Immersive Setup of Autonomous Forklifts in Factories

Joon Hyub Lee
DRB-KAIST SketchTheFuture Research Center
KAIST
Daejeon, Republic of Korea
joonhyub.lee@kaist.ac.kr

Sang-Hyun Lee
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
sang-hyun.lee@kaist.ac.kr

Hanbee Jang
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
hanbee.jang@kaist.ac.kr

Siripon Sutthiwanna
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
siripon.s@kaist.ac.kr

Hyelim Hwang
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
hyelim.hwang@kaist.ac.kr

Seok-Hyung Bae
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
seokhyung.bae@kaist.ac.kr



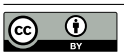
Figure 1: While newly constructed factories increasingly incorporate logistics automation from the outset, most existing factories still rely on manned forklifts, and there is a strong demand for gradual automation. We propose a novel interactive system in VR that facilitates an efficient setup workflow of autonomous forklifts in such factories. In a 3D reconstructed factory, the user immersively annotates (a) operating areas, (b) pallet locations, and (c) automatic doors, by selecting grid points on a handheld plane that corresponds to a remote plane aligned with features in the environment. The user can (d) simulate AFL task execution and inspect it at real scale and in real time.

Abstract

Autonomous forklifts (AFLs) are being introduced to address labor shortages and safety concerns in factory logistics, but existing setup methods rely on sparse LiDAR point clouds with insufficient spatial context. We propose a novel interactive system in VR for immersively setting up AFLs within a 3D reconstruction of the target factory. Our workflow leverages users' sense of presence and situational awareness, enabling them to quickly and easily annotate information essential for AFL operation through projective plane interaction. Users can then simulate the resulting 3D node-link task map at real scale and in real time. We conducted a case study in a factory and received positive initial feedback from stakeholders.

Keywords

Autonomous forklift; 3D scan; hand gesture; VR



This work is licensed under a Creative Commons Attribution 4.0 International License.
UIST Adjunct '26, Detroit, MI, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2855-6/26/11
https://doi.org/10.1145/3830397.3841856

ACM Reference Format:
Joon Hyub Lee, Sang-Hyun Lee, Hanbee Jang, Siripon Sutthiwanna, Hyelim Hwang, and Seok-Hyung Bae. 2026. Immersive Setup of Autonomous Forklifts in Factories. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '26), November 02–05, 2026, Detroit, MI, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3830397.3841856

1 Introduction

In factories, forklifts are indispensable to internal logistics, as they can lift heavy objects on pallets, operate from floor level to high shelves, and maneuver easily in tight spaces. Recently, autonomous forklifts (AFLs) are being introduced into factories to alleviate labor shortages and safety concerns. Instead of a driver's seat, AFLs are equipped with sensors and control systems for map-based localization and navigation [1, 6]. Although advances in physical AI are expected to enable more flexible and intelligent operations [5, 7], most AFLs commercially available today primarily operate according to predefined paths and actions.

Currently, to set up an AFL, engineers first use LiDAR to construct a 2D localization map of the target factory. They then manually define paths and actions on this map in the form of nodes and links [2, 6] (Figure 2a). However, sparse 2D point clouds provide limited spatial context, making it difficult to know in advance whether

the initially configured nodes and links will translate into the intended physical paths and actions. The configuration must therefore be verified onsite using an actual AFL. If the desired behavior is not achieved, engineers must iterate between map configuration and onsite verification through repeated trial and error. In large-scale factories, this process can take up to several weeks.

2 Interactive System

To address this bottleneck, we propose a novel interactive system for setting up AFLs in a factory through a three-step workflow: 3D scan, immersive annotation, and simulation (Figure 1). Instead of configuring task maps on barebones 2D point clouds, users perform the setup immersively within a rich, detailed 3D reconstruction of the factory, producing a 3D node-link task map as the final output (Figure 2).

