Embodied Data Annotation Technology Based on Scene Perception

Zhang Bojun
12211615
Zheng Feng Research Group
12211615@mail.sustech.edu.cn

Ye Hangajin
12212012
Zheng Feng Research Group
12212012@mail.sustech.edu.cn

ABSTRACT

Context. Research and develop a semi-automatic, network-based tool for automated 3D point cloud annotation, supporting panoramic annotation of multimodal data streams, focusing on image and point cloud data processing.

Goal. The development of a 3D point cloud automated annotation tool aims to provide a semi automated, network-based platform that can greatly improve the processing efficiency of 3D data.

Method. We have combined the efficient processing capability of the SIBR algorithm for 3D point clouds and utilized the Vue+Lite framework in front-end applications to provide an efficient user experience. In addition, the cloud architecture of the system combined with panoramic annotation of multimodal data streams enables our annotation tool to provide high-quality and efficient annotation support in fields such as autonomous driving and intelligent transportation.

Results. This tool focuses on the processing of image and point cloud data, providing an efficient solution. We have completed the construction of scene perspective cameras and orthogonal cameras, point cloud annotation, implementation of annotation boxes, UI interaction, loading and annotation saving of point cloud data, front-end and back-end linkage interaction, etc., to support embodied intelligent training and other applications based on 3D data analysis.

Conclusions. This study successfully developed a semi-automatic 3D point cloud automated annotation tool based on Gaussian segmentation and network, significantly improving the efficiency and accuracy of data annotation. By supporting panoramic annotation of point cloud data, this tool provides strong technical support for the embodied intelligence field and can meet the annotation needs of complex traffic scenes.

ACM Reference Format:

Zhang Bojun and Ye Hangajin. 2025. Embodied Data Annotation Technology Based on Scene Perception . In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

This document presents a detailed report on the development of a semiautomatic, network-based tool for automated 3D point-cloud annotation, specifically designed for the processing of image and point-cloud data. The domain of this experiment lies at the intersection of autonomous driving technology, computer vision, and 3D data processing. The technologies relevant for this experiment include point cloud processing, Gaussian segmentation, machine learning, and network-based automation techniques, which are essential for the development and improvement of 3D data annotation tools.

The primary motivation behind this experiment is to address the growing need for efficient, accurate, and scalable solutions for annotating large-scale 3D point cloud datasets, which are widely used in autonomous driving and traffic perception systems. Annotating these datasets manually can be a time-consuming and error-prone task, especially in complex traffic scenes. Existing tools often lack the necessary automation to handle the intricacies of 3D data and can struggle with the demands of large-scale annotation tasks. For example, manual annotation processes can result in inconsistencies, inaccuracies, and high costs, making them impractical for large datasets required for training autonomous systems.

This experiment aims to develop a semiautomatic annotation tool that leverages Gaussian segmentation and network-based techniques to improve the efficiency and accuracy of the annotation process. The tool supports panoramic annotation of point clouds and is specifically tailored for handling complex 3D scenes, such as those encountered in traffic environments. By providing an automated solution for annotating 3D data, this tool significantly reduces time, cost, and errors associated with manual annotation, making it a valuable asset for researchers and developers in the field of autonomous driving.

The developers and researchers involved in this experiment will gain valuable insights into how to create and implement network-based tools for 3D data annotation. In addition, they will learn how such tools can be used to process large datasets with high accuracy, which is crucial to developing intelligent systems capable of understanding and interpreting complex environments, such as autonomous vehicles in real-world traffic scenarios.

2 RELATED WORK

Although there are already some 3D point cloud annotation related platforms or systems on the market, but most of them who are relatively mature and easy to use are closed-source projects, while those open source projects are not easy to use. SO base on this situation, we decides to design out own 3D point cloud annotation platform inorder to better fitting the needs of the laboratory.

related projects:

1.SUPERVISELY

SUPERVISELY is an open-source project that provides a comprehensive workflow solution for data annotation and team collaboration in computer vision tasks. Through this platform, developers, researchers, and data scientists can efficiently create, manage, and share high-quality image and video annotation data.

At its core, Supervisely is a web-based interface with an intuitive and easy-to-use UI design that supports a wide range of sophisticated annotation tools. These tools include, but are not limited to, points, lines, polygons, rectangles, 3D boxes, etc., to meet the needs of various object detection, segmentation, and key point recognition. In addition, the platform integrates real-time preview and playback functions so that users can quickly verify the results during the annotation process.

The project uses a microservices architecture that allows for flexible scaling and customization. Its backend is based on Node.js and GraphQL APIs, ensuring efficient data processing and communication. The front-end is built with React, which ensures the responsiveness of the interface and the user experience.

2.3D-BAT

3d-bat proposes a novel annotation system with tools for efficient,accurate 3D localization and tracking of objects using full-surround multi-modal data streams. It contains many powerful functions like Full-surround annotations, AI assisted labeling,Batch-mode editing,3D to 2D label transfer and so on. However, when we actually use it to do some annotation, we found it a little bit hard to use, especially the movement, switching of the 3D scene and editing of the bounding box are not very smooth.

