level of immersion should be combined with adapted interaction techniques and visual aids in order for it to achieve the desired effects on the annotation process. In this work, we present *Immersive-Labeler*, a fully immersive, VR-based 3D point cloud annotation environment. During the annotation process, the user is offered different navigation and data manipulation techniques. Furthermore, we assist the user with three types of visual aids. We start the evaluation of our environment with a user study (N=20) to investigate the effect of the level of immersion and the visual aids on the speed of the annotation process. The results revealed that the annotation is faster for higher immersion levels. Furthermore, our proposed visual aids significantly increase the speed of the process.

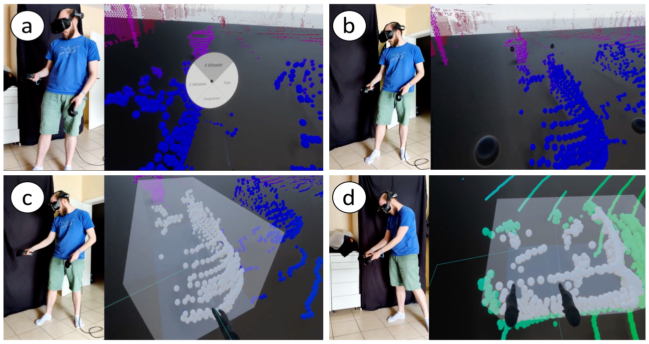

Figure 1: The annotation process consisted of (a) semantic label selection, (b) placing four points to create a base of a bounding box, (c) auto-generation of a bounding box, and (d) a manual adjustment of a bounding box.

2 IMMERSIVE-LABELER

In this paper, we present an annotation environment that is adapted to the specific requirements of annotating large-scale 3D point clouds scenes of urban environments, captured by commercial LIDARs. During the annotation process, the user is fully immersed in

SIGGRAPH '22 Posters, August 07-11, 2022, Vancouver, BC, Canada
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9361-4/22/08.
https://doi.org/10.1145/3532719.3543249

the 3D point cloud scene. The system allows the user to perceive objects at their real-life scale, which gives a better context to decide on the semantic label of the object of interest and its physical extent. To navigate the scene the user has 3 possibilities: (1) walking around the scene for near destinations, (2) teleportation for far but visible destinations, and (3) "as a giant walking" for even further destinations. The three locomotion modalities will allow the use of *Immersive-Labeler* in large as well as small spaces. To annotate an object of interest, the user first chooses the semantic label of the object and places four points around it. These points will set the edges of the bounding box. The user can easily resize and adjust the orientation of the bounding box by using the controllers to grab it and physically stretch or minimize the distance between the hands, as depicted in figure 1. Furthermore, we implement three types of visual aids to further help the user spot all the details of the scene. The implemented visual aids are: (1) distance- and (2) region-based coloring, and (3) object-based highlighting. With the *distance-based coloring,* the surrounding data points were colored based on their distance to the center of the scene, according to a predetermined color palette. the colors ranged from green for the nearest points to red for the points located at the borders of the scene. In the *region-based coloring*, we split the 3D point cloud scene into regions with radii of four meters each. Each region is assigned a unique color and does not account for similarities of objects within them. Finally, for the *object-based* highlighting, we used bounding boxes, generate by a pre-trained neural network, around objects in the scene to provide precise semantic indications and location for objects of interest.

