Indoor 3D Reconstruction Using Camera, IMU and Ultrasonic Sensors

Desire Burume Mulindwa ⓘ

Tshwane University of Technology, Pretoria, South Africa
Email: deziburume@gmail.com

Abstract

The recent advances in sensing and display technologies have been transforming our living environments drastically. In this paper, a new technique is introduced to accurately reconstruct indoor environments in three-dimensions using a mobile platform. The system incorporates 4 ultrasonic sensors scanner system, an HD web camera as well as an inertial measurement unit (IMU). The whole platform is mountable on mobile facilities, such as a wheelchair. The proposed mapping approach took advantage of the precision of the 3D point clouds produced by the ultrasonic sensors system despite their scarcity to help build a more definite 3D scene. Using a robust iterative algorithm, it combined the structure from motion generated 3D point clouds with the ultrasonic sensors and IMU generated 3D point clouds to derive a much more precise point cloud using the depth measurements from the ultrasonic sensors. Because of their ability to recognize features of objects in the targeted scene, the ultrasonic generated point clouds performed feature extraction on the consecutive point cloud to ensure a perfect alignment. The range measured by ultrasonic sensors contributed to the depth correction of the generated 3D images (the 3D scenes). Experiments revealed that the system generated not only dense but precise 3D maps of the environments. The results showed that the designed 3D modeling platform is able to help in assistive living environment for self-navigation, obstacle alert, and other driving assisting tasks.

Keywords

3D Point Cloud, Position Estimation, Iterative Closest Point (ICP),
Ultrasonic Sensors Distance Measurement, 3D Indoor Reconstruction

1. Introduction

The fusion of multiples ultrasonic sensors and a camera to produce a better 3D

reconstruction of an indoor environment is a challenging task even though it helps to extend the possibilities in robotics. The main idea is to build up a platform combining the advantages of low-cost compensatory sensors towards acceptable indoor environment 3D reconstruction, with a satisfactory quality for mobile platforms (such as a wheelchair) navigation and driving assistance.

This paper implements the data fusion at two levels. Firstly, to fuse the data from the ultrasonic sensor array to compute the distance to any objects (or obstacle) in the scene with accuracy after removing measurement noise and drift, and secondly, to fuse the ultrasonic depth measurements and the 3D point clouds generated from multiple 2D calibrated images from the mounted web camera using existing methods.

There are few methods that depend on a regular camera output to reconstruct 3D scenes [1]. These techniques are known as structure from motion (SfM). Simultaneous localization and mapping (SLAM) is one such application in the domain of incremental structure from motion.

This paper is arranged as follows. In Section 2, the related work is discussed to assess the most common techniques for 3D point clouds generation. Section 3 explores the background to this research by providing the motivations for the proposed algorithm while Section 4 explains in detail the different steps taken to reach the results by exposing the proposed methodology. This is followed by the experimental results in Section 5 as well as the conclusions in Section 6.

2. Related Work

Several authors have attempted to reconstruct the 3D indoor environments using non-invasive sensors such as camera, lidar… Structure-from-motion (SFM) is a well-known approach to recover the 3D shape of an object in the presence of numerous sequences [2]. Tomasi and Kanade [3] introduced a factorization method to recuperate the contour of rigid objects using an orthographic camera.

With these sensor fusion applications, a mandatory step for extrinsic calibration is the relative transformation between the involved sensors. Qilong Zhang and Robert Pless [4] explain both the experimental and theoretical results for the extrinsic calibration of a 2D laser range finder and a camera.

The Iterative Closest Point (ICP) algorithm normally uses the features of the equivalent points in the registration process to locate the closest neighbor matching points [5]. As an algorithm for point clouds registration, the Iterative Closest Point needs some little improvement since it presents numerous limitations such as the inability to define acceptable correspondence between two point clouds especially for sparse point clouds [6]. Accordingly, Gressin *et al.* [7] define some low-level features to tackle the correspondence by adopting a hybrid approach that combines surface reconstruction and feature extraction for 3D point clouds registration.

