Automated Room-Level Labelling of 360 Degree Video Frames Using SLAM

Samuel A. Prieto[1[0000-0001-8341-2630]], Eyob T. Mengiste[1[0000-0002-4841-7476]], Mohamed Benaich[1], and Borja García de Soto[1[0000-0002-9613-8105]]

[1] S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), United Arab Emirates (UAE)
samuel.prieto@nyu.edu

Abstract. This paper introduces an end-to-end workflow that automatically assigns room-identifiers to every frame of 360-degree video collected on construction sites. Manual tagging of imagery is a major bottleneck for progress-tracking and safety analytics, where inspectors may capture thousands of panoramas per walk and then spend hours linking them to rooms on the floor plan. To eliminate that overhead, we couple visual-inertia SLAM with a simplified, BIM-derived floor plan. Once the SLAM trajectory is rigidly aligned, each trajectory point inherits the label of the room it falls inside, yielding a fully-georeferenced image set ready for analysis. The approach was validated on a real interior renovation project, where an operator completed three site walks per week with a 360-degree camera. The resulting high-granularity, spatially coherent imagery accelerates inspection, progress quantification and safety audits without extra sensors, pre-training or constrained capture routes, making the workflow deployable from early construction stages onward. Future work will explore drift-correction and real-time labeling of construction elements to widen adoption across diverse building typologies.

Keywords: Automated labelling, 360-degree video, visual-inertia SLAM, progress monitoring

1 Introduction

Monitoring progress and ensuring safety on construction sites are critical for maintaining schedules, controlling costs, and upholding compliance with safety standards. However, the current practices for collecting and labeling spatial data during inspections are often labor-intensive and prone to human error. Inspectors typically rely on manual methods to record and organize visual data, requiring them to annotate which images correspond to specific rooms or locations. This process is not only time-consuming but also introduces inefficiencies that can hinder timely decision-making.

To address these challenges, this study introduces an automated method to georeference 360-degree video footage and assign spatial labels to video frames within a building's floorplan. By leveraging Simultaneous Localization and Mapping (SLAM) techniques, the trajectory of an operator's path through the construction site is

reconstructed and aligned with a provided 2D floor map. This alignment enables the automatic assignment of room-level labels to each video frame, eliminating the need for manual annotation and significantly streamlining the data collection workflow.

The proposed approach is particularly valuable for construction progress and safety monitoring, where timely and accurate spatial data are essential. By automating the room-level labeling process, the system enhances both the efficiency and accuracy of data collection. Moreover, this method reduces the cognitive load on inspectors, allowing them to focus on critical evaluation tasks rather than administrative duties.

This paper details the development and validation of the proposed method, highlighting its potential to improve spatial data organization and progress tracking in construction. The remainder of the paper is organized as follows: Section 2 reviews the state of the art in spatial data collection and SLAM-based techniques. Section 3 describes the methodology, including the integration of 360-degree video footage, SLAM trajectories, and floor map alignment. Section 4 presents the experimental setup and results, demonstrating the accuracy and efficiency of the system. Finally, Section 5 discusses the conclusions, and future directions for this research.

2 State of the art

The advancements in spatial data organization and indoor mapping technologies have significantly enhanced the efficiency of capturing, processing, and interpreting data for various applications, particularly in construction and building management. State-of-the-art methods leverage techniques such as visual simultaneous localization and mapping (VSLAM) [1], 360-degree imagery [2], sensor fusion [3], and semantic segmentation [4] to automate spatial referencing and floor plan generation [5], [6], [7]. These approaches range from floor plan estimation using 360-degree images and stereo matching for acoustic modeling to the integration of BIM data for automated labeling and semantic understanding. Table 1 presents the summary of the key contributions and constraints of methods proposed in recent literature specifically for applications in the construction and building environments.

Table 1. Summary of recent approaches proposed for construction and built environments.