3D scan. Users create a high-precision, high-fidelity 3D reconstruction of the factory with a handheld industrial 3D scanner in a single walkthrough. The scan produces 3D Gaussian Splatting (3DGS) data [4, 9] used for immersive annotation in the next step, and corresponding 3D LiDAR point cloud data used as the localization map for the actual AFLs [8] after the setup.

Immersive annotation. Users enter the completed 3D reconstruction in VR, observe the environment at real scale, and annotate three types of spatial information: first, the operating areas within the factory where the AFLs can move; second, candidate locations where pallets can be placed; and third, various kinds of automatic doors that the AFLs must pass through along their routes.

To do so, users perform intuitive hand gestures based on projective plane alignment [3]. Originally proposed for interior architectural design, this interaction technique allows users to experience an environment at a 1:1 scale while quickly and easily drawing various prismatic forms with precise dimensions using a grid system. This aligns with our goal of helping users fully leverage their sense of presence and situational awareness to make better informed decisions in complex factory environments.

Specifically, users align a remote plane with recognized orthogonal features in the 3D reconstruction, adjust the grid size on the handheld plane, and then pinch the desired grid points to create straight lines and closed polylines. In our system, a polygon drawn on the floor specifies an operating area (Figure 1a); a horizontal line drawn on a vertical surface specifies pallet locations (Figure 1b), with as many pallets as fit, placed at equal intervals along the line; and a rectangle drawn on a vertical surface specifies an automatic door (Figure 1c).

Simulation. Based on the essential information annotated by users in the 3D reconstruction, a task map is generated in the form of a 3D node-link diagram. First, nodes and links representing "main roads" are laid out along the center of each operating area, and "side streets" branching toward each pallet location are added. Next, nodes encoding maneuvering routines for picking up or placing pallets, as well as waiting routines for passing through automatic doors, are inserted into this road network (Figure 2b).

Users can create custom tasks on the task map and observe a virtual AFL executing them within the 3D reconstruction at real scale and in real time (Figure 1d). The task map can be loaded onto the AFL together with the LiDAR localization map for deployment in the actual factory.

3 Implementation

To implement the proposed system, we used the XGRIDS LixelKity K1 handheld 3D scanner, which weighs approximately 1 kg and features a LiDAR sensor with a measurement rate of 200,000 pts/s, a measurement range of 40 m, rated accuracy of ±1.2 cm, and two 48 MP cameras that together provide 360° coverage. Point cloud data were processed using LixelStudio, and 3DGS data were processed using Lixel CyberColor.

The VR client was implemented in Unity and used XGRIDS' official 3DGS plugin for Unity and a Meta Quest 3 VR headset. The client ran on a PC equipped with an Intel Core i9-14900HX CPU, an NVIDIA GeForce RTX 4070 GPU, and 32 GB of RAM.

4 Case Study

We applied our system to a factory of an industrial materials and components manufacturer. The target process takes place in an indoor-outdoor area measuring approximately 35 m × 50 m. It involves receiving semi-finished products placed on pallets from a preprocessing area, storing them in two warehouses, transporting the pallets to a postprocessing area at specified times, and returning the empty pallets to the preprocessing area after use. To address labor shortages and safety concerns, the company sought to replace a manned forklift with an AFL.

We 3D scanned the facility for 15 minutes and processed the scan data into 3DGS and 3D LiDAR point cloud formats for 3 hours. We then annotated the 3D reconstruction for 30 minutes, generated a task map in the form of a 3D node-link diagram, and simulated the result in VR. The engineers compared our system favorably with existing methods, which required them to work with limited 2D point clouds and often involved more than two weeks of repeated trial and error. The managers also gained confidence in the planned AFL operation scenarios through the lifelike visualization and requested that the system be applied to more processes.