3. Background to the Study

Independent mobility of the handicapped particularly the visually impaired in

their daily indoor environments could be improved with a wheelchair equipped with an obstacle sensing and avoidance system.

The main problem considered here is the use of the depth dimension of the 3D point clouds generated by the ultrasonic sensors and IMU system to acquire a better 3D point cloud. At the end of the day, a precise 3D real-time reconstruction of the indoor environments for mobile navigation and driving assistance is obtained.

Figure 1 depicts the setup of the entire system setup presented with the grid of the four ultrasonic sensors, where only the active ultrasonic sensor can transmit and receive ultrasonic wave, all the other 3 receiver ultrasonic sensors(with the transmitter blocked) only receive ultrasonic wave.

4. Proposed Methodology

This article is aimed at using a modified but robust iterative closest point algorithm that combines the structure from motion generated 3D point clouds with the ultrasonic sensors and IMU generated 3D point clouds to derive a much more precise point clouds using the depth from the ultrasonic sensors.
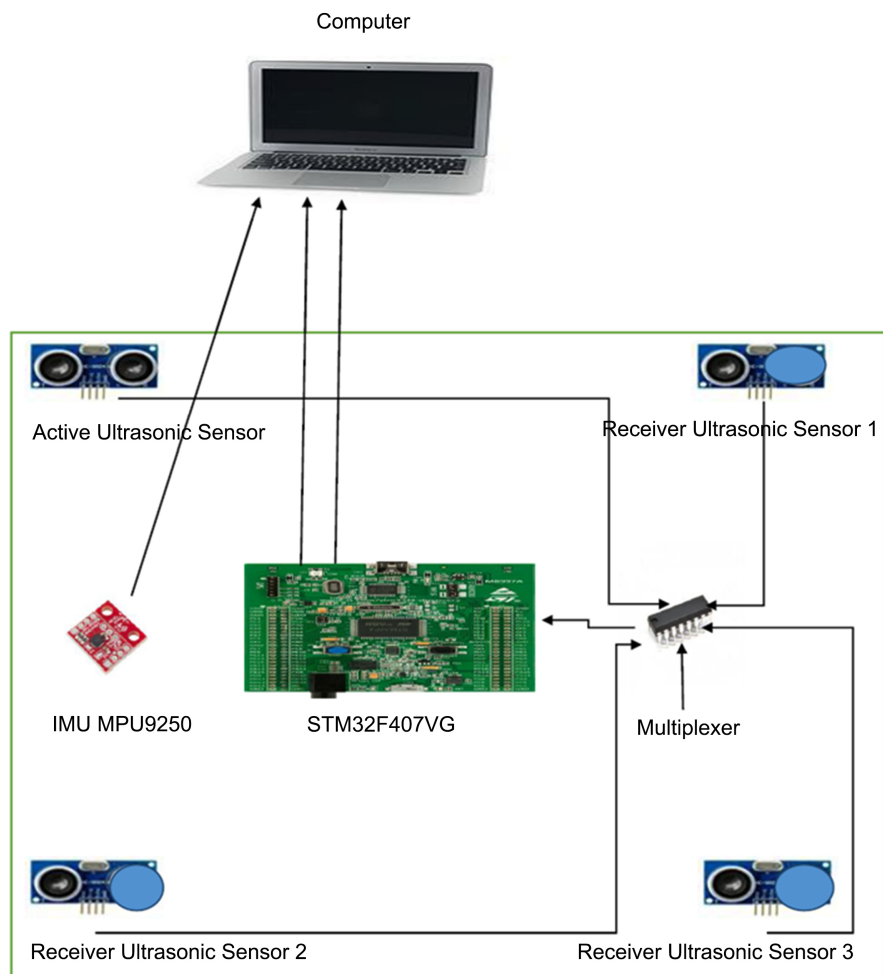

Figure 1. System structure of the 3D sensing platform.