3 EVALUATION AND RESULTS

We report the results of a preliminary user study we conducted to investigate the effects of the level of immersion and the visual aids on the speed of annotation. For this, we used *Immersive-Labeler* in 9 different configurations by varying two independent factors: (1) the level of immersion, and (2) the visual aids. In our study, we adopt three different levels of immersion: non-immersive, semi-immersive, and fully immersive. In the non-immersive variant, the users visualized the scenes on a 2D monitor while sitting on a chair with no degrees of freedom. In the semi-immersive variant, The participants used an HMD for the visualization. We limited the VR system to only have 3 degrees of freedom. The user sits on a rolling chair and is allowed to only perform rotations. The navigation in the scene is then done through teleportation. In the fully-immersive variant, the user wears an HMD to visualize the scene. A total of 6 degrees of freedom were provided to exactly map the users' motion in the scene, thus totally immersing him in the activity. For the study, we recruited 20 participants aged between 18 and 32 (M=25,4, SD=3,137). The task assigned to the participants consisted on visualizing the 3D point cloud environment, recognizing a class instance from a list of semantic classes that were shown at the start of the task, and annotating the object with the highest possible accuracy. For each participant and each experiment condition, we logged the Task Completion Time (TCT), defined as the start of the task and the participant's confirmation of the annotation. Furthermore, the participants were asked to provide their feedback, at the end of the experiment, in the form of a semi-structured interview.

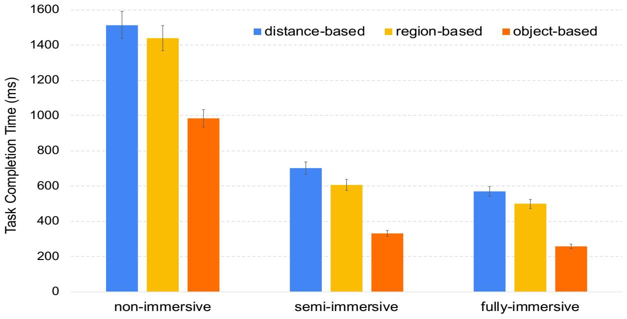

Figure 2: Mean and standard error for different visual aids and level of immersion combinations

For the analysis of the results, we use two-way repeated-measures (RM) ANOVAs with the level of immersion and visual aids as the two independent factors. We tested the sphericity assumption using Mauchly's test. For the cases where the sphericity is violated, we used the Greenhouse-Geisser correction. Additionally, we used the Bonferroni corrected pairwise t-tests for post-hoc analyzes.

Furthermore, we used the eta-squared η^2 to estimate the effect size based on Cohen's classification. The analysis shows a significant influence of the level of immersion on the TCT with a large effect size ($F(1.37, 26.038) = 340.991, p < 0.001$, $\eta^2 = 0.648$). The post-hoc analysis backs up this finding. In fact, the TCT decreases significantly when the level of immersion increases (non-immersive: EMM μ = 1310.84 ms, σx = 140.994, semi-immersive: EMM μ = 540.583ms, σx = 100.104, fully-immersive: EMM μ = 440.126ms, σx = 60.784). The more the participant is immersed in the scene, the lower the TCT value is. Similarly, the visual aids proved to have a significant influence on the TCT with a large effect size as well ($F(1.465, 27.828) = 9.251, p = 0.002$, $\eta^2 = 0.327$). Post-hoc analysis confirmed lower TCTs for visual aids with finer context indication (distance-based: EMM μ = 920.830ms, σx = 140.104, region-based-based: EMM μ = 840.745ms, σx = 90.8, object-based: EMM μ = 520.318ms, σx = 60.945). The interaction effects between the level of immersion and the visual aids were not statistically significant ($F(1.465, 27.828) = 0.442, p = 0.642$, $\eta^2 = 0.023$). The results are depicted in the figure 2. During the interview, all participants showed a strong interest in the idea of a fully immersive annotation environment. When asked about the aspects that made the annotation task easier, all participants mentioned that, in addition to the visual aspect of the fully-immersive configuration, the scale mapping between the real world and the virtual world provided an intuitive way to map real world-objects to their 3D point cloud counterparts.

ACKNOWLEDGMENTS

This work has been funded by National Research Center for Applied Cybersecurity ATHENE and LOEWE initiative (Hesse, Germany).

REFERENCES

Florian Wirth, Jannik Quehl, Jeffrey Ota, and Christoph Stiller. 2019. Pointatme: efficient 3d point cloud labeling in virtual reality. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1693–1698.

Walter Zimmer, Akshay Rangesh, and Mohan Trivedi. 2019. 3d bat: A semi-automatic, web-based 3d annotation toolbox for full-surround, multi-modal data streams. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1816–1821.