Use of Kinect Azure for BIM reconstruction: Establishment of an acquisition protocol, segmentation and 3D modeling

Nada El Haouss[a,*], Rawia Makhloufi[a], Ishraq Rached[a], Rafika Hajji[a], Tania Landes[b]

[a] College of Geomatic Sciences and Surveying Engineering, Institute of Agronomy and Veterinary Medicine Rabat 6202 Morocco – elhaoussnada@iav.ac.ma – makhloufirawia@iav.ac.ma – ishraq.rach@gmail.com – r.hajji@iav.ac.ma

[b] ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, National Institute of Applied Sciences (INSA Strasbourg), 24, Boulevard de la Victoire, 67084 Strasbourg Cedex, France – tania.landes@insa-strasbourg.fr

* Corresponding author

Technical Commission II

KEY WORDS: Kinect Azure, RGB-D Camera, Indoor, Acquisition Protocol, 3D Reconstruction, Segmentation, 3D Modeling, BIM

ABSTRACT:

With the popularization of RGB-D cameras, access to the third dimension is now possible at low cost. However, these systems have a lower accuracy compared to other technologies such as terrestrial laser scanners (TLS) or mobile laser scanners (MLS). RGB-D cameras have proved their potential for 3D indoor mapping, particularly for Building Information Models reconstruction (Li et al., 2020). This paper aims to investigate the acquisition protocol and propose a method for BIM reconstruction by using an RGB-D camera (Kinect Azure). First, an acquisition protocol is established with the aim of improving the quality of 3D reconstruction of indoor scenes. Different scene cases are considered, namely a single wall, a corridor, a room (with different levels of detail) and two adjacent rooms. After having extracted the best acquisition scenarios for each case of the studied scenes, an image processing is performed for the most complex scenes. Then the 3D reconstruction is performed and the resulting point clouds are subsampled and cleaned. Next, an evaluation of the geometric quality of the 3D reconstruction is performed, by making a comparison between the point clouds from the acquisition protocol (room and corridor) and the reference point clouds from an MLS. The results of this comparison shows that the differences between the two point clouds have an absolute average deviation that doesn't exceed 4.8mm, which proves that the proposed method has reached competitive accuracy. Finally, segmentation and 3D modeling of the studied scenes are proceeded to extract the BIM objects.

1. INTRODUCTION

BIM (Building Information Modeling) is one of the most notable innovations in construction engineering that allows efficiency, accuracy, and quality in project management (Cheng & al., 2020).

It is a process based on a common 3D digital model that connects the construction professionals so that they can design, construct and operate buildings and infrastructure more efficiently and collaboratively. A BIM model can either be generated based on a CAD (Computer Aided Design) model that describes the "As-Designed" state of the building or created after construction is complete; we refer to this as the "As-Built" BIM.

Point clouds from TLS (Terrestrial Laser Scanner), and MLS (Mobile Laser Scanner) are common inputs for generating BIM models. TLS has the advantage of allowing very accurate acquisition of a large volume of data, while MLS combines both accuracy and mobility (Wang & al., 2019). However, access to laser scanning has a major limitation due to its price which is not affordable to all users.

Recently, there has been a growing interest in the use of low-cost RGB-D (Red Green Blue-Depth) cameras for 3D acquisition and indoor reconstruction of buildings. Although the signal-to-noise ratio remains rather weak, the reasonable price of this sensor is very motivating for its use in 3D reconstruction of indoor scenes (Li & al., 2020), digitization of cultural heritage (Herban & al., 2022) and forestry applications (McGlad & al., 2022). Several researches have addressed the use of RGB-D cameras for 3D building reconstruction, namely Zhou & al. (2022) and Wahbeh & al. (2021). However, to our knowledge, there is no work dealing with the acquisition protocol to be respected in order to reach a good quality of 3D reconstruction.

This research aims to fill in this gap by proposing a methodological acquisition protocol that has been tested for several indoor scenes in order to extract, in a reliable way, BIM objects (walls, floors, ceilings, doors, windows) using the Kinect Azure.

The reminder of this paper is organized as follows: the methodology adopted is presented in section 2. Section 3 is devoted to the results of the acquisition protocol, the geometric reconstruction of the studied scenes and to the segmentation and 3D modeling. The paper ends with a conclusion in section 4.

2. METHODOLOGY

In this section, we detail the general methodology followed during this work from the establishment of an acquisition protocol to the extraction of BIM objects.

The acquisition protocol addresses four cases of scenes including a single wall, a corridor, a room (with different levels of detail) and two adjacent rooms. According to each scene, several experimentations were performed to find out the optimal way for accurate data acquisition. Next, an image processing was performed for the room (the least cluttered) and the corridor which were the most problematic. Two paths were carried out: one included filtering the depth images and another one that worked with raw images, in order to test the efficacy of depth images filtering. Then, we performed a 3D reconstruction followed by point cloud subsampling and denoising. Afterwards, an evaluation of the geometric quality of the 3D reconstruction was carried out. After validating the relevance of the acquisition

protocol, a workflow has been developed to generate 3D indoor models of the studied scenes and to extract BIM objects. The workflow is illustrated in figure 1 and will be commented in the next parts.