4.1. 3D Point Clouds Generated from the Fusion of the IMU and the Ultrasonic Sensors

4.1.1. Reliable Distance Measurements

The detection of the 3D reflection point implies two tasks such as the range measurements from the facing obstacle as well as the wheelchair displacements detection in the indoor environment. As the wheelchairs carrying the system (ultrasonic sensors grid, STM32F407VG, MPU9250, camera…) is moving along a wall, an accurate 3D point cloud is expected to reflect this reality.

1) *Distance Estimation Procedure*

One ultrasonic sensor only provides a single measurement output for each transmitted pulse [8]. Hence, if facing multiple obstacles in the detection range, an ultrasonic sensor detects one object only, unlike other range measurement sensors such as the lidar. In reality, unexpected data jumps in the ultrasonic sensor measurements cause an unreliability concern [9].

To alleviate the vulnerabilities of relying on single ultrasonic measurements, the research used an array of multiple ultrasonic sensors (**Algorithm 1**).

Algorithm 1. Ultrasonic sensors array distance estimation algorithm.

Distance measurement algorithm of the four ultrasonic sensors array
% Data availability in the communication channel
If data available **continue; end**;
If not, wait for more data.
% Data validity is expressed through the condition that the reading should be between 10 cm and 1000 cm
If data < 10 || data > 1000
then data valid, continue; **end**;

% Addition of channel selected as they are changing since the system is using 4 channels (4 ultrasonic sensors) as organized by a multiplexer
counter (channel_ID) = counter (channel_ID) + 1;
% Local reflection point position where only the Z component (depth) is read while X and Y (platform displacements) are set to zeros here
    x = zeros(size(d));
    y = zeros(size(d));
    z = (0.5*d-p./(2*d));
where d = ultrasonic sensor distance
      p = coordinates position of the ultrasonic sensor in the array with respect to the active ultrasonic sensor

---

2) *3D Point Cloud Generation Procedure*

Figure 2 explains the steps taken to obtain the 3D point clouds from the fusion of the ultrasonic sensors and the IMU (accelerometer). The data from the Z axis are discarded since the mobile platform (wheelchair) is not expected to move neither upward nor downward in its path. In order to achieve the generation of the reflection point in 3D coordinates, the only missing component is the Z axis. The data from the 4 ultrasonic sensors provide the range (distance) from any obstacle. This range is used as the Z component data for the reflection point.
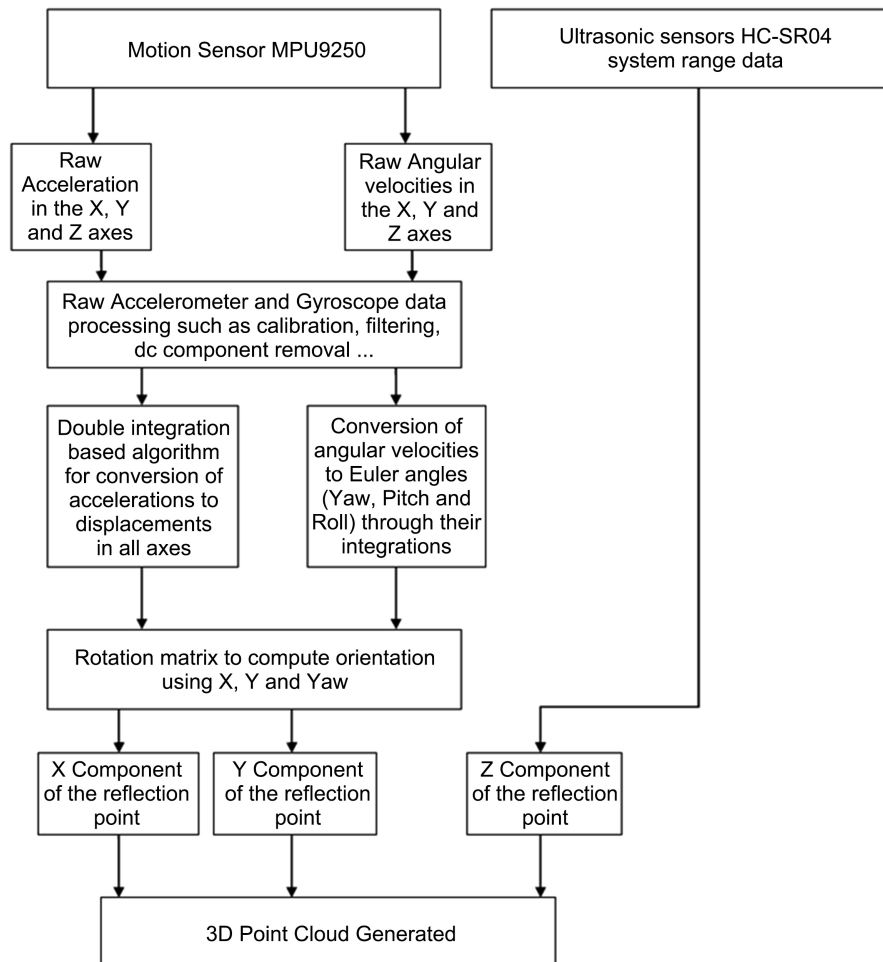
Figure 2. Schematic of the IMU and ultrasonic sensors 3D Point Cloud generation process.