Ref.	Method	Contribution	Limitation
[8]	360-DFPE	Floor plan estimation using VSLAM and 360 images for space reconstruction; identifies rooms, tracks transitions, and outputs 2D maps	Applicable only to already built spaces; limited in early construction stages
[9]	SegNet and Stereo Matching	Models room acoustics using 360-degree stereo images; predicts acoustic properties with no need for traditional setups	Focused on acoustic modeling; does not support spatial segmentation or construction site applications
[10]	Sensor Fusion Algorithms	Combines video and sensor data (gyroscope, accelerometer) to create indoor floor plans; supports segmentation by spaces	Requires user input and crowdsourced data, which may lead to inconsistencies
[11]	Seg2Reg	Combines segmentation and regression to create accurate room layouts, handling occlusions effectively	Limited to visible regions in panoramic images; relies on differentiable rendering
[12]	Object SLAM	Creates 3D semantic maps with spatial layout and semantic consistency; robust loop closure detection	Requires pre-trained models for object detection; focuses on static environments

[13]	BIM Projection and Inverse Photogrammetry	Automates labeling of construction site images using BIM data; integrates semantic and temporal information	Relies on accurate 4D BIM models; less effective in areas without detailed BIM data
[14]	SfM (Structure-from-Motion)	Provides 3D reconstructions and semantic labeling for indoor spaces; integrates object-to-object constraints in mapping	Requires RGB-D data and extensive manual labeling for dataset creation
[15]	Mask RCNN and ORB	Enhances VSLAM with semantic understanding for high-precision indoor 3D mapping	Computationally intensive; struggles in environments with sparse features
[16]	ORB-SLAM, DynaSLAM	Surveys vSLAM advancements; highlights their applications in mapping, AR, and wayfinding	A survey paper; does not propose new algorithms or implementations
[17]	Panoptic NeRF	Introduces a method for transferring labels from 3D to 2D by combining coarse 3D bounding primitives with noisy 2D semantic predictions	Requires significant computational effort, Its focus on static, outdoor scenes limits its applicability in dynamic or indoor environments. High reliance on 3D annotations and pre-trained models
[18]	Cross-Modal 360° Depth Completion and reconstruction	Introduces a novel framework for 360-degree depth completion and reconstruction, leveraging cross-modal inputs of sparse depth maps and RGB panoramic images addressing challenges of distortion and unequal sparsity in panoramic data.	This method is mainly for static already built cases, which makes it challenging to implement in dynamic construction scenarios
[19]	ORB-SLAM3	Presents a novel adaptation of the ORB-SLAM3 system to support 360-degree panoramic video, enabling autonomous positioning and orientation using a fisheye image calibration method.	Although their method provided accurate relative positioning within the environment, it is still not used to link and label the localized position with floorplans

While the methods presented on Table 1 address critical challenges in spatial data processing and accurate referencing, limitations such as dependency on pre-trained models, extensive manual labeling, computational intensity, and restricted applicability in dynamic or early construction environments highlight the need for further improvements.

Researchers such as [20] and [21] proposed datasets that can be used to address data limitations in 3D scene labeling.[20] specifically compiled the datasets to enable accurate SLAM performance in dynamic and real world construction scenarios. Although their dataset is valuable in advancing construction monitoring and automation, the dataset is mainly collected from a single construction site, pausing a limitation in generalizability. Similarly, [22] developed a large datasets containing more than 71 thousand panoramic 360 images from more than 1500 unfurnished homes. These datasets are linked with floor plans, localized with windows and doors annotations. The process of creating this dataset includes manual annotations and ground truth inputs. Hence, there is a potential of error. Moreover, implementation of this dataset involves training processes. Moreover, this dataset represents a finished home that are not subjected to dynamic change or active construction processes.

This paper addresses key limitations identified in these methods, specifically the reliance on manual data labeling, static environments, and limited support for early construction stages. By introducing a novel method that automates room-level labeling and spatial referencing of 360-degree video frames within buildings, this work bridges these gaps. Using SLAM-generated trajectories to align operator paths with 2D floor maps, the proposed approach eliminates the need for manual labeling and enhances georeferencing accuracy, streamlining data collection and labeling processes. These

advancements significantly contribute to progress monitoring and spatial data organization in dynamic construction site environments, with the goal of improving efficiency and automation in construction and building management applications.

3 Methodology

The methodology employed in this study aims to streamline and automate the process of room-level labeling for 360-degree video frames within construction sites. By leveraging advanced SLAM techniques and integrating them with floor plan data, the approach systematically combines spatial and visual information to enhance data organization and reduce manual effort. The methodology is represented in a Business Process Model and Notation (BPMN) diagram, as shown in Fig. 1, which outlines the sequential steps of the process. The different steps used in the methodology are described in more detail in the following subsections.