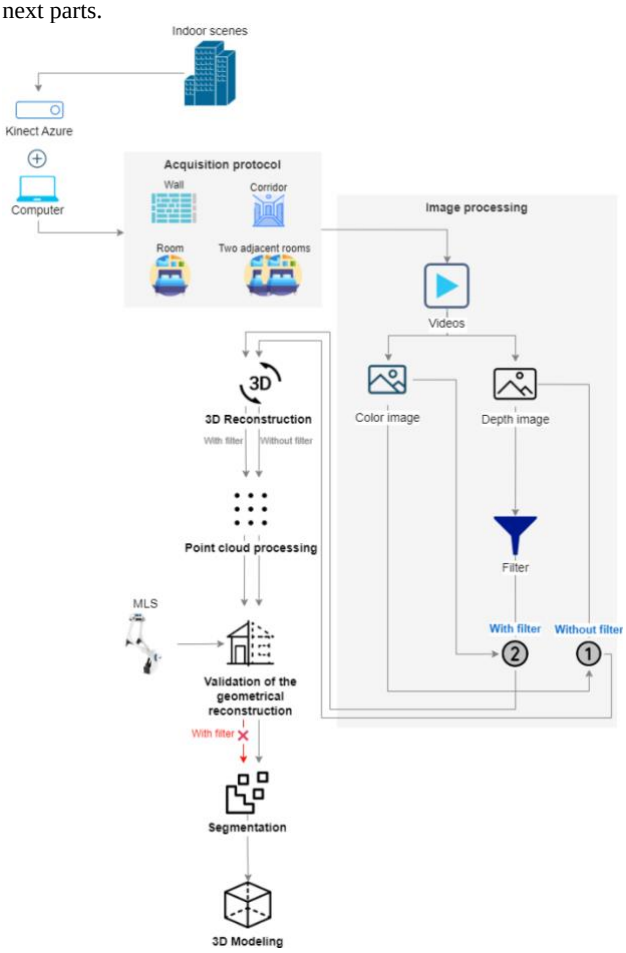


Figure 1. Flow chart of the process developed in this work.

2.1 Acquisition protocol

The process is initiated by recording a video of the scene using the RGB-D camera. RGB images and depth images are then extracted from this video sequence and aligned to generate a 3D model. This alignment step is error prone and depends on the adopted acquisition method. In this context, several experimentations regarding the trajectory, the position of the scanner etc., were performed in a way to capture the entire scene. Therefore, the number of images generated differs from one test to another, and consequently affects the alignment step. Moreover, to align two consecutive images, we used the RGBD-ICP (Iterative Closest Point) color that relies on both RGB and depth information. But even with features with different colors and depth information, the trajectory adopted by the operator affects the quality of the reconstructed 3D model. That's why we tried a variety of possibilities and compared their results to find the most adequate scenario for each particular scene (see section 3.1).

2.2 Image processing

Once the 3D models are generated, an image processing is performed for depth images, i.e. a bilateral filter applied to the depth images. Unlike the other filters, the bilateral filter is used to smooth the images and reduce noise, while preserving the edges. The model of the bilateral filter can be formulated as shown in equation (1), after Durand et al. (2002).

$$BF[I]p = (1/Wp) \sum q \in S \, G\sigma s (\|p - q\|) \, G\sigma r (\|Ip - Iq\|) \, Iq \qquad (1)$$

Where:

1/Wp : Normalization factor

Gσs (‖p - q‖): The space weight that denotes the spatial extent of the kernel,

Gσr (‖Ip - Iq‖): The range weight that denotes the minimum amplitude of an edge.

The bilateral filter includes new factors compared to other filters: the normalization factor and the range weight. The latter ensures that only pixels whose intensity values are similar to those of the center pixel are considered for blurring, while sharp intensity changes are maintained.

2.3 3D Reconstruction

The reconstruction workflow of the scenes is performed with the Open3d library including the extraction of the color images and depth aligned images, the creation of fragments, their registration, their fine-registration and finally the integration of the scene.

In order to obtain more reliable information about the local geometry of the surfaces, we divide the input RGB-D video into fragments of k=100 frames. Then, we adopt the same number of frames per fragment.

To address the odometry problem, adjacent images are initialized by an identity matrix while non-adjacent images are initialized by a sparse baseline match.

Subsequently a pose graph is constructed and optimized for multidirectional registration of all RGB-D images in this sequence. Once the pose graph is created, the multidirectional registration is performed to estimate the poses of the RGB-D images. Next, RGB-D integration is used to reconstruct a colored fragment from each RGB-D sequence.

Once the scene fragments are created, the next step is to align them in a global coordinate system.

The global registration between the different fragments is performed by the RANSAC (RANdom SAmple Consensus) algorithm. Once the pose graph is built and optimized, a multidirectional registration is performed to align all the fragments in the same global system of the scene.

Regarding the fine registration we opted for the color option which uses both geometry and color for registration. Finally, a global registration between the different fragments is performed a second time to align all the fragments in the same global system of the scene.

The final step of the workflow is to integrate all RGB-D images into a single TSDF (Truncated Signed Distance Function) volume and extract a mesh as a result.