4.1.2. Implementation of the 3D Scene Modeling Algorithm

1) *Data Collection and Processing*

The design of a Simulink diagram (**Figure 3**) was a critical task in order to manage both the connections and the data communication in the system. For UART communication between the proposed system and the computer, a Simulink diagram had to be designed in order to dictate the format, the length and the nature of the data passing through the system.

2) *Camera Calibration*

The detected reflecting points are projected to the global coordinates system to have a cumulative point cloud for the environment. The locally reflected point will then be the mean value of all these Z values found according to their specific ultrasonic sensor on the mobile platform. In order to convert local to global coordinates, the displacement of the mobile platform has to be acquired accurately in all three dimensions namely the X, Y and Z planes. In order to gather the displacements, a double integration of the processed acceleration values from the Sparkfun MPU9250 is performed before the inherent drift is removed.

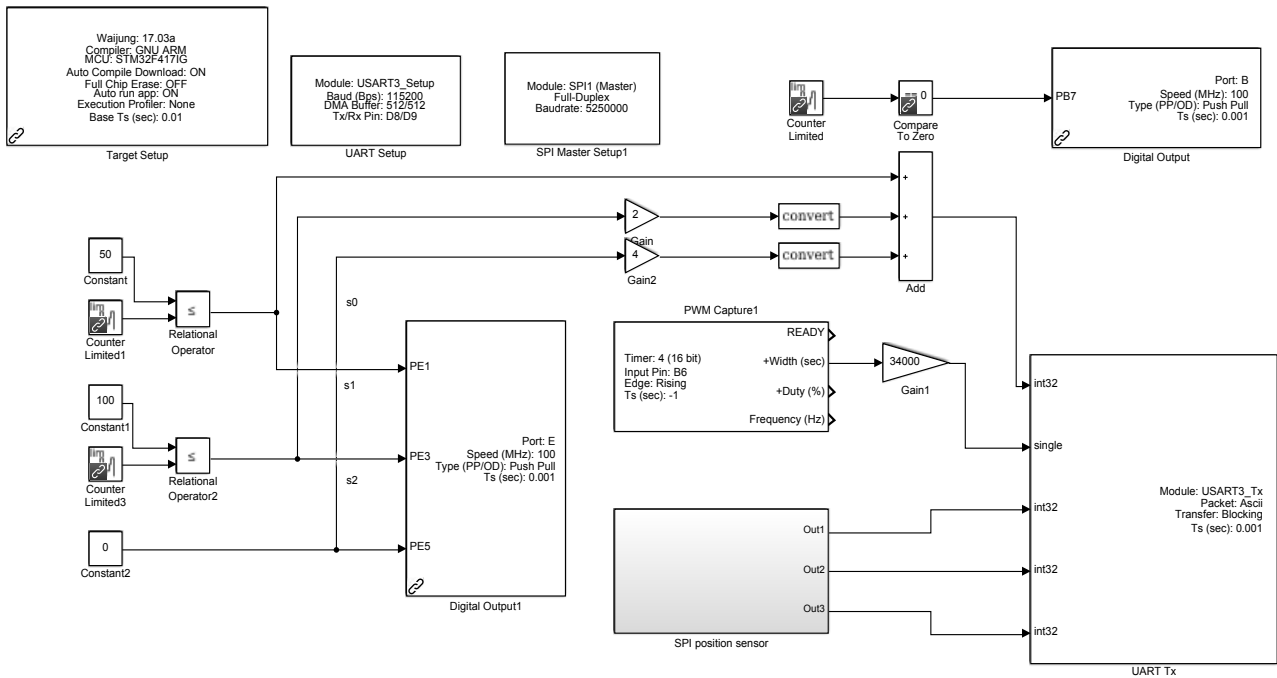**Figure 4** illustrates the relationship between local and global coordinate systems:

Figure 3. Simulink Blockset for ultrasonic sensors with multiplexer using Serial Peripheral Interface (SPI).
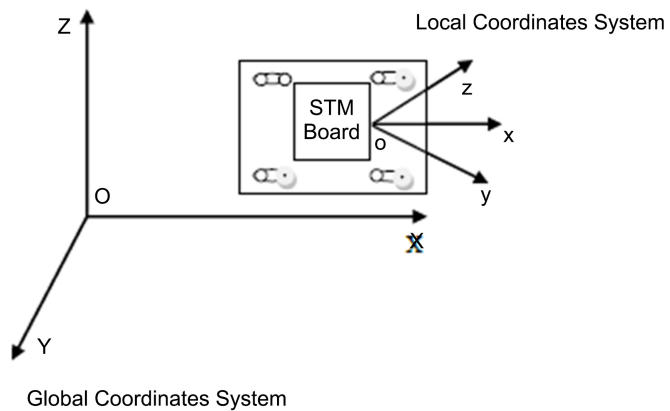

Figure 4. Global and local coordinates of the system.

The transformation matrix entails the translation of the 2D camera coordinates frame to the ultrasonic sensors 3D point clouds coordinate frame by using extrinsic transformation as shown in the figure below (**Figure 5**).

The camera is first calibrated using a checkerboard pattern (**Figure 6**). At that moment, the position of the plane in the camera coordinates system is computed. The ultrasonic sensors system also scans the same plane. As the camera catches the plane images, the presence of some of the ultrasonic sensors scanned points on the pattern plane allows estimating the pose between the sensors.