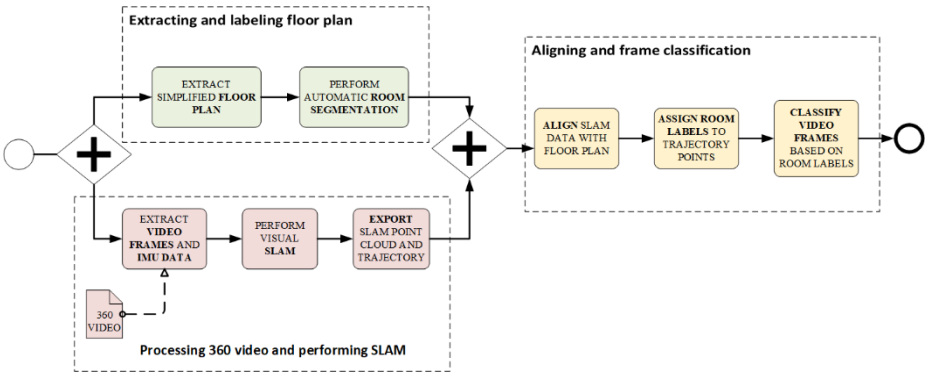


Fig. 1. BPMN diagram of the overall methodology

3.1 Processing 360 video and generating SLAM outputs

The initial step in our methodology focuses on processing the 360-degree video footage and associated Inertial Measurement Unit (IMU) data to reconstruct a three-dimensional (3D) point cloud and generate a detailed trajectory of the camera's movement throughout the environment. The process begins by splitting the 360-degree footage into discrete frames for each recorded position. Each position includes six frames corresponding to the different perspectives of the 360-degree imagery. The frame extraction rate, an adjustable input parameter, determines the number of frames processed per second. This parameter directly affects the density of the resulting point cloud and the granularity of the trajectory.

The SLAM algorithm integrates the extracted frames and IMU data to construct the 3D point cloud, which represents the spatial structure of the environment, and a camera trajectory that traces the operator's path. The trajectory comprises a sequence of positional data points, where each point corresponds to a camera position and contains the six associated 360-degree frames. These outputs form the foundation for subsequent

processing stages, as the point cloud will later be aligned with the floor plan to establish spatial references. Simultaneously, the trajectory will serve as the basis for automatic labeling of video frames.

This step necessitates balancing computational efficiency and accuracy. Higher frame extraction rates yield denser point clouds and more detailed trajectories but increase computational demands. Conversely, lower frame rates reduce processing requirements at the expense of spatial resolution. By carefully selecting this parameter, the methodology ensures that the outputs are both accurate and computationally feasible, enabling efficient downstream processing.

3.2 Extracting simplified floor plan and performing room segmentation

The next step involves generating a simplified floor plan derived from the Building Information Model (BIM). This floor plan retains only the structural outlines of the building, focusing on the boundaries of enclosed spaces, such as rooms. This simplified representation is crucial for minimizing computational complexity while ensuring sufficient detail for subsequent alignment processes.

Once the floor plan is extracted, an automated segmentation process identifies each enclosed space and assigns it a unique identifier (i.e., a label). These labels can be further aligned with predefined room identifiers provided by the construction firm or project developer, ensuring consistency with existing documentation and nomenclature. While this labeling process is automated, the methodology allows for manual refinement if necessary, providing flexibility to address any ambiguities or inconsistencies in the initial segmentation.

Importantly, this segmentation and labeling process is performed only once during the project's setup phase. Once completed, the labeled floor plan becomes a stable reference for subsequent tasks, including the alignment of SLAM-generated trajectories and the classification of video frames.

3.3 Aligning SLAM data with the floor plan and assigning room labels

In this step, the reconstructed 3D point cloud and the operator's trajectory are aligned with the simplified floor plan to establish a consistent spatial reference system. Initially, the SLAM-generated point cloud, which represents the site's layout, is superimposed onto the simplified floor plan. This alignment ensures that the spatial coordinates of the point cloud correspond to the reference system of the floor plan. A sufficiently dense point cloud is essential for accurately representing the site's structure and achieving a reliable alignment.

Once the point cloud is aligned, the same transformation is applied to the SLAM-generated trajectory. This process maps the operator's path (composed of discrete positions, each containing six 360-degree frames) onto the labeled floor plan. With the trajectory superimposed on the labeled floor plan, each position's coordinates are matched to the corresponding room label from the segmentation process. Consequently, the room label is associated with all six frames at each position.