2.4 Point cloud processing

The point cloud processing consists of two steps: subsampling followed by denoising. The first step is essential in any processing phase, it aims to reduce the number of points to simplify the segmentation step which will be done later.

The subsampling of our data was performed automatically thanks to a C++ algorithm based on PCL (Point Cloud Library). The removal of outliers from our point cloud was done manually.

2.5 Evaluation of the geometric quality of the 3D reconstruction

After applying an acquisition protocol that allows a good correspondence of the images based on a visual evaluation, a validation of the geometric reconstruction is required. We therefore consider two representative scenes of the tests performed: the uncluttered room scene and the corridor scene.
The acquisition was first performed by the Kinect with a resolution of 2048x1536 for the color camera combined with the NFOV (Narrow Field of View) Unbinned depth mode, respecting the 3D reconstruction workflow presented earlier. The scenes are then acquired by an MLS based on SLAM technology (NAVVIS VLX). The point cloud from the MLS was considered as the reference cloud since it has a better accuracy.
To make this comparison, an alignment of the two clouds is essential. This was done mainly by an ICP algorithm. Firstly, we visually evaluated the odometry effect on the two scenes to know if our acquisition protocol allows to reduce this error. Then, we computed the relative distances separating part of the two point clouds.
We also simultaneously evaluated the effect of depth image filtering on the resulting point cloud.

2.6 Segmentation and 3D modeling

Before generating the BIM model, a segmentation process of the point clouds was performed. This step consists in using the histogram of altitudes to detect walls, ceilings, and grounds, then to carry out a manual extraction of the openings (doors and windows). Before, a preparation of the segmentation data is mandatory. It consists of a 2D projection of the walls' segments to extract the coordinates (X, Y) of the edges, and then the height of the walls. The same process is applied to the opening segments (doors and windows). All these data were used as input for the automatic generation of the BIM model via algorithms developed in this work; the first one is for the 3D modeling of the scene (walls, ceilings, and floors) and the second one concerns 3D modeling of the openings within the first model.

3. RESULTS

In this section, we analyse the results of our approach, with regards to the acquisition and 3D reconstruction, the image processing, the point cloud processing, the evaluation of the geometric reconstruction, segmentation and 3D modeling.

3.1 Acquisition and 3D reconstruction

Concerning the wall scene, the multiplicity of performed experiments shows that the best way would be to capture the whole scene when it's possible, regarding the camera range (figure 2); otherwise, we can proceed fragment by fragment.

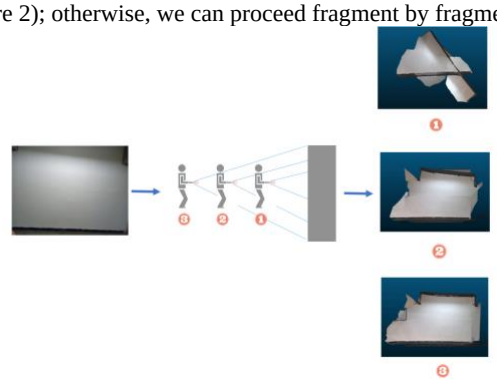


Figure 2. Comparison of results from different acquisition positions.

In this last case, we have two options: we can either adopt a vertical trajectory to acquire the first portion of the wall before moving to the next one ("up and down" method) or we can adopt a horizontal trajectory to acquire the bottom of the wall before moving to the upper level ("left and right" method). Both methods were unsuccessful (figure 3), even if the "left and right" approach seems to slightly improve the final result.

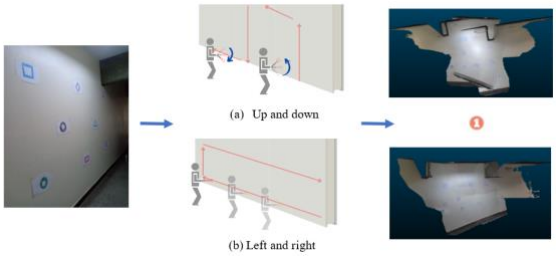


Figure 3. Acquisition result using the "up and down" (a) and "left and right" (b) approaches.

When using the "left and right" approach, there are two ways to navigate between consecutive levels. Firstly, by rotating the camera and secondly by performing a translation of the camera (figure 4). This last option was unable to differentiate between the ceiling and the floor levels. Therefore, we concluded that the best way to acquire an object is by adopting a horizontal trajectory (by level) to acquire the bottom of the wall before rotating the camera and acquiring the next level. This process should be repeated until the whole scene is captured.

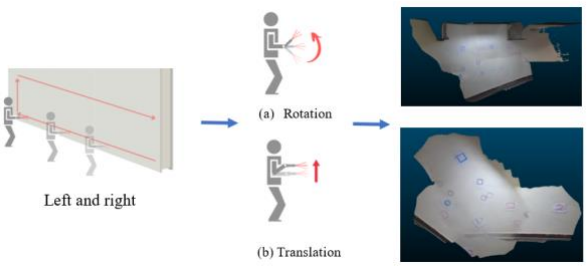


Figure 4. Result of the acquisition using a rotation (a) and a translation (b) of the camera.

To solve the image alignment problem, several experimentations were made. We tried placing on the wall a variety of 2D objects (stickers) with different RGB information. Unfortunately, that didn't help improving the quality of the 3D reconstruction (figures 5 and 6).

