3D indoor modeling with low-cost RGB-D camera and iPad Pro: A Comparative Study

Ishraq Rached[1] ,Badr ed-dine Ouzougarh[2],Hassan Radoine[3], Rafika Hajji[4],Rashid Haffadi[5]

[1]College of Geomatic Sciences and Surveying Engineering, Institute of Agronomy and Veterinary Medicine, Rabat 6202, Morocco - Ishraq.rach@gmail.com
[2]College of Geomatic Sciences and Surveying Engineering, Institute of Agronomy and Veterinary Medicine, Rabat 6202, Morocco - ouzougarhbadred-dine@iav.ac.ma
[3]Mohammed VI Polytechnic University, UM6P-CITINNOV SA, Rabat, Morocco -hassan.radoine@um6p.ma
[4]College of Geomatic Sciences and Surveying Engineering, Institute of Agronomy and Veterinary Medicine, Rabat 6202, Morocco - r.hajji@iav.ac.ma
[5]GEOPTIMA, B4, Med El Amraoui Street, Corner of Sebou Street, Office 4, K´enitra, Morocco - rashidhaffadi@gmail.com

Keywords: RGB-D sensor, iPad Pro LiDAR, Kinect Azure, Indoor 3D reconstruction, 3D modeling.

Abstract

The rapid advancement in indoor 3D building modeling has led to increased interest in low-cost solutions for 3D data acquisition. While Terrestrial Laser Scanning (TLS) and Mobile Mapping Systems (MMS) produce detailed 3D models, their high cost and complex workflows make them impractical for many applications. In this paper, we investigate the effectiveness of using low-cost sensors, specifically RGB-D camera and iPad Pro for 3D modelling. Through a series of experiments, we evaluate these devices in terms of data accuracy, processing speed, and qualitative analysis using 3D point clouds and heat map visualizations, comparing the results with MMS data as the ground truth. Three distinct environments: an office room, a corridor, and a staircase were scanned to assess performance across varying levels of scene complexity. The results show that both devices are effective for indoor 3D modeling, but the RGB-D camera was more accurate, with an average C2C distance of 0.0245 meters compared to the iPad's 0.0465 meters. However, the iPad Pro was faster, completing scans 30% quicker, making it better suited for tasks that require speed over precision.