This approach eliminates the need for manual labeling or predefined trajectories during data collection, significantly streamlining the workflow. By automating the assignment of spatial labels to video frames, the system ensures accurate and efficient room-level data organization, enabling users to focus on analysis and evaluation rather than administrative tasks.

4 Experimentation

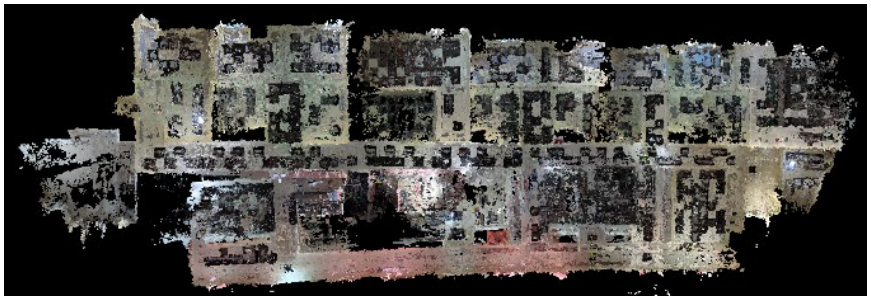
The experimentation was conducted at an ongoing construction site located within a university campus. The site is currently undergoing repurpose work, which involves creating new partitions to transform the existing space for updated functionality. The construction area spans approximately 966 $m^{2,}$ providing a varied environment suitable for testing the proposed methodology by performing progress monitoring walks with the 360 camera three times per week. The site's dynamic nature, with ongoing modifications and structural changes, presents an ideal scenario for evaluating the system's robustness and adaptability to real-world conditions.

4.1 Processing 360 video and generating SLAM outputs

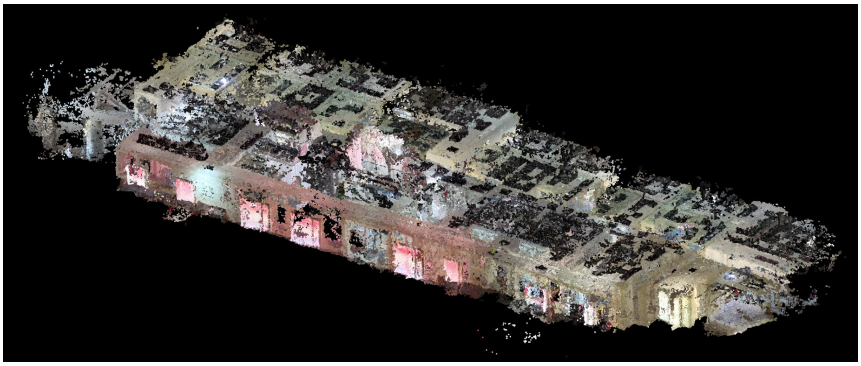
To reconstruct the 3D point cloud and generate the camera trajectory, we processed 360-degree video footage and associated IMU data collected from the experimentation site. The video was collected using an Insta360 X4 camera, recorded at 4K resolution and 100fps. The video was split into discrete frames based on a configurable frame rate of 25 frames per second, a parameter which was selected to balance computational efficiency and spatial resolution. Using the SLAM algorithm Stella Vslam Dense [23], the visual data and IMU readings were integrated to produce two key outputs: a dense 3D point cloud representing the spatial layout of the site, and a trajectory mapping the camera's movement throughout the space.

As shown in Fig. 2, the reconstructed point cloud provides a detailed enough representation of the site's structural elements, enabling accurate alignment with the simplified floor plan in subsequent steps.

Similarly, the trajectory, illustrated in Fig. 3, demonstrates the operator's path through the construction site. Each position along the trajectory corresponds to a set of coordinates and contains the six 360-degree frames previously extracted. The density of the point cloud and the granularity of the trajectory were directly influenced by the frame extraction rate, allowing the level of detail to be adjusted according to the needs of the application.