3) *Position Estimation*

Since X and Y constitute the first two components of the 3D point clouds (X, Y, Z), it is very important to keep track of the wheelchair movements at all times. These displacements are acquired by integrating the acceleration data twice.
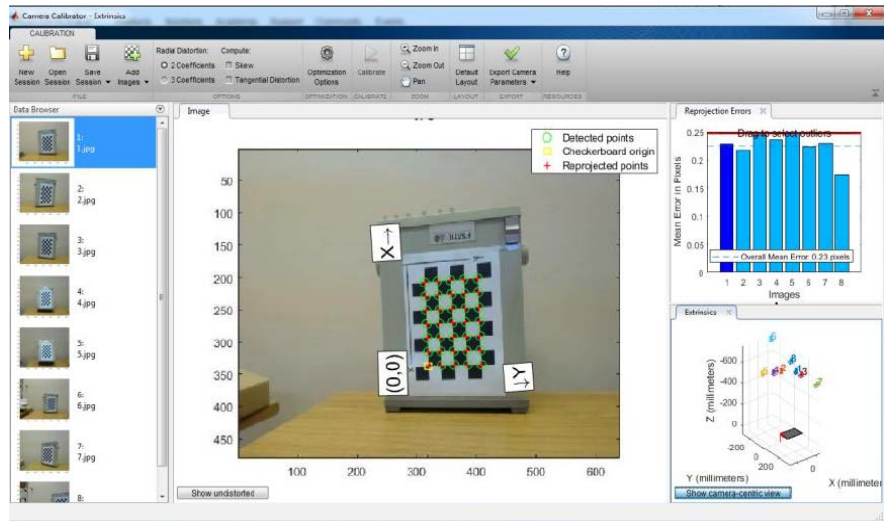
Figure 5. Extrinsic calibration of camera and ultrasonic sensors.
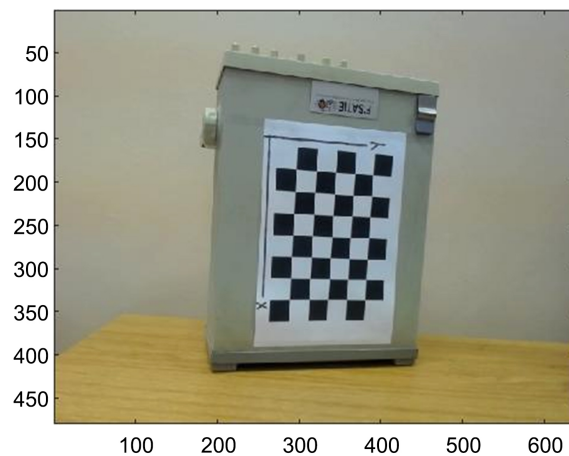

Figure 6. Checkerboard patterns of the fiducial target.

In reality, as the experiment begins, all three parameters (namely the acceleration, the velocity and the position) are zero (null). As the wheelchair starts to move, the acceleration is expected to change quickly as much as the velocity while the position also shifts numerically. The problem is that a drift appears and affects the accuracy of the position.

In order to tackle the drift problem, an algorithm computes the drift rate which will derive the range of the drift in the velocity data. At the end of the day, this velocity drift will be subtracted from the velocity.

The value of the rotating angle results from the integration of the angular velocity obtained from the gyroscope readings over time to obtain the Euler angles as shown in **Figure 7**.

Experiments have shown that the gyroscope data also have a drift that is constantly accumulating errors resulting in very erroneous data. The solution could be fusion with the accelerometer to correct the drift through a complimentary filter.

4.2. Modified Iterative Closest Point Algorithm

The process of enhancing the quality of the resulting point cloud is depicted in Figure 8. First, both the generated point clouds (structure from motion and ultrasonic sensors) are projected on a two-dimensional grid to determine neighborhood between points in 2D. This process is followed by a surface analysis that performs a feature extraction of the scene targeted and an alignment of the two-point clouds. Finally, the point clouds are registered and fused in order to improve the precision of the 3D model.

4.2.1. 2D Projection

The two consecutive 3D points (from Sfm and from ultrasonic sensors array) are projected onto a two-dimensional grid at the same time. The idea is to be able to
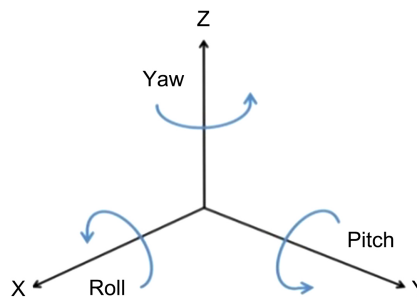