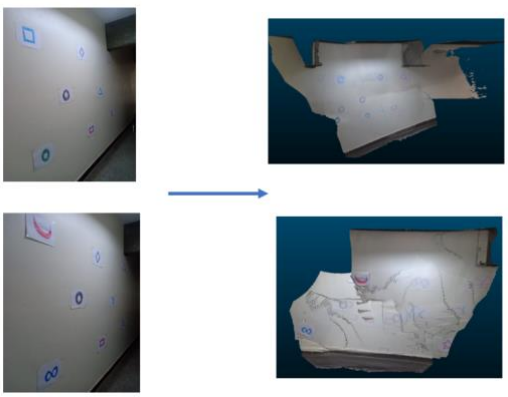


Figure 5. Effect of changing the layout of stickers.

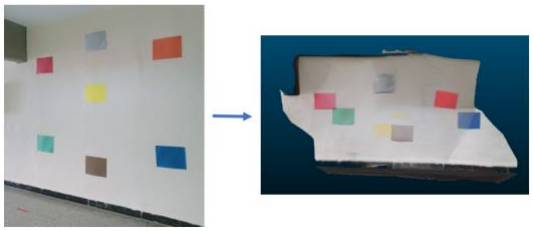


Figure 6. Effect of the colored sheets.

Since the RGB-D camera uses depth images to reconstruct the scene in 3D, we tried another method to solve the image alignment problem by placing on walls 3D objects with different shapes and colors (figure 7). This method solves perfectly the alignment problem.

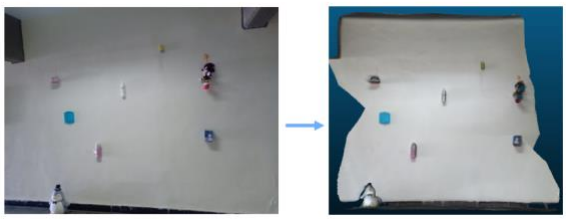


Figure 7. Impact of the use of volumetric objects on the 3D reconstruction of the wall.

For the corridor scene, we can state, through experimentation (figure 8), that it is necessary to capture it by moving according to the trajectory of the corridor and making a left and right sweep to capture all the details. Despite the presence of furniture and 3D stickers, the acquisition of the corridor by the "wall by wall" approach failed. It is therefore recommended to move along the trajectory of the corridor and acquire all the details. This method facilitates the matching of the images afterwards. Two possibilities were tested: fixing the camera on a point of view or sweeping it left and right to acquire the details of the wall. Although the second one has less occlusions since the acquisition angle has been changed to lift the details of the adjacent walls, however the generated point cloud is noisier compared to the one resulting from an acquisition with a fixed angle.
We therefore proposed to adopt the sweeping approach and perform a point cloud denoising afterwards.

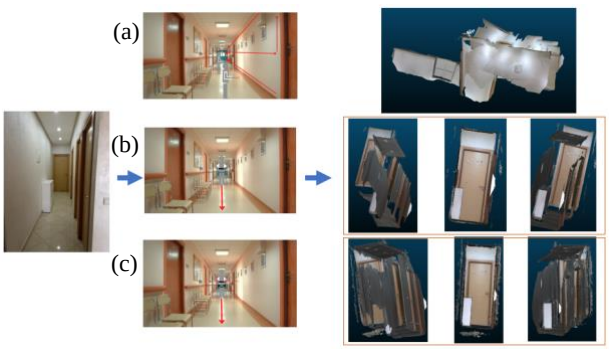


Figure 8. Results of acquisition using the “wall by wall” approach (a), moving according to the trajectory without changing the camera position (b) and moving according to the trajectory by making a left and right sweep (c).

For the room case, we explored three cases of scenes with different levels of clutter. For the cluttered room case, we firstly performed the left and right approach from the center of the room. Then we compared between the translation and rotation of the camera between consecutive levels to confirm prior results (figure 9).

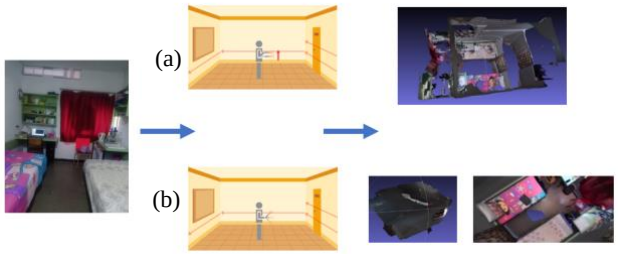


Figure 9. Result of the acquisition by translation (a) and by rotation (b) of the camera.

Even though this method gives satisfactory results, the problem of occlusions persists since the room is filled with furniture. So, we tried acquiring the room by following its details. We can distinguish between a scan of each level of the room and a scan following all levels wall by wall (figure 10).
After performing several experimentations, we can say that for a cluttered room, the best method is to scan it wall by wall.