(a)



(b)

Fig. 2. (a) Top view and (b) perspective view of the SLAM reconstructed point cloud

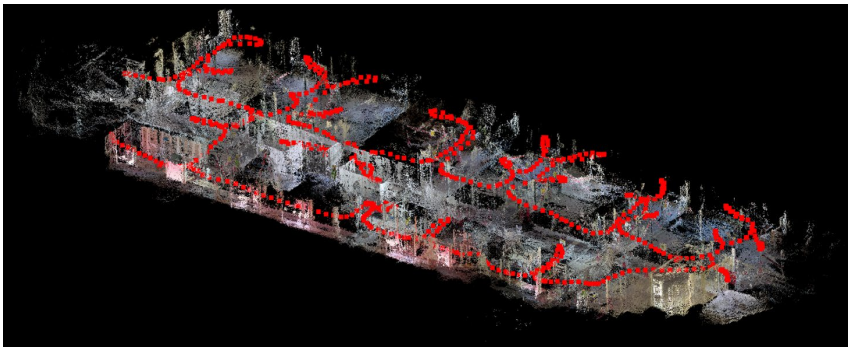


Fig. 3. Reconstructed operator's trajectory (in red) superimposed to the point cloud (ceiling has been removed for visualization purposes).

4.2 Extracting simplified floor plan and performing room segmentation

For this stage, the simplified floor plan was derived directly from the Building Information Model (BIM) of the experimentation site. The extraction process involved simplifying the original BIM-derived floor plan by removing non-structural elements such as furniture, fixtures, and other details, retaining only the layout of walls and partitions. The resulting floor plan, shown in Fig. 4, provides a clean and efficient representation of the site's structural boundaries, essential for the subsequent labeling and alignment processes.

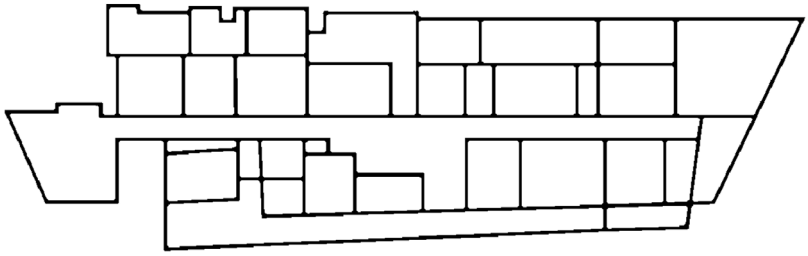


Fig. 4. Simplified floorplan extracted from the BIM.

The simplified floor plan was then processed using a segmentation algorithm implemented in MATLAB to perform automatic room segmentation and labeling. This algorithm identifies enclosed shapes within the floor plan, corresponding to individual rooms or spaces, and assigns a unique identifier to each. The labeling process relied on geometric analysis of the floor plan, detecting closed polygons to ensure comprehensive segmentation of all enclosed spaces. The output of this process is shown in Fig. 5, where each identified room is distinctly labeled.

Fig. 5. Segmented rooms out of the simplified floor plan. Each room is assigned a different number and color for visualization purposes.

4.3 Aligning SLAM data with the floor plan and assigning room labels

The final stage involves processing the generated point cloud and trajectory to align them with the labeled floor plan, enabling automatic labeling of the trajectory points.

The point cloud, which is not aligned with any axis, undergoes principal component analysis (PCA) in MATLAB to determine its primary directions. The principal components are identified using the eigenvectors of the point cloud, with the largest one used to align the point cloud with the XY plane. This step ensures that the point cloud is properly oriented with the horizontal plane, facilitating further processing.

The same transformation is applied to the trajectory points, aligning them consistently with the adjusted point cloud. Next, a slice of the point cloud is extracted to remove extraneous elements such as the ceiling, floor, and clutter, leaving only the structural layout of the walls. Both the trajectory and the processed point cloud are then projected onto a 2D plane (Fig. 6), making them ready for alignment with the simplified floor plan generated earlier.

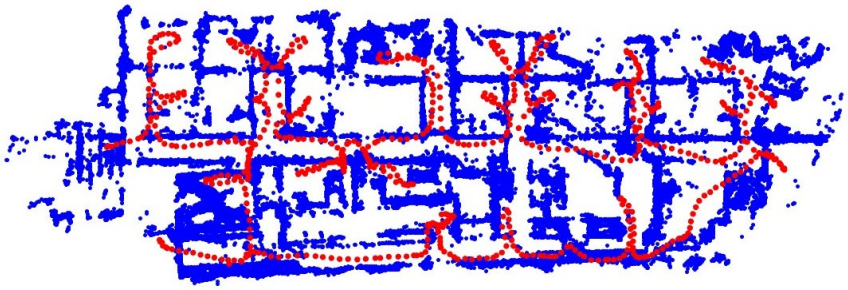


Fig. 6. Projected trajectory (in red) and sliced point cloud (in blue) onto the XY plane.