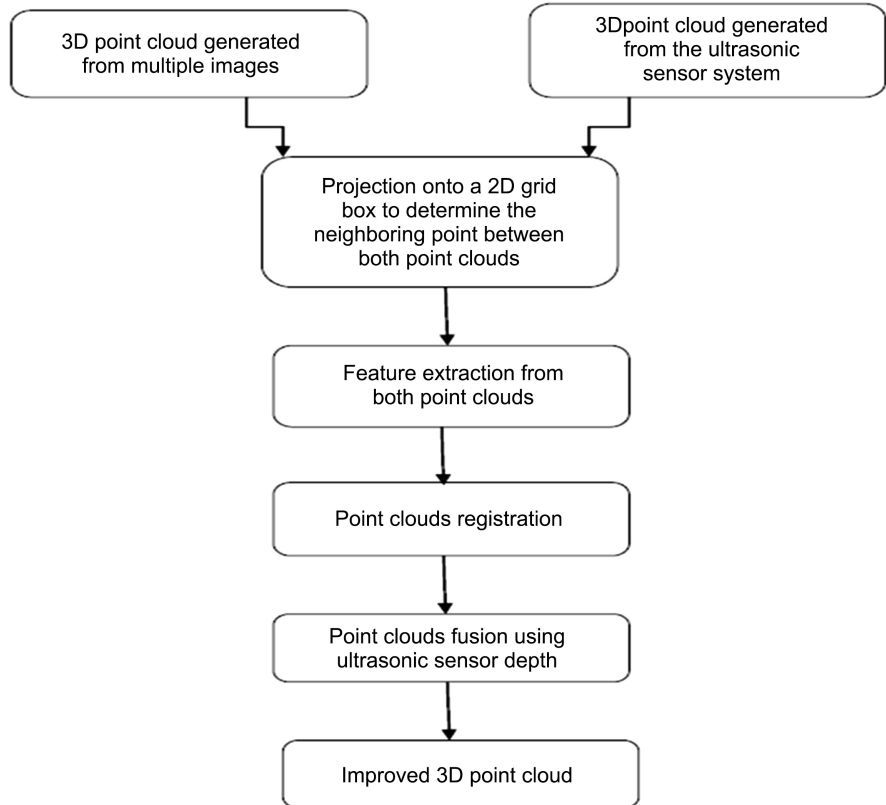**Figure 7.** Euler angles: roll, pitch and yaw.

Figure 8. Overview of the system.

view which point are going to be found in the same grid to determine neighborhood. This method uses the closeness of two pixels in the 2D projection to conclude that two points are neighbors in 3D.

4.2.2. Feature Extraction

Targeted scene 3D reconstruction depends largely on the geometrical features as a result of the choice of the neighborhood selected. Because of the accuracy of the ultrasonic sensor, the points selected in the same neighborhood will define the geometrical features of the scene. These features are the result of the depth measurements gathered by reading the distances from the system.

4.2.3. Local Registration

Most solutions to detect the alignment of two 3D point clouds lie on the iterative closest point algorithm. A difficulty lies in the fact that the ICP approach presents stern limitations. For instance, it is difficult to detect acceptable equivalence between two consecutive point clouds, as often, the corresponding pairs found do not illustrate the same point in reality leading to inaccurately aligned 3D point clouds and to wrongly projected sensor poses.

The advantage of using the 3D point clouds produced by the ultrasonic sensors is that they are accurate even though sparse. This accuracy in distance measurement will be used to improve the quality of the resulting 3Dpoint cloud. The improved ICP algorithm steps are explained in Algorithm 2 below.

Algorithm 2. Modified Iterative Closest Point algorithm.