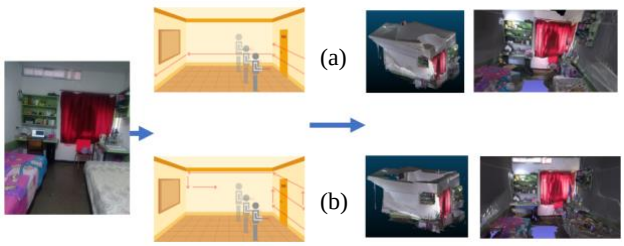


Figure 10. Result of the "level by level" (a) and "wall by wall" (b) scanning.

For the lightly cluttered (figure 11) or even uncluttered room (figure 12), a single level sweep scan with maximum distance from the target is the most optimal solution.

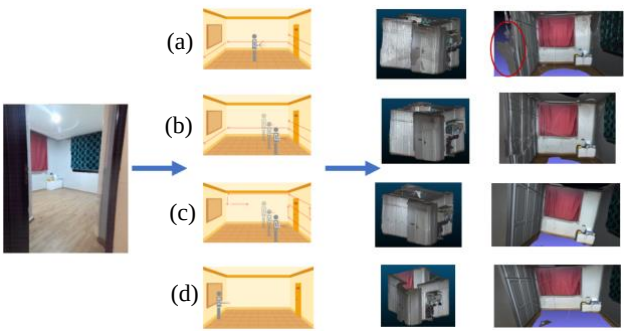


Figure 11. Results the center (a), the "level by level" (b), the "wall by wall" (c) and the single level sweep scan (d), for the lightly cluttered room.

Figure 12. Confirmation of the single level sweep method for the uncluttered room.

Lastly, we studied the case of two adjacent rooms, but we got unsuccessful results. We tried at first to capture both the interior and the exterior of a room but the alignment algorithm failed to identify any matching features between them. As a result, the model wasn't well reconstructed (figure 13).

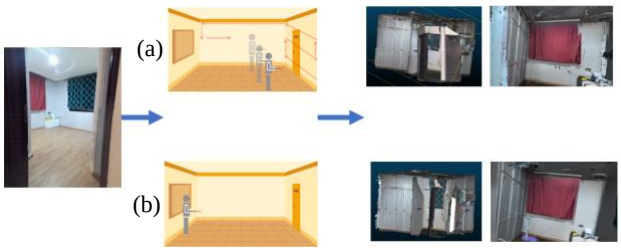


Figure 13. Results of acquisition two adjacent rooms by using single level (a) and wall by wall scanning (b).

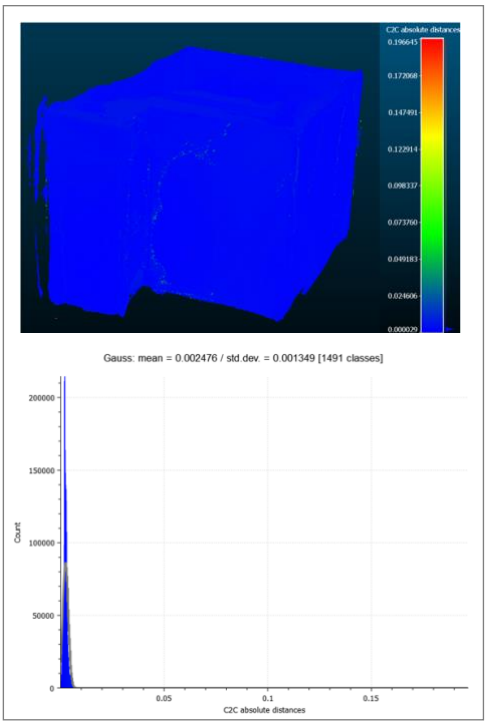
We therefore suggest making two videos; the first covers the interior of the room and the second covers its exterior, then assemble the resulting point clouds.

3.2 Image processing

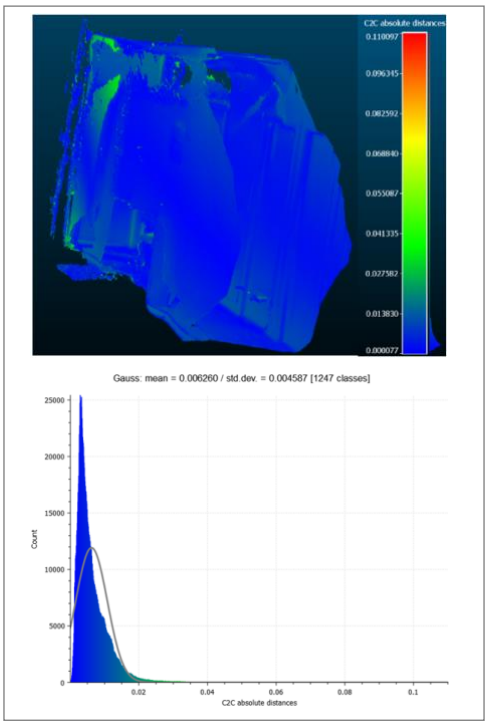
After establishing an acquisition protocol that allows a good correspondence of the images, we chose the two most problematic scenes to test the geometric quality of their reconstruction: the uncluttered room and the corridor.
Since the uncluttered room does not represent any distinctive character, it is possible that despite the loop closure, an odometry effect remains. However, before proceeding to this comparison an image processing of the results is performed.
In this step we studied the effect of the pre-processing of the depth images on the quality of the 3D reconstruction.
First, by comparing the raw and filtered point clouds, we noticed almost no significant change for the room (figure 14) and a small reduction in noise for the corridor (figure 15). However, filtering does not only act on the noise but also on the location of the points in the cloud since it modifies the depth images and consequently changes the measured values. To confirm this, a comparison of the point clouds two by two was carried out (figure 16).