The projection is aligned with the labeled floor plan by matching the structural elements visible in both datasets (Fig. 7). This alignment ensures that the trajectory is accurately superimposed on the floor plan, with each trajectory point corresponding to a specific room or space. Based on the 2D coordinates of the trajectory points, room labels are assigned automatically by referencing the labeled floor plan.

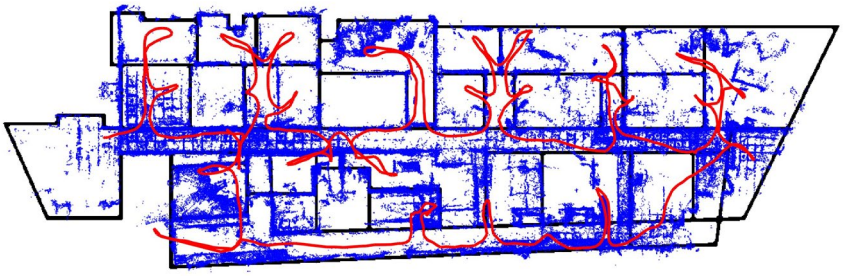


Fig. 7. Projected point cloud (in blue) and trajectory (in red) aligned with the simplified floor plan (in black).

The final result, shown in Fig. 8, demonstrates the labeled trajectory points, completing the process of automatic room-level labeling for the 360-degree video frames.

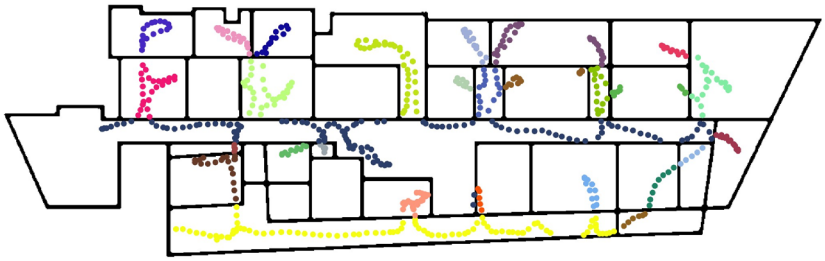


Fig. 8. Final labeled trajectory points.

4.4 Discussion and limitations

One of the primary challenges encountered in this methodology is the inaccuracy introduced by the SLAM algorithm. Visual SLAM techniques are inherently prone to cumulative drift, particularly in environments with repetitive or sparse visual features, where the algorithm struggles to maintain precise localization. This drift can result in misalignment between the trajectory and the reference floor plan, as evidenced in Table 2, which shows the number of trajectory points associated with each room. Each trajectory point corresponds to six 360-degree frames (top, bottom, front, back, left, right), providing a measure of how well-defined each visited room is. While room 21 lacks data because it was not visited by the operator, rooms like 19 and 14 exhibit gaps due to trajectory drift, where SLAM inaccuracies caused some points to be incorrectly mapped outside the room boundaries. Drift inaccuracies are introduced at the end of the path most likely due to the fact that room 2 is a long corridor with similar features, resulting in the SLAM algorithm not being as precise as in the rest of the trajectory where the operator visited multiple rooms with distinctive features. Although addressing these inaccuracies is beyond the scope of this paper, it highlights an important limitation that future work could address, potentially through drift correction techniques or more robust SLAM algorithms.

Table 2. Summary of the amount of trajectory points per room.

Room	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Trajectory points	88	6	0	0	29	129	5	11	8	5	14
Room	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
Trajectory points	19	0	10	5	4	24	0	29	0	34	8
Room	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
Trajectory points	16	12	15	12	21	13	30	10	17	16	13

Another limitation lies in the reliance on the structural features within the point cloud for accurate alignment with the simplified floor plan. Early stages of the construction could present environments with minimal features, incomplete structures, or significant clutter, that may reduce alignment reliability. Moreover, the methodology depends on

the initial quality and accuracy of the simplified floor plan extracted from the BIM model. Errors or ambiguities in the floor plan, such as missing or mislabeled spaces, can propagate through the system and affect the reliability of room labeling.

Despite these challenges, the methodology demonstrates practical advantages. By automating the association of 360-degree video frames with their respective rooms, it eliminates the need for labor-intensive manual labeling. In traditional workflows, operators either follow predefined routes to systematically capture images in every room or manually assign collected data to rooms afterward. Both approaches are time-consuming and prone to disruption if site conditions change or rooms are inaccessible. In contrast, our approach allows operators to freely walk through the site while recording a continuous video, with frames automatically associated with their corresponding rooms. This significantly reduces time and effort while ensuring immediate access to labeled data.