1. Introduction

In recent years, 3D modeling using low-cost devices has gained growing interest. Numerous authors have proposed studies using RGB-D sensors for indoor 3D modeling. For instance,(Wang et al., 2012)introduced a method that uses multiple structured light-based depth cameras, (Henry et al., 2012) focused on dense 3D modeling of indoor environments. Their contributions served as the cornerstone for employing RGB-D devices in the field of building reconstruction. (Li et al., 2020)further advanced this research area by automating the generation of as-built Building Information Models (BIM) using RGB-D devices. Such automation simplifies the process of creating detailed models for architectural and construction purposes. More recent research has continued to focus on the potential of RGB-D devices for 3D building reconstruction.(Zhou et al., 2022) and (Wahbeh, 2021) explored methods to improve model accuracy in complex environments, while (Delasse et al., 2022) and (El Haouss et al., 2022) examined the potential and limitations for RGB-D sensors. (Rached et al., 2024) highlighted the importance of enhancing model robustness, particularly when reconstructing intricate details in challenging environments. These studies collectively emphasize that while low-cost sensors show significant promise in 3D building reconstruction, more research is needed to enhance the precision and robustness of the generated models. There has also been an increasing interest in combining the advantages of multiple low-cost sensors, such as RGB-D devices and iPad equipped with LiDAR, to complement one another in 3D modeling tasks. For example, IPad Pro's LiDAR capabilities have been explored by (Ancona et al., 2015) in the context of mobile, user-friendly devices for interior design applications. Other low-cost alternatives, such as panoramic multi-camera systems, have proven to be effective for photogrammetric modeling of indoor spaces. These systems offer the advantage of capturing 360-degree views, so reducing the number of images needed to cover a scene (Barmpoutis et al., 2020). Although panoramic cameras have been used for indoor modeling, RGB-D cameras, which combine depth information with RGB images, offer a more comprehensive solution for generating textured 3D models without the need for key points. Furthermore combining data from TLS and RGB-D devices can improve the quality of 3D models, especially in cases where the TLS cannot access certain areas (Yang et al., 2018). While low-cost sensors provide ease of use and reduce cost, additional data processing is often required to achieve a finished 3D model. Traditionally, manual modeling techniques were necessary to produce high-accuracy models, but automatic approaches, such as surface-based methods, volumetric primitives, and shape grammars (Yang et al., 2018), have gained prominence for their efficiency and cost-effectiveness. The advancements in deep learning and automated reconstruction techniques have also contributed to this field. Neural networks have been applied to 3D depth estimation from single images , and methods like 2D semantic segmentation and 3D volumetric segmentation are increasingly used for RGB-D data modeling. Despite these advancements, challenges such as occlusions, lack of semantic information, and varying point densities remain key issues in automatic 3D modeling, particularly in indoor environments (Yang et al., 2018). The literature suggests that RGB-D cameras and other low-cost sensors hold significant potential for 3D modeling applications. However, complex architectural environments are still a challenge when using these sensors. Therefore combining the strengths of different sensors, such as RGB-D cameras and LiDAR-equipped devices, could lead to more accurate and complete 3D models, offering an accessible and cost-effective solution for various modeling tasks. In this paper, we aim to fill this gap by conducting a comparative analysis between the Microsoft Kinect Azure camera (RGB-D device) and iPad Pro (equipped with a LiDAR sensor). To this end, we evaluate the performance of both devices across different indoor environments with varying complexities, such as office rooms, corridors, and staircases. By analyzing their accuracy, processing speed, and overall model quality, we aim to determine whether these low-cost devices can be used complementarily or if one outperforms the other in certain contexts. This paper provides valuable insights into the feasibility of using these tools for 3D indoor modeling and contributes to improving the quality and robustness of models generated by low-cost sensors. The remainder of this paper is organized as follows: Section 2 presents the methodology, section 3 presents the experiment which encompasses the case study project, data acquisition protocols, devices used for scanning, the data processing workflow, and the 3D reconstruction process. Section 4 focuses on the results and discussion, providing a comparative analysis of the performance of each device. Finally, Section 5 concludes the study and outlines potential directions for future research.

2. Methodology

The methodological workflow, as illustrated in Figure 1, outlines the steps to evaluate and compare the performance of the RGB-D Camera and the iPad Pro LiDAR for 3D building reconstruction. The workflow begins with the data acquisition phase, where both devices capture 3D data in three distinct environments: an office room, corridor, and staircase. In the data processing phase, the RGB-D data is processed using Open3D to extract RGB and depth images, while the iPad Pro data is processed through the Sitescape, a mobile application that leverages the iPad Pro's LiDAR capabilities to capture and export high-resolution 3D point clouds for use in modeling and analysis. The 3D reconstruction phase involves generating point clouds from both data sources, followed by the comparison and analysis phase, where the generated point clouds are evaluated and compared to the MMS ground truth (MS-96). Both quantitative and qualitative assessments are performed, including the calculation of mean and standard deviation of Cloud-to-Cloud (C2C) distances, as well as a heat map visualization for model accuracy. Lastly, processing speeds are compared, and the results are analyzed to draw conclusions regarding the accuracy, speed, and overall quality of the models produced by both devices.

3. Experiments

In this study, we focus on three distinct environments as shown in Figure 2, each representing varying levels of complexity: an office room filled with equipment, a corridor with architectural features, and a staircase with intricate spatial arrangements. These environments were chosen to reflect common indoor settings where accurate 3D modeling is essential for applications such as facility management, renovation planning, and virtual tours. The office room served as a controlled environment, allowing for the assessment of the devices' capabilities in capturing detailed features, such as furniture and equipment, within a confined space. The corridor was selected for its linear structure, providing an opportunity to evaluate the devices' performance in a longer, narrow space with limited visibility. Finally, the staircase was included to challenge the devices with vertical elements and varying perspectives, requiring effective handling of occlusions and depth variations.