Figure 14. Room case: raw point cloud on the left and point cloud from filtered depth images on the right.

Figure 15. Corridor case: raw point cloud on the left and point cloud from filtered depth images on the right.



(a)



(b)

Figure 16. Values and histograms of C2C (cloud to cloud) distances [in m] between the point cloud from a pre-processing of the depth images and the raw point cloud for the case of the scene: (a) Room, (b) Corridor.

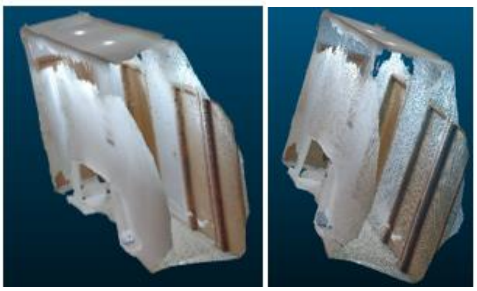
The result of the comparison of the point clouds two by two shows that the room scene has not undergone a great change unlike the corridor scene. Indeed, the average distance calculated between the two clouds is 2.5mm with a standard deviation of 1.3mm for the room case, and 6.3mm with a standard deviation

of 4.6mm for the corridor case. It is also important to note that the most pronounced deviations in the corridor point cloud are located at the corners of the walls.
We can so conclude that although the bilateral filter is intended to smooth the image and reduce noise without distorting the edges, it would change the measured values and so increase the measurement errors. In order to evaluate its impact on our result, we compared a filtered point cloud with an accurate reference point cloud. We chose the corridor scene to perform this comparison since it represents a larger number of deviations between its two compared point clouds (figure 16). Before proceeding to this step, a point cloud processing is performed.

3.3 Point cloud processing

Figures 17 and 18 show the result after applying the subsampling algorithm and removing the noisy areas manually to the room and corridor scene respectively (with filter or without filter).

Figure 17. Subsampled and cleaned point cloud of the room scenes.



(a) (b)

Figure 18. Subsampled and cleaned point cloud of the corridor: (a) without filter and (b) with filter.

3.4 Evaluation of the geometric quality of the 3D reconstruction

In this step, we focus on comparing the former results to a more precise reference point cloud from an MLS (NAVVIS VLX), with a relative precision of 6mm as supplied by the manufacturer. In order to evaluate the presence of the odometry effect, the point clouds of the two scenes from the RGB-D camera were overlayed on the clouds from the MLS taken as reference (figures 19 and 20).

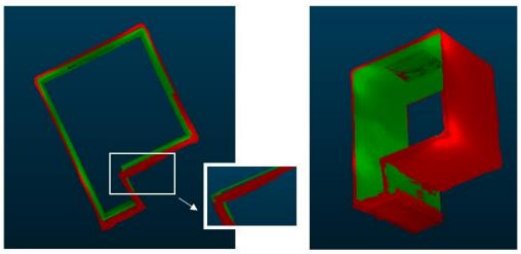


Figure 19. Superposition of the point clouds of the room scene from MLS (in green) and from the Kinect Azure (in red).

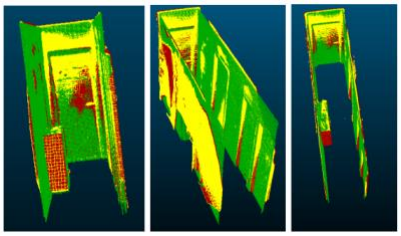
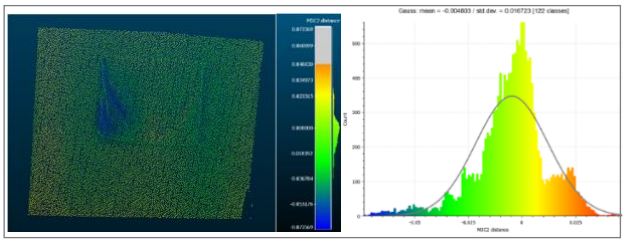
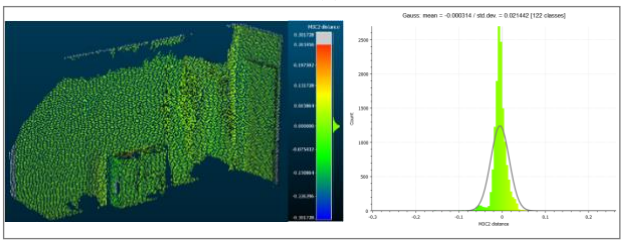


Figure 20. Superposition of the point clouds of the corridor scene from MLS (in green), from the Kinect Azure without depth image processing (red) and Kinect Azure with depth image processing (in yellow).