Compared to these projects, we believe that out project is more user-friendly and smoother to use.

3 PROJECT PLANNING

3.1 Team Members and Division of Labor

This project is jointly participated by two developers, each responsible for development work in different fields

Zhang Bojun: Mainly responsible for the development of visual management systems, including visual management of perspective cameras and orthogonal cameras, BBox annotation operations, design and development of user-friendly interfaces (UI), optimization of 3D visualization interfaces and their visual layouts, etc.

Ye Hangjian: Mainly responsible for front-end interface , including multi-file selection function, graphical interface of labels, related logic of annotation, and so on. Also responsible for front-end and back-end communication, such as designing back-end API and data transmission.

3.2 Project Design

In the project initiation phase, the team first conducted in-depth research and learning on an open-source project -3D Bat on GitHub. This project provides a basic front-end framework and demonstrates how to implement basic annotation tool functions, especially for the annotation execution logic of BBox (bounding box). Through analysis and learning of the project, the team has mastered the application of relevant technology stacks and tools, and determined the technical roadmap of the project.

In the initial stage of the project, we will focus on the following aspects of work:

3.2.1 BBox annotation operation. As one of the core functions of this project, the team first implemented the BBox annotation tool. This tool supports users to box select objects in point cloud data and generate corresponding annotation information.

3.2.2 Optimization of Visual Display Issues. The visual display in the project is particularly critical, and the team has deeply optimized the image display to ensure accurate presentation of annotated information and smooth user interaction.

3.2.3 Point Cloud Data Import and Label Management System. The visual display in the project is particularly critical, and the team has deeply optimized the image display to ensure accurate presentation of annotated information and smooth user interaction.

3.2.4 Front end and backend API interaction. The project has completed the API communication function between the front end and backend, ensuring data synchronization and interaction between the front-end user interface and the backend database.

As the project progressed, the team continued to expand and optimize the system's functionality, including but not limited to a more sophisticated user interface (UI), more efficient annotation management tools, and further improvements to 3D visualization effects.

3.3 Function of our system

As of now, the project has achieved the following phased results:

3.3.1 BBox annotation function. Successfully implemented BBox annotation function based on point cloud data, allowing users to visually select target objects in 3D views and generate annotation data.

3.3.2 3D visualization function. The project has completed a 3D visualization function based on point cloud data. Users can view, rotate, and scale point cloud data through an interactive interface, and analyze it from multiple perspectives.

3.3.3 UI management system. A complete user interface management system has been developed, providing an intuitive and easy-to-use interactive experience, including modules for annotation operations, data import, file management, etc.

3.3.4 Point cloud data import and management. A point cloud data import system has been established, which supports uploading, storing, and managing multiple data formats to ensure data integrity and usability.

3.3.5 Front end and back-end API interaction. Completed the API interaction design between the front-end and back-end, and implemented functions such as data transmission, annotation information storage, and data synchronization.

3.4 Detailed description

This figure shows how our playform looks like and how many functions it contains.SO next let's show each function with images.

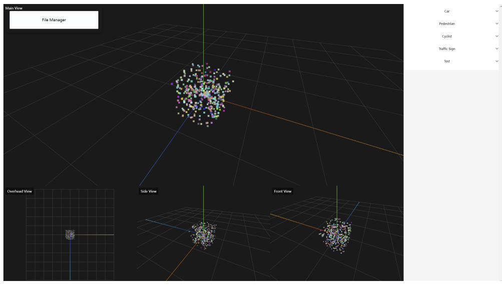

Figure 1: Home page of the playform
Description: This is the home page of our annotation platform, which contains 1:Home/Main View 2:File selector: 3.label controllor.

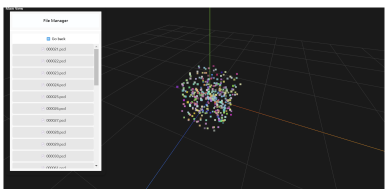

Figure 2: File Selector
Description: Opens the file management panel, it can import and manage point cloud files.

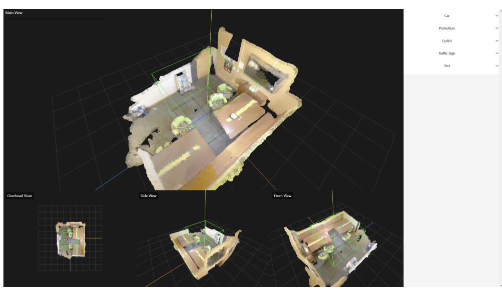

Figure 3: add annotation box
Description: The add annotation box feature allows users to create bounding boxes around specific regions

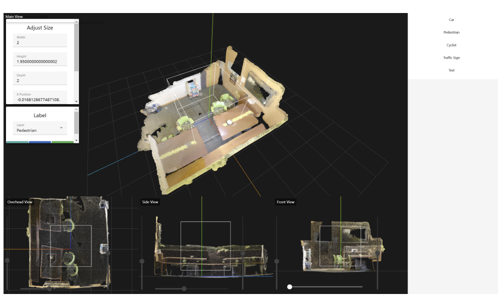

Figure 4: edit annotaion box
Description: In this interface, we can change the size, position, labels, and other properties of the boundingbox

Through the implementation of these achievements, the project has made significant progress in visual presentation, data management, and user experience, laying a solid foundation for functional expansion and performance optimization in subsequent stages.