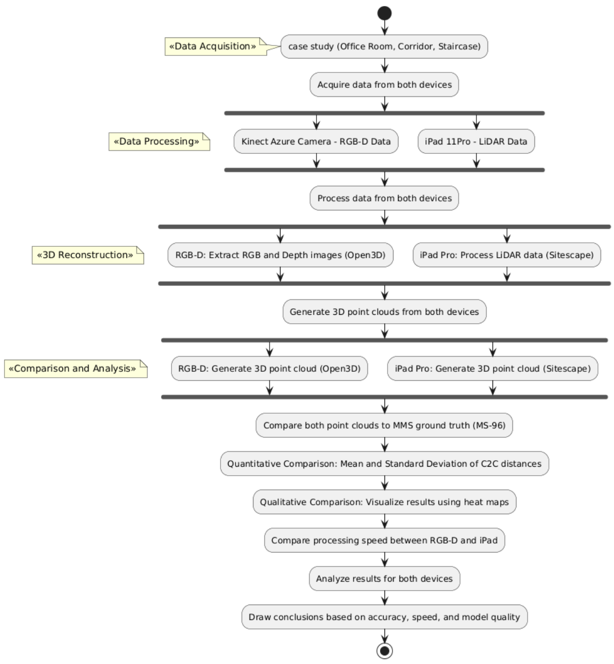

Figure 1. "The Methodological workflow ".

Figure 2. Selected indoor environments for 3D modeling - Office Room, corridor, and staircase.

3.1 Data acquisition

For this study, two low-cost devices were used for data acquisition: the Microsoft Kinect Azure RGB-D device and the iPad Pro 11, equipped with a LiDAR sensor and dual RGB cameras. Both devices captured RGB images. Additionally, a Mobile Mapping System (MMS) MS-96 Mobile Mapping, Viametris was used as a reference for comparing the accuracy and quality of the point clouds generated by the two low-cost devices. The MMS provided high-precision 3D data, which served as a benchmark for evaluating the performance of the Kinect Azure and iPad Pro. Detailed specifications for each device are presented in Table 1.

3.1.1 Microsoft Kinect Azure RGB-D aamera The Kinect Azure, the successor to the widely-used Kinect v2, represents a significant advancement in 3D modeling and indoor mapping technologies. This device integrates a high-resolution

RGB camera with a depth sensor, allowing for the simultaneous capture of color and depth data. This capability is crucial for creating detailed 3D models, aligning with the acquisition pipeline described by (Bernardini and Rushmeier, 2002), which emphasizes the importance of capturing both geometric and appearance data. Traditionally, 3D modeling methods required separate processes for acquiring geometry and texture, leading to increased complexity and costs. The introduction of RGB-D cameras, like the Kinect Azure, has streamlined this process, making it more accessible and affordable. Researchers such as(Zollhöfer et al., 2018) have explored the advancements and challenges of RGB-D technology, highlighting its expanding role across various fields. The Kinect Azure's ability to effortlessly gather comprehensive 3D data has unlocked new opportunities for applications such as indoor mapping and 3 modeling (Rached et al., 2024) and (El Haouss et al., 2022), and augmented reality, as noted in the survey (Xu et al., 2019).

3.1.2 Ipad 11 Pro The iPad Pro, equipped with a LiDAR sensor and dual RGB cameras, offers a mobile and cost-effective solution for 3D model generation, as shown in research conducted by (Teo and Yang, 2023), the iPad Pro's LiDAR sensor has demonstrated effectiveness in generating highly accurate point clouds, with minimal deviations when scanning is performed with proper distance and stability. Furthermore,(Ingman et al., 2020) compared various low-cost sensor systems, including an RGB-D device, a low-end terrestrial laser scanner, and a panoramic camera, for automatic cloud-based indoor 3D modeling. Their findings emphasized the potential of LiDAR-equipped devices like the iPad Pro in providing accurate and efficient 3D data acquisition for a range of applications.

3.1.3 MS-96 Viametris To establish a robust ground truth reference for evaluating the low-cost devices, this study utilized an MS-96 Mobile Mapping, Viametris. The system typically comprises a combination of high-accuracy laser scanners, inertial measurement units, and high-resolution cameras. During data acquisition, the MMS is mounted on a mobile platform and moved through the environment, systematically capturing point cloud data with precise positional information.