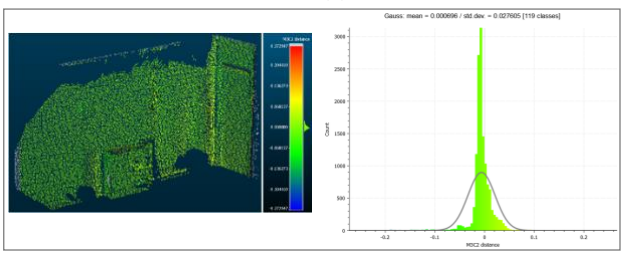
Concerning the first scene (the room), we noticed that despite the loop's closure and the good superposition of the cloud with the reference, a small effect of odometry is visible on a corner of the scene. The use of the single level acquisition method, which is more suitable for a room with empty walls, reduced this error but did not eliminate it. It would therefore be more relevant to add 3D objects in the scene to improve the geometric quality of the reconstruction.
Concerning the corridor scene, the overlay with the reference point cloud was performed satisfactory and no odometry effect was detected for both point clouds, with or without depth image processing. This is due to the presence of a distinctive set of details that allows a good 3D reconstruction.
Then, a pairwise comparison of the point clouds (figure 21) of a portion of the two scenes has been performed to quantify their differences.



(a)



(b)



(c)

Figure 21. Values and histogram of the relative distances M3C2 [in m] between the clouds from the Kinect Azure and MLS of the scene (a) Room, (b) Corridor without filter, (c) Corridor with filter. The blue represents the smallest distances and the red represents the largest distances.

We could note that the majority of the large gaps are at the corners. This may be due to the multipath phenomenon. However, it should be noted that these deviations do not greatly affect the geometric quality of the 3D reconstruction.
The histogram from the two-by-two point cloud comparison of a part of the room has an average of -4.8 mm with a standard deviation of 16mm. While that of a part of the corridor presents an average of 0.3 mm with a standard deviation of 21.4 mm in the absence of a processing of the depth images and an average of 0.7 mm with a standard deviation of 27.6 mm in the opposite case. It should be noted that these values are well within the accuracy range specified by the manufacturer (11 mm).
From these results, we concluded that the filtering of the depth images has slightly altered the measured values.
In order not to risk altering the measurements, we therefore opted for a processing of the point cloud since this approach only acts on the density of the points without affecting their location.

3.5 Segmentation and 3D modeling

The extraction of floors, ceiling, and walls is done by using the altitude histogram (figure 22), based on the value of the altitude along the Y axis.
The choice of the Y-axis is made simply because the vertical axis of the camera is pointing downwards and follows the Y-axis. The altitude of the part is therefore inversely proportional to the scalar values of Y.

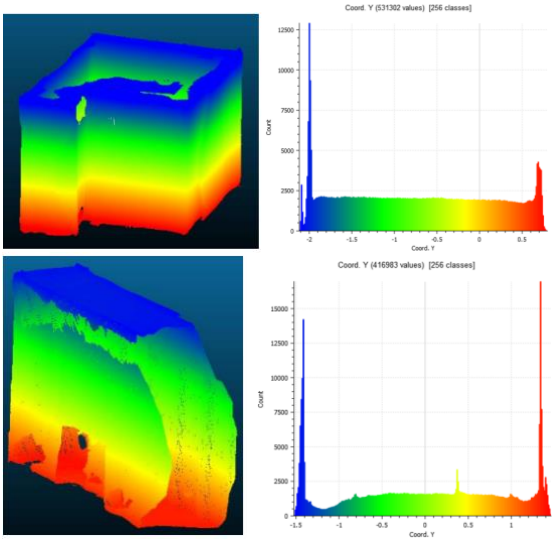
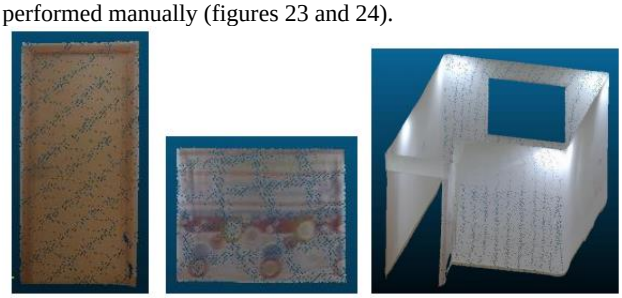


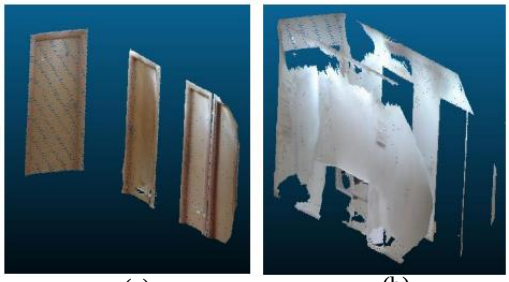
Figure 22. Point cloud exported in scalar value of Y for each case on the left and their altitude histogram on the right.

Then, the openings extraction from the wall segment was performed manually (figures 23 and 24).



(a) (b) (c)

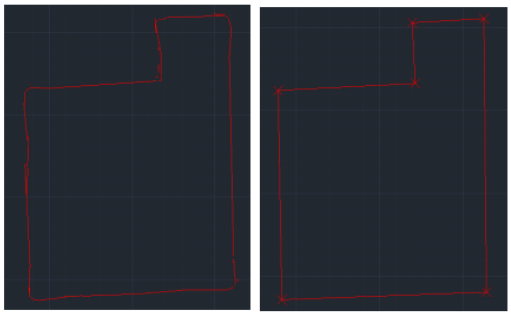
Figure 23. Room case: Segmentation of (a) door, (b) window, (c) wall without openings.