4 STRUCTURE OF SOME CONPONENTS

- Relationship diagram

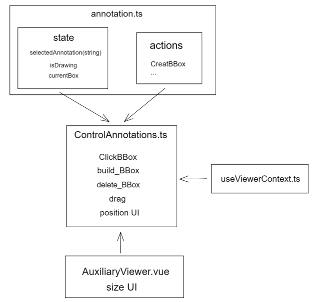

Figure 5: Basic Operation Structure
Description: This structural diagram is an introduction to the basic annotation operation process, with the core being ControlAnnotations.ts, which includes various basic annotation functions such as click selection, creating BBox, deleting BBox, dragging and dropping, and the implementation of various operation UI.

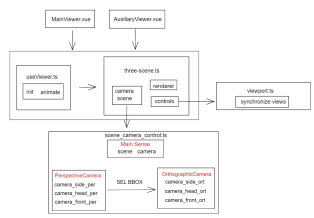

Figure 6: Camera Structure
Description: This diagram shows the camera control and operating system. The core of the camera operating system is scene camera control, which is responsible for storing various cameras in the main and auxiliary views, and managing the types of camera currently in use. The viewport is responsible for managing controls and assisting camera sync in executing camera synchronization visual operations.

5 DISCUSSION

5.1 Overview

The focus of this study is to develop a semi-automatic 3D point cloud annotation tool based on Gaussian segmentation and network technology. This tool addresses the inefficiencies and inaccuracies of existing annotation tools when processing large-scale 3D point cloud data. By implementing features such as BBox annotation, 3D visualization, UI management system, point cloud data import and management, and front-end and back-end API interaction, we successfully built a platform capable of supporting embodied intelligent training and other 3D data analysis applications.

5.2 Main Research Achievements and Impacts

5.2.1 Improved Annotation Efficiency and Accuracy. By integrating a Gaussian segmentation algorithm with automatic annotation functionality, the tool significantly reduces the time and labor required for manual annotation. This also enhances the accuracy of annotations, making it more reliable for complex data processing.

5.2.2 Support for Panoramic Annotation. The tool offers panoramic annotation capabilities for point cloud data, an essential feature for analyzing complex traffic and other 3D scenes. This broadens the applicability of the tool in diverse scenarios.

5.2.3 User Friendliness. Special attention has been given to user interface design and interactivity, ensuring that the tool is intuitive and accessible even for users without professional backgrounds. This lowers the barrier to entry and promotes wider adoption.

5.2.4 Scalability. The tool is built with a modular design, enabling seamless expansion of functionalities and optimization of performance. Future developments could include supporting more annotation types and incorporating deep learning models for enhanced assistance.

5.3 Future Research Directions

5.3.1 Deep Learning Integration. Future iterations of the tool will incorporate deep learning models for automatic and auxiliary annotation, further enhancing efficiency and accuracy.

5.3.2 Support for Advanced Annotation Types. Expanding the tool to include semantic segmentation, instance segmentation, and other annotation types will address a wider range of application needs.

5.3.3 Optimized 3D Visualization Effects. Improvements in point cloud rendering speed and visual effects will enhance the user experience, ensuring smoother and more effective interactions with the data.

5.3.4 Cloud Platform Development. Deploying the tool on a cloud platform will facilitate convenient remote access, enable collaborative annotation, and support a distributed workflow for users across various locations.

5.4 Conclusion

The 3D point cloud annotation tool developed in this study provides robust technical support for the field of embodied intelligence, addressing the annotation needs of complex traffic scenes and beyond. By focusing on efficiency, accuracy, user-friendliness, and scalability, this tool contributes to the advancement of 3D data analysis applications. Future work will focus on refining the tool to make it even more effective, aiming to drive further progress in embodied intelligence technology.

6 REFERENCES

1.Pointcept. (2023). SAM3D: Segment Anything in 3D Scenes. ICCV 2023 Workshop.

2.A Codebase for Point Cloud Perception Research. Includes recent works: PTv3 (CVPR 2024 Oral), PPT (CVPR 2024), OA-CNNs (CVPR 2024), and MSC (CVPR 2023). 3.Jain, A. (2024).

ODIN: A Single Model for 2D and 3D Segmentation. CVPR 2024.

4.Facebook Research. (2023). VLPart: Going Denser with Open-Vocabulary Part Segmentation. ICCV 2023.

5.UX-Decoder. (2024). Semantic-SAM: Segment and Recognize Anything at Any Granularity. ECCV 2024.

6.Reference details are not provided but can be listed as: Author(s). (Year). Hierarchical Open-Vocabulary 3D Scene Graphs for Language-Grounded Robot Navigation. Conference or journal name (if known).