Type of device	Sensing technology	Range (m)	RGB resolution	Field of view
Kinect Azure	Structured light	0.5–5	1920x1080 pixels	120°
iPad Pro 11	Time-of-flight (LiDAR)	0.2–5	12 MP (wide), 10 MP (ultra-wide)	70° LiDAR, 120° ultra-wide RGB camera
MS-96 Viametris	Laser scanner + GNSS/IMU	0–120	4×24 MP	360° scanning

Table 1. Technical characteristics of used sensor systems

3.2 Scanning and data processing

We captured videos using the Kinect Azure camera, systematically scanning the walls and details by moving in a unidirectional manner and incorporating upward movements to ensure comprehensive coverage of the floor and ceiling. Similarly, for the iPad Pro, the LiDAR sensor was used to scan the environment with a handheld motion, covering the key areas from various angles to fill in data gaps and ensure detailed depth capture.

After the data collection phase for both devices, initial preprocessing steps were applied, including the removal of noise and outliers from the generated point clouds. This step was essential in improving the quality of the 3D models by eliminating artifacts that may have been introduced due to environmental factors or minor inconsistencies in the scanning process. These preprocessing actions helped enhance the accuracy and reliability of the point clouds from both the Kinect Azure and the iPad Pro.

3.2.1 RGB-D image processing The raw data consists of a video stream that combines RGB images and depth maps. To begin the processing workflow, we decomposed the captured video into its constituent frames, each containing an RGB image paired with a depth map. Once the frames were extracted, depth maps were aligned with the RGB images to form a consistent dataset. We used Open3D python library's for 3D reconstruction (Zhou et al., 2018)

3.2.2 iPad data processing For the data acquired using the iPad Pro 11, we utilized SiteScape. SiteScape enables the efficient import and refinement of point cloud data collected via the iPad's LiDAR sensor, allowing for noise reduction and filtering to enhance data quality. The software provides interactive 3D visualization, enabling users to explore the captured environment and identify areas needing attention.

3.3 3D reconstruction

3D reconstruction from RGB-D data is performed using open3D (Zhou et al., 2018). The reconstruction process consists of four key steps:
1.Make fragment: build local geometric surfaces (referred to as fragments) from short subsequences of the input RGBD sequence. This part uses RGBD Odometry, Multiway registration, and RGBD integration.
2.Register fragments: the fragments are aligned in a global space to detect loop closure. This part uses Global registration, ICP registration, and Multiway registration.
3.Refine registration: the rough alignments are aligned more tightly. This part uses ICP(Iterative closest point) registration, and Multiway registration.
4.Integrate scene: integrate RGB-D images to generate a mesh model for the scene. This part uses RGBD integration. For the iPad Pro data, the 3D reconstruction was handled using SitScape, which processes the LiDAR data from the iPad to generate point clouds.

3.4 Model evaluation

Following 3D reconstruction, each model underwent a rigorous evaluation to assess its quality and accuracy. Key metrics included:
Model quality: Quantitative metrics were employed to evaluate model accuracy by comparing the processed models against the MMS ground truth data. Cloud-to-cloud (C2C) Distance Analysis was used to compute the distances between the constructed point clouds and the MMS ground truth point cloud. From this analysis, key metrics were extracted to quantify the accuracy of the point cloud reconstructions. Specifically, we focused on the following metrics:
•**Mean C2C Distance:** This metric represents the average deviation between corresponding points in the two point clouds. A lower mean C2C distance indicates higher geometric precision and better alignment with the MMS ground truth.
•**Standard Deviation of C2C Distances:** This metric captures

the variability of the distances, providing insight into the consistency of the reconstruction. A lower standard deviation suggests more uniform accuracy across the model.

The standard deviation (σ) of the C2C distances was computed using the formula:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \mu)^2}$$

where N is the total number of points in the point cloud, d_i represents the individual C2C distance for the i-th point, and μ is the mean of the C2C distances, calculated as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i$$

4. Results and analysis

In this section, we present the findings from our comparative study of the RGB-D camera and the iPad Pro for indoor 3D modeling. We analyze the performance of both devices in terms of data accuracy, processing speed, and model quality across three distinct environments: an office room, a corridor, and a staircase.