(a) (b)

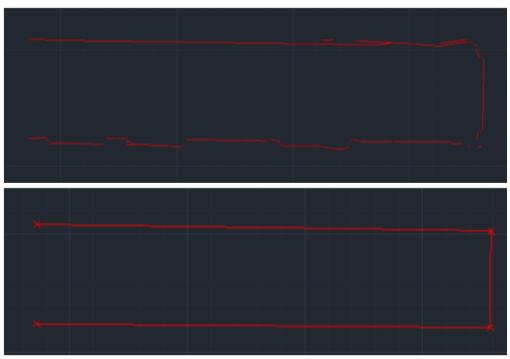
Figure 24. Corridor case: Segmentation of (a) door, (b) wall without openings.

Before 3D modeling, we first performed a 2D projection of the wall segment for both scenes. Then, it is necessary to adjust and connect the lines before exporting the coordinates of the edges (figures 26 and 27). The height of the walls and the doors is measured directly from the point cloud and then exported in CSV (Comma-separated values) format.



(a) (b)

Figure 25. Room case: 2D projection lines of the walls (a) before adjustment and connection, (b) after adjustment and connection and marking of the edges.



(a)

(b)

Figure 26. Corridor case: 2D projection lines of the walls (a) before adjustment and connection, (b) after connection, adjustment and edge marking.

Concerning the openings, we followed the same approach; we made a 2D projection of the openings for both scenes and we exported the coordinates of the edges in CSV format.
After extracting all this data, the 3D model of both scenes was generated automatically (figure 27).

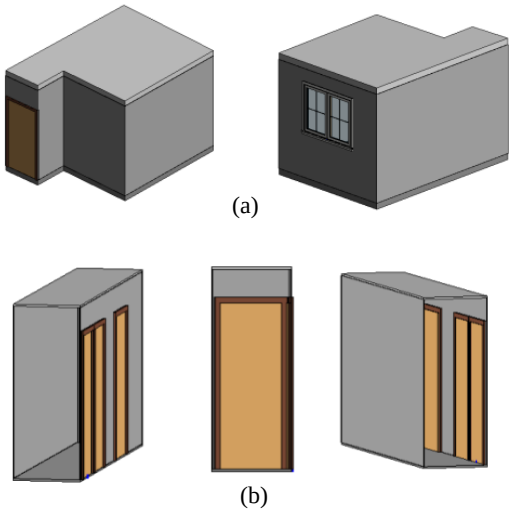


Figure 27. Results of 3D modeling for the room (a) and the corridor (b).

4. CONCLUSION

This paper sets two milestones: the first one is to design, experiment and validate an acquisition protocol by using the Kinect Azure camera for indoor 3D modeling. The second one aims to extract BIM objects through a process of segmentation and 3D modeling. The results achieved by this research show that the proposed protocol has significantly improved the 3D reconstruction of most scenes (the wall, the corridor, and the room). However, significant efforts are still required to study the case of two adjacent rooms and other cases such as a large corridor, a building façade, and a multistorey building. It is then essential to compare the resulting point clouds with a reference point cloud produced with an accurate MLS or a static TLS. Furthermore, the developed process needs to be fully automated to reach an integrated workflow that can be applied to RGB-D data for BIM reconstruction.

AKNOWLEDGEMENTS

The authors would like to thank the Geoptima company for giving access to the material used in this work.

REFERENCES

Cheng, J. C. P., Chen, K., & Chen, W. (2020). State-of-the-Art Review on Mixed Reality Applications in the AECO Industry. Journal of Construction Engineering and Management, 146(2), 03119009. doi:10.1061/(asce)co.1943-7862.0001749

Durand, F., Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. ACM Transactions on Graphics, 21(3). doi:10.1145/566654.566574

Herban, S., Costantino, D., Alfio, V. S. & Pepe, M. (2022). Use of Low-Cost Spherical Cameras for the Digitisation of Cultural Heritage Structures into 3D Point Clouds. Journal of Imaging, 8(1), 13. https://doi.org/10.3390/jimaging8010013

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. https://doi.org/10.3390/s20010293

McGlade, J., Wallace, L., Reinke, K., Jones, S. (2022). The Potential of Low-Cost 3D Imaging Technologies for Forestry Applications: Setting a Research Agenda for Low-Cost Remote Sensing Inventory Tasks. Forests, 13, 204. https://doi.org/10.3390/f13020204

Wahbeh, W. (2021). Parametric Modelling Approach to Reconstructing Architectural Indoor Spaces from Point Clouds. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 251-257.

Wang, Y., Chen, Q., Zhu, Q., Liu, L., Li, C., Zheng, D. (2019). A Survey of Mobile Laser Scanning Applications and Key Techniques over Urban Areas. Remote Sens., 11, 1540. https://doi.org/10.3390/rs11131540

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. https://doi.org/10.1016/j.patcog.2021.108468