A further benefit is the system's ability to map data to as-planned spaces, even if construction is incomplete. For instance, rooms 10 and 11, which lack a separating wall at the time of data collection, are still distinctly labeled based on the as-planned floor plan. This capability is particularly valuable for tracking progress in dynamic construction environments, where spaces may evolve over time but still need to be monitored.

4.5 Conclusion and future work

This paper introduced a methodology for automating room-level labeling of 360-degree video frames using SLAM-generated trajectories and alignment with a simplified floor plan. By eliminating the need for manual data labeling, this approach significantly enhances the efficiency and accuracy of spatial data organization in construction site monitoring. The results demonstrated the potential of the proposed system to streamline progress monitoring and building management workflows while addressing common challenges such as misalignment and labeling inefficiencies. The methodology's ability to function effectively in dynamic construction environments without a need for pre-training and extra data requirements further highlights its applicability in real-world scenarios, particularly for progress tracking and data management.

Despite its advantages, the approach is not without limitations. Issues such as SLAM drift and reliance on distinct features for alignment underscore areas where the system could benefit from further refinement.

Future work will focus on addressing the identified limitations, particularly through the integration of advanced SLAM drift correction techniques and sensor fusion to improve alignment accuracy. Additionally, efforts will be made to adapt the methodology for early construction stages and environments with sparse or incomplete structural features. Expanding the system's capabilities to incorporate real-time labeling and integration with augmented reality tools will also be explored. Finally, validating the methodology across a broader range of construction sites and scenarios will ensure its generalizability and robustness, paving the way for widespread adoption in the construction industry.

Acknowledgments. This research was partially supported by different Centers at NYUAD. In particular, the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) funded by Tamkeen under the NYUAD Research Institute Award CG013, the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001, and the Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. I. Abaspur Kazerouni, L. Fitzgerald, G. Dooly, and D. Toal, "A survey of state-of-the-art on visual SLAM," *Expert Syst. Appl.*, vol. 205, p. 117734, Nov. 2022, doi: 10.1016/j.eswa.2022.117734.
2. Y. Shinde, K. Lee, B. Kiper, M. Simpson, and S. Hasanzadeh, "A Systematic Literature Review on 360° Panoramic Applications in Architecture, Engineering, and Construction (AEC) Industry," *J. Inf. Technol. Constr. ITcon*, vol. 28, no. 21, pp. 405–437, Aug. 2023, doi: 10.36680/j.itcon.2023.021.
3. A. Eryomin, R. Safin, T. Tsoy, R. Lavrenov, and E. Magid, "Optical Sensors Fusion Approaches for Map Construction: A Review of Recent Studies," *J. Robot. Netw. Artif. Life*, vol. 10, no. 2, pp. 127–130, 2023, doi: 10.57417/jrnal.10.2_127.
4. S. A. Prieto, E. T. Mengiste, U. Menon, and B. G. de Soto, "Current State and Trends of Point Cloud Segmentation in Construction Research," *Int. Symp. Autom. Robot. Constr. ISARC Proc.*, vol. 2024 Proceedings of the 41st ISARC, Lille, France, pp. 972–979, Jun. 2024, doi: 10.22260/ISARC2024/0126.
5. X. Zhang, Z. Fang, Z. Lu, J. Xiao, X. Cheng, and X. Zhang, "3D Reconstruction of Weak Feature Indoor Scenes Based on Hector SLAM and Floorplan Generation," in *2021 IEEE 7th International Conference on Virtual Reality (ICVR)*, May 2021, pp. 117–126. doi: 10.1109/ICVR51878.2021.9483856.
6. J. D. Tascón -Vidarte, "Floor plans from 3D reconstruction of indoor environments," in *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, Aug. 2016, pp. 1–7. doi: 10.1109/STSIVA.2016.7743310.
7. D. Farin, W. Effelsberg, and P. H. N. de With, "Floor-plan reconstruction from panoramic images," in *Proceedings of the 15th ACM international conference on Multimedia*, in MM '07. New York, NY, USA: Association for Computing Machinery, Sep. 2007, pp. 823–826. doi: 10.1145/1291233.1291420.
8. B. Solarte, Y.-C. Liu, C.-H. Wu, Y.-H. Tsai, and M. Sun, "360-DFPE: Leveraging Monocular 360-Layouts for Direct Floor Plan Estimation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6503–6510, Jul. 2022, doi: 10.1109/LRA.2022.3173730.
9. H. Kim, L. Remaggi, S. Fowler, P. J. Jackson, and A. Hilton, "Acoustic Room Modelling Using 360 Stereo Cameras," *IEEE Trans. Multimed.*, vol. 23, pp. 4117–4130, 2021, doi: 10.1109/TMM.2020.3037537.
10. S. Chen, M. Li, K. Ren, and C. Qiao, "Crowd Map: Accurate Reconstruction of Indoor Floor Plans from Crowdsourced Sensor-Rich Videos," in *2015 IEEE 35th International Conference on Distributed Computing Systems*, Jun. 2015, pp. 1–10. doi: 10.1109/ICDCS.2015.9.