4.1 Quantitative assessment

After aligning the point clouds captured by the devices with the MMS ground truth, a Cloud-to-Cloud (C2C) distance analysis was conducted. The C2C distances were calculated to measure the deviations between the model and the ground truth, helping to assess the accuracy of each device in capturing the geometry of the environment. The histograms for the Cloud-to-Cloud (C2C) absolute distances as shown in Figure 3 provide a clear representation of the average deviations between point clouds captured by both the iPad Pro LiDAR and RGB-D camera systems across different environments. In terms of C2C distances, the RGB-D camera consistently shows lower mean values, indicating higher accuracy in aligning point clouds with minimal discrepancies. For instance, in both the office and stairway environments, the RGB-D camera achieves a mean C2C distance of around 0.0245 m, demonstrating its ability to generate accurate point clouds that are closely aligned with the reference. On the other hand, the iPad Pro LiDAR exhibits higher mean C2C distances, ranging from 0.0465 m in the office and stairway environments to 0.0511 m in the corridor. These higher mean values suggest that the iPad Pro is less precise, producing larger deviations in the captured data, which may result in less accurate point cloud models when compared to the RGB-D camera. However, the iPad LiDAR still captures a significant portion of the data within an acceptable range, making it useful for capturing larger areas quickly, even with more deviations.

The analysis of standard deviation in figure 4 provides insight into the spread or variability of the C2C distances, reflecting how consistently each device captures point cloud data. Across the board, the RGB-D camera shows lower standard deviations, indicating a tighter clustering of points around the mean. This suggests that the RGB-D camera consistently captures points with minimal variation, offering a more reliable and precise representation of the environment. For example, the standard deviation in the corridor environment is around 0.0243 m, reflecting minimal variation in the point cloud, and similar tight spreads are observed in other environments. Conversely, the iPad Pro LiDAR shows higher standard deviations, with values such as 0.0375 m in the stairway environment and 0.0336 m in the corridor. These higher standard deviations indicate more variability in the captured data, meaning that the point cloud contains more outliers or points with larger deviations from the mean. This variability could be attributed to the iPad's less precise LiDAR system, which is more prone to errors in capturing fine details or complex geometries. In conclusion, the RGB-D camera provides a more consistent point cloud with fewer deviations, while the iPad, though more variable, offers broader coverage in less controlled environments.

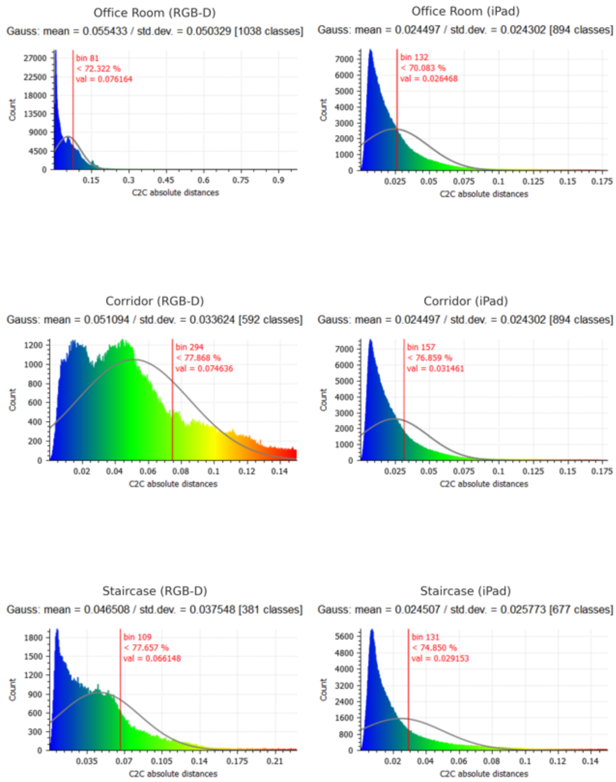

Figure 3. "Comparative analysis of C2C absolute distances for RGB-D and iPad liDAR in Office Room, corridor, and staircase environments".

4.2 Processing speed

The processing speed for 3D data acquisition and model generation was systematically evaluated for both the RGB-D camera and the iPad Pro across the three environments. The acquisition times are summarized in Table 2 below, detailing the time taken for each environment, along with the total processing times for both devices. This summary allows for a clear comparison of the efficiency of each device in capturing and processing 3D data.