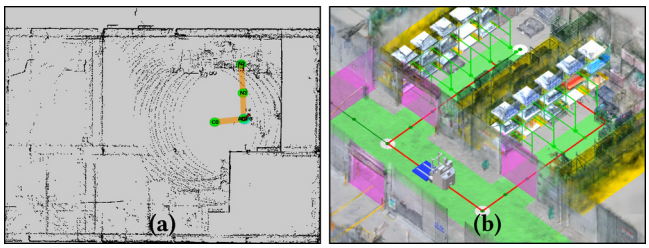


Figure 2: (a) In existing methods, a 2D point cloud of a factory provides insufficient spatial context for creating a node-link task map. (b) Our system's 3D node-link task map in isometric view, showing the operating areas (green), pallet locations (yellow), and automatic doors (pink) overlaid on the 3D Gaussian Splatting (3DGS) reconstruction.

5 Conclusion & Future Work

AFLs perform relatively simple macros, such as picking up a pallet at point A and placing it at point B, yet setting them up on 2D point clouds with insufficient spatial context requires guesswork and repeated onsite validation of whether the configured nodes and links produce the intended paths and actions.

We aimed to eliminate this bottleneck by combining the latest 3D scanning technology with immersive VR interaction. In our system, users can intuitively annotate essential information directly on the 3DGS of the factory environment. Meanwhile, the precise alignment of the 3D LiDAR point cloud with the 3DGS ensures that AFLs navigate paths and execute actions as intended, thereby reducing costly trial and error during onsite validation.

Furthermore, users can inspect how AFLs would operate in the environment at real scale and in real time during the early planning stage. In the case study, such visualization facilitated communication and decision-making among factory engineers and managers. Future work should deploy the proposed system on real AFLs, collect further feedback from stakeholders, and improve its usability.

Acknowledgments

This research was supported by the DRB-KAIST SketchTheFuture Research Center, and the Technology Commercialization Support Program through the Korea Innovation Foundation funded by the Ministry of Science and ICT. We thank Minho Seo and Kijung Kim for their help with 3D scanning the factory.

References

[1] Guillem Alenyà, Josep Escoda, Antonio B. Martínez, and Carme Torras. 2005. Using laser and vision to locate a robot in an industrial environment: a practical experience. In *Proc. ICRA '05*. 3539–3544.

[2] BlueBotics. 2022. Installing an ANT driven vehicle in 5 steps. Retrieved July 10, 2026 from https://youtu.be/p-l4a_ZG8g0

[3] Hanbee Jang, Seung-Jun Lee, and Seok-Hyung Bae. 2026. Projective walls, floors, and windows: aligned plane interactions for interior architectural design in VR. In *Proc. UIST '26*. To appear.

[4] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian splatting for real-time radiance field rendering. *TOG* 42, 4, Article 139 (2023), 14 pages.

[5] Pierrick Lorang, Hong Lu, Johannes Huemer, Patrik Zips, and Matthias Scheutz. 2025. Few-shot neuro-symbolic imitation learning for long-horizon planning and acting. In *Proc. CoRL '25*. 2501–2518.

[6] Humberto Martínez Barberá, Juan Pedro Cánovas Quiñonero, Miguel Ángel Zamora Izquierdo, and Antonio F. Gómez Skarmeta. 2003. i-Fork: a flexible AGV system using topological and grid maps. In *Proc. ICRA '03*. 2147–2152.

[7] Koshi Oishi, Teruki Kato, Hiroya Makino, and Seigo Ito. 2025. Visual-based forklift learning system enabling zero-shot Sim2Real without real-world data. In *Proc. ICRA '25*. 4915–4921.

[8] Yue Pan, Xingguang Zhong, Liren Jin, Louis Wiesmann, Marija Popović, Jens Behley, and Cyrill Stachniss. 2025. PINGS: Gaussian splatting meets distance fields within a point-based implicit neural map. In *Proc. RSS '25*.

[9] Benjamin Reynolds, James Hobin, Yuanzhi Cao, Martin Klaudiny, Daniel A Dangond, Kaize Xie, Keranie Theodosiou, Rik De Leeuw, Shiva Kashalkar, and Valentin Heun. 2024. The pop-up metaverse: a multi-user, multi-tasking spatial computing environment for collaborative spatial problem-solving. In *Proc. CHI EA '24*. Article 424, 5 pages.