11. C. Sun *et al.*, "Seg2Reg: Differentiable 2D Segmentation to 1D Regression Rendering for 360 Room Layout Reconstruction," Nov. 30, 2023, *arXiv*: arXiv:2311.18695. doi: 10.48550/arXiv.2311.18695.
12. X. Ji, P. Liu, H. Niu, X. Chen, R. Ying, and F. Wen, "Object SLAM Based on Spatial Layout and Semantic Consistency," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023, doi: 10.1109/TIM.2023.3316258.
13. A. Braun and A. Borrmann, "Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning," *Autom. Constr.*, vol. 106, p. 102879, Oct. 2019, doi: 10.1016/j.autcon.2019.102879.
14. J. Xiao, A. Owens, and A. Torralba, "SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels," in *2013 IEEE International Conference on Computer Vision*, Dec. 2013, pp. 1625–1632. doi: 10.1109/ICCV.2013.458.
15. C. Tao, Z. Gao, J. Yan, C. Li, and G. Cui, "Indoor 3D Semantic Robot VSLAM Based on Mask Regional Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 52906–52916, 2020, doi: 10.1109/ACCESS.2020.2981648.
16. C. Theodorou, V. Velisavljevic, V. Dyo, and F. Nonyelu, "Visual SLAM algorithms and their application for AR, mapping, localization and wayfinding," *Array*, vol. 15, p. 100222, Sep. 2022, doi: 10.1016/j.array.2022.100222.
17. X. Fu *et al.*, "Panoptic NeRF: 3D-to-2D Label Transfer for Panoptic Urban Scene Segmentation," in *2022 International Conference on 3D Vision (3DV)*, Sep. 2022, pp. 1–11. doi: 10.1109/3DV57658.2022.00042.
18. R. Liu, G. Zhang, J. Wang, and S. Zhao, "Cross-Modal 360° Depth Completion and Reconstruction for Large-Scale Indoor Environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 25180–25190, Dec. 2022, doi: 10.1109/TITS.2022.3155925.
19. X. Zhao, Q. Hu, X. Zhang, and H. Wang, "An ORB-SLAM3 Autonomous Positioning and Orientation Approach using 360-degree Panoramic Video," in *2022 29th International Conference on Geoinformatics*, Aug. 2022, pp. 1–7. doi: 10.1109/Geoinformatics57846.2022.9963855.
20. M. Trzeciak *et al.*, "ConSLAM: Periodically Collected Real-World Construction Dataset for SLAM and Progress Monitoring," in *Computer Vision – ECCV 2022 Workshops*, L. Karlinsky, T. Michaeli, and K. Nishino, Eds., Cham: Springer Nature Switzerland, 2023, pp. 317–331. doi: 10.1007/978-3-031-25082-8_21.
21. P. Selvaraju *et al.*, "BuildingNet: Learning To Label 3D Buildings," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10397–10407. Accessed: Jan. 24, 2025. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Selvaraju_BuildingNet_Learning_To_Label_3D_Buildings_ICCV_2021_paper.html
22. S. Cruz, W. Hutchcroft, Y. Li, N. Khosravan, I. Boyadzhiev, and S. B. Kang, "Zillow Indoor Dataset: Annotated Floor Plans With 360° Panoramas and 3D Room Layouts," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 2133–2143. doi: 10.1109/CVPR46437.2021.00217.
23. H. Surmann, M. Thurow, and D. Slomma, "PatchMatch-Stereo-Panorama, a fast dense reconstruction from 360° video images," Nov. 29, 2022, *arXiv*: arXiv:2211.16266. doi: 10.48550/arXiv.2211.16266.