As shown in Figure 5, the RGB-D Camera spends significantly more time on processing than the iPad Pro, with 25.4% of the total time allocated to processing (18 minutes), compared to 12.7% for the iPad Pro (9 minutes). Both devices

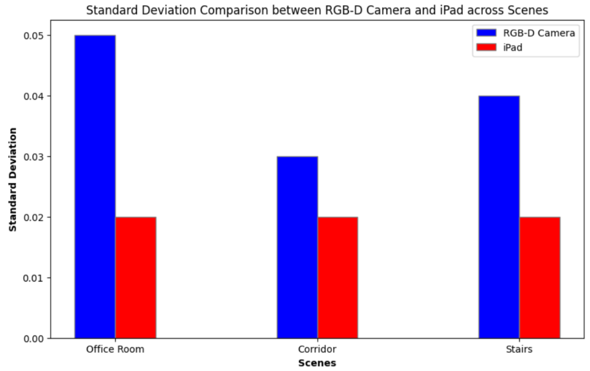

Figure 4. "Standard deviation comparison for RGB-D Camera and iPad Pro across the three environments".

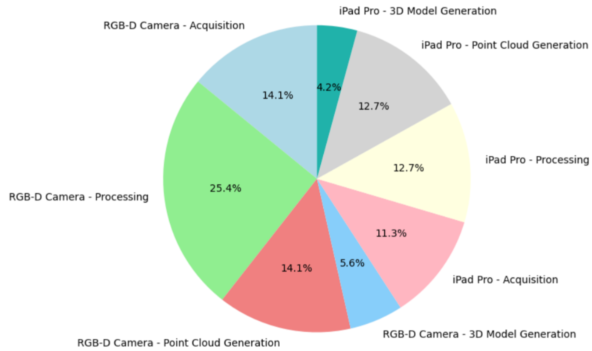

Figure 5. "Total processing time distribution for RGB-D Camera and iPad Pro across the three environments".

Device	Environment	Acquisition time (min)	Processing time (min)	Point cloud generation (min)	3D Model generation (min)	Total time (min)
RGB-D Camera	Office Room	4	6	4	2	16
	Corridor	3	6	3	1	13
	Stairs	3	6	3	1	13
iPad Pro	Office Room	4	3	3	1	13
	Corridor	2	3	3	1	9
	Stairs	2	3	3	1	9

Table 2. Acquisition, processing, and 3D model generation times for RGB-D camera and iPad Pro

have a relatively balanced distribution of time spent on acquisition and point cloud generation. For acquisition, the RGB-D Camera takes 14.1% of the total time (10 minutes), while the iPad Pro is slightly faster, spending 11.3% of the time on acquisition (8 minutes). For point cloud generation, both the RGB-D Camera and the iPad Pro allocate 14.1% and 12.7% of their time, respectively, which translates to 10 minutes for each device. The 3D model generation stage is the quickest for both devices, with the iPad Pro completing it in 4.2% of the total time (3 minutes) and the RGB-D Camera in 5.6% of the time (4 minutes). In summary, the RGB-D Camera spends more time on processing (18 minutes), enhancing precision and accuracy but making it slower overall compared to the iPad Pro. The iPad Pro is quicker across all stages, making it ideal for time-sensitive tasks, although it may not provide the same level of detail or accuracy as the RGB-D Camera.

4.3 Qualitative assessment

To enhance the visual inspection process, a heat map was employed. This heat map was generated as shown in Figures 6,7 and 8 to provide a more comprehensive view of the spatial distribution of errors across the scanned environments. The heat map visualizes the density of point-to-point distances, indicating areas where the reconstructed models from the RGB-D camera and iPad Pro deviated most significantly from the MMS ground truth. Office room case: As shown in Figure 6 in the office room scenario, the RGB-D Camera continues to show better accuracy, with most of the heat map consisting of blue and green areas, indicating that it captures the room's general structure well. Some yellow and orange spots appear around the edges of the furniture and other complex geometries, suggesting minor inaccuracies in capturing detailed features like desks and corners. However, the overall model remains fairly accurate. The iPad Pro, on the other hand, displays larger deviations, with more yellow and red areas, particularly around the furniture and along the walls. This indicates that the iPad has difficulty with fine details and sharp edges, leading to a less accurate point cloud overall, especially around the desks and other small objects. The iPad Pro's limitations in capturing intricate features are more apparent in this scenario, making its model less precise compared to the RGB-D Camera.

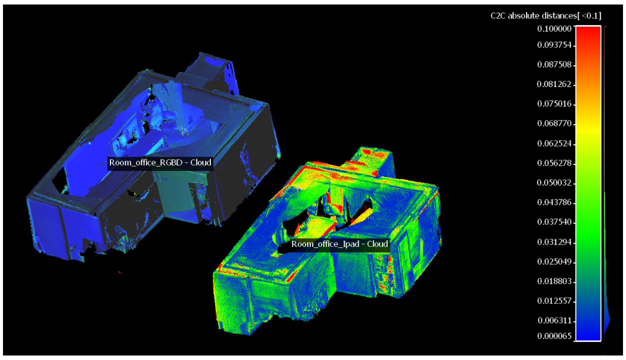

Figure 6. "Heat map for the RGB-D and iPad pro models compared to MMS ground truth in office room".

Corridor Case: As shown in Figure 7, In the corridor scenario, the RGB-D Camera again performs well, producing a heat map with mostly blue and green regions, indicating a high level of accuracy in capturing the walls and ceiling. The model is consistent, with sharp edges and minimal deviations, although there are small patches of yellow near the junctions between the ceiling and walls, where slight inaccuracies are observed. On the other hand, the iPad Pro shows more significant deviations, especially along the ceiling and wall junctions, where red and yellow areas are present. These deviations suggest that the iPad has difficulty accurately capturing the linear structures of the corridor, particularly at the boundaries where walls meet the ceiling. As a result, the iPad's point cloud is less reliable in capturing the fine details of this environment compared to the RGB-D Camera. Stairs case: In the staircase scenario as shown in Figure 8, the RGB-D Camera shows a high level of accuracy, with mostly blue and green regions on the heat map, indicating

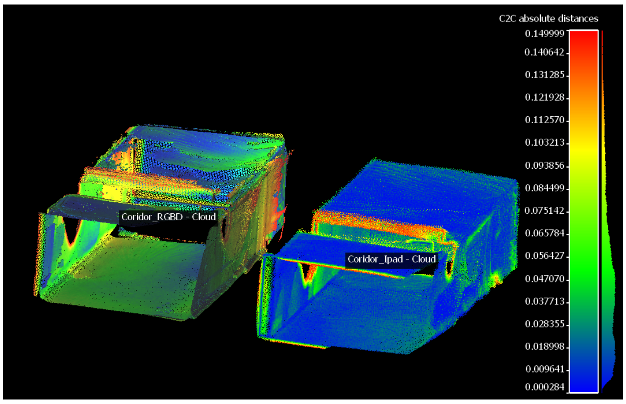

Figure 7. "Heat map for the RGB-D and iPad pro models compared to MMS ground truth in corridor".

minimal deviation between the point cloud and the reference model. The steps are well-defined, and the edges are captured with precision. Only small areas near the top of the staircase exhibit yellow and red spots, which suggests slight deviations due to either geometric complexity or possible occlusions. In contrast, the iPad Pro shows more significant deviations, with yellow and red areas appearing more prominently, especially near the top of the staircase and along the edges. The iPad struggles with edge definition, particularly where the steps meet the wall, resulting in less accurate point cloud generation overall. The larger deviations in the iPad's model indicate that it has difficulty capturing the fine details and edges in this environment.

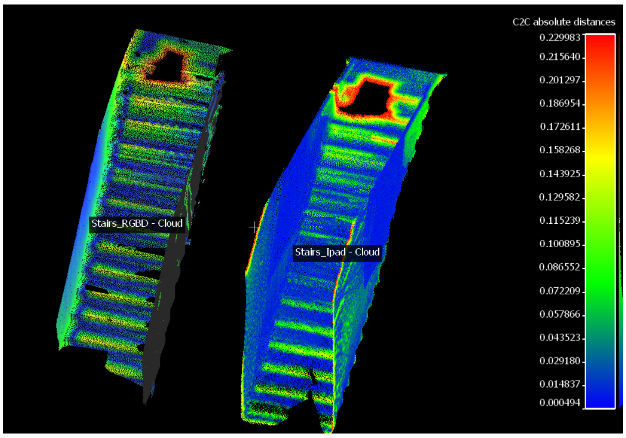

Figure 8. "Heat map for the RGB-D (a) and iPad pro (b) models compared to MMS ground truth in stairs".

5. Conclusion

This study demonstrates the strengths and limitations of the RGB-D camera and the iPad Pro LiDAR in indoor 3D modeling tasks. Through the analysis of heat maps and C2C absolute distances, it is evident that the RGB-D camera consistently outperforms the iPad Pro in terms of accuracy, especially in environments with complex geometries such as staircases and office rooms. The RGB-D camera excels at capturing sharp edges, fine details, and linear structures, making it more suitable for applications that demand high precision, such as heritage building information modeling (HBIM) or detailed architectural documentation. Conversely, the iPad Pro offers advantages in terms of speed and ease of use, performing significantly faster in acquisition and processing times across all environments. However, the iPad Pro shows larger deviations, particularly in regions with complex features, such as furniture or ceiling-wall junctions, reducing its suitability for precision-critical tasks. In conclusion, the RGB-D camera is a potential choice for projects that demand good precision while the iPad Pro may be more appropriate for quick scanning tasks where speed is prioritized over precision. Future work could explore hybrid approaches, combining the strengths of both systems to enhance overall efficiency and accuracy in 3D modeling workflows. Additionally, further research could evaluate the integration of these systems into building information modeling (BIM) platforms to streamline the process of updating or maintaining architectural models in real-time.

References

Ancona, M., Clini, P., Dellacasa, A., Falzone, P., La Camera, A., Quattrini, R., Sommariva, E., Stephens, J., 2015. Extending a mobile device with low-cost 3d modeling and building-scale mapping capabilities, for application in architecture and archaeology. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 453–459.

Barmpoutis, P., Stathaki, T., Dimitropoulos, K., Grammalidis, N., 2020. Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures. *Remote Sensing*, 12(19), 3177.

Bernardini, F., Rushmeier, H., 2002. The 3d model acquisition pipeline. *Computer graphics forum*, 21number 2, Wiley Online Library, 149–172.

Delasse, C., Lafkiri, H., Hajji, R., Rached, I., Landes, T., 2022. Indoor 3D reconstruction of buildings via azure Kinect RGB-D camera. *Sensors*, 22(23), 9222.

El Haouss, N., Makhloufi, R., Rached, I., Hajji, R., Landes, T., 2022. Use of Kinect Azure for BIM reconstruction: Establishment of an acquisition protocol, segmentation and 3D modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 87–94.

Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D., 2012. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *The international journal of Robotics Research*, 31(5), 647–663.

Ingman, M., Virtanen, J.-P., Vaaja, M. T., Hyyppä, H., 2020. A comparison of low-cost sensor systems in automatic cloud-based indoor 3D modeling. *Remote sensing*, 12(16), 2624.

Li, Y., Li, W., Tang, S., Darwish, W., Hu, Y., Chen, W., 2020. Automatic indoor as-built building information models generation by using low-cost RGB-D sensors. *Sensors*, 20(1), 293.

Rached, I., Hajji, R., Landes, T., Haffadi, R., 2024. Deep Learning Based Semantic Segmentation for BIM Model Generation from RGB-D Sensors. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 10.

Teo, T.-A., Yang, C.-C., 2023. Evaluating the accuracy and quality of an iPad Pro's built-in lidar for 3D indoor mapping. *Developments in the Built Environment*, 14, 100169.

Wahbeh, W., 2021. Parametric modelling approach to reconstructing architectural indoor spaces from point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 251–257.

Wang, J., Zhang, C., Zhu, W., Zhang, Z., Xiong, Z., Chou, P. A., 2012. 3D scene reconstruction by multiple structured-light based commodity depth cameras. 5429–5432.

Xu, H., Xu, J., Xu, W., 2019. Survey of 3D modeling using depth cameras. *Virtual Reality & Intelligent Hardware*, 1(5), 483–499.

Yang, Y., Jin, S., Liu, R., Kang, S. B., Yu, J., 2018. Automatic 3d indoor scene modeling from single panorama. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3926–3934.

Zhou, H., Qi, L., Huang, H., Yang, X., Wan, Z., Wen, X., 2022. CANet: Co-attention network for RGB-D semantic segmentation. *Pattern Recognition*, 124, 108468.

Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847*.

Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R., Kolb, A., 2018. State of the art on 3d reconstruction with rgb-d cameras. *Computer graphics forum*, 37number 2, Wiley Online Library, 625–652.