% Improved Iterative Closest Point algorithm
% Point Clouds from structure from motion and ultrasonic sensors array
availability in the system
If 2 point clouds data are available,
If area is reliable, **continue; end;**
else, wait for more data.
% Area is unreliable when it is at the corner of a number of objects or when it reflects a complex object.
% Correspondence estimation
If both point clouds are projected in 2D space
Then corresponding point are the closest points in the 2D grid
% The points with the smallest distance between them in the 2D space are chosen.
% Distance from the ultrasonic reading to determine the range to the obstacle
For all points in the point clouds with similar neighborhood,
use the distance from the ultrasonic reading to determine the range to the obstacle
since ultrasonic sensors generated point clouds are sparse but more reliable;
% Transformation estimation
for all points:
1. apply transformation on both point clouds which consist of reducing
drastically the point-to-point distance
2. match points of the corresponding point cloud while changing the Z component
of the structure from motion point cloud with the distance from the obstacle
gathered by the ultrasonic sensors.
3. transform the reference point cloud using the estimation of the present location.
% Fusing with the second registered point cloud.

end;

---

5. Experimental Results

5.1. IMU and Ultrasonic Sensors 3D Point Clouds Generation

To achieve accurate 3D scans of the apparent scenes, this paper used a platform mounted with four ultrasonic sensors, camera, STM32F407VG, motion sensor MPU9250 (IMU). The system was mounted over a mobile platform such as a wheelchair in order to assist the living conditions for the handicapped or impaired vision individuals.

5.1.1. System Practical Setup

The practical setup as illustrated with the wiring diagram (**Figure 9**) is as follow:

The main role of the setup was twofold:

- to determine the 3D coordinates of the reflected points as the platform is moving around the room.
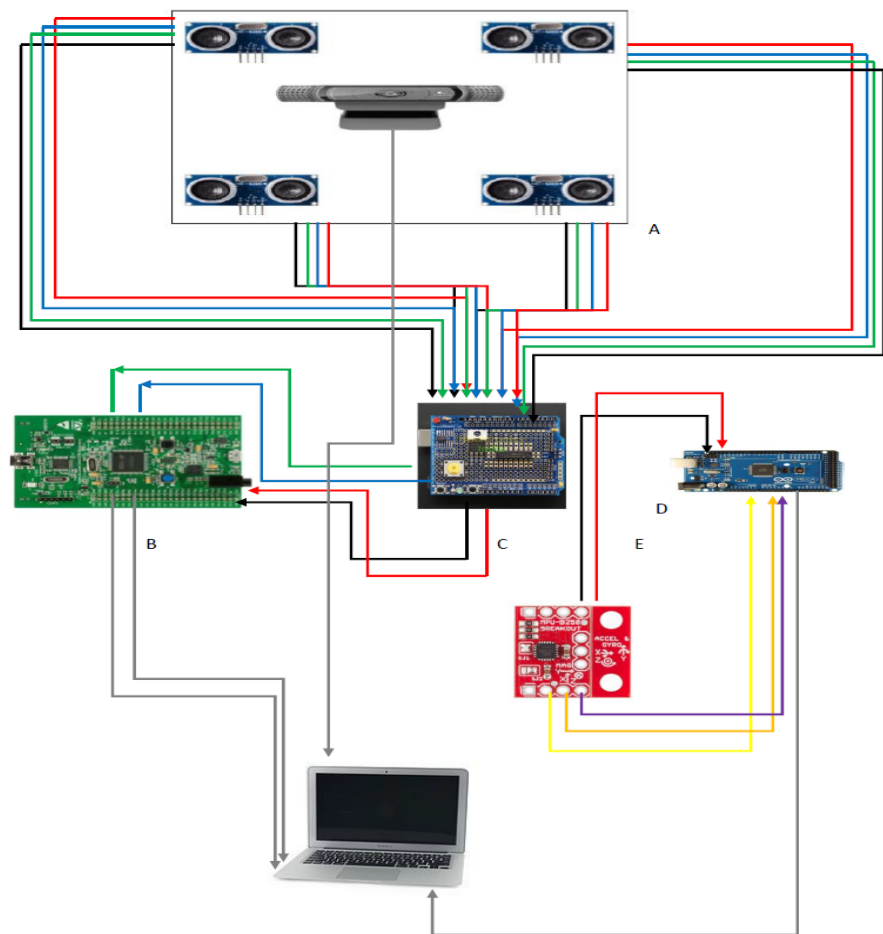- to allow the 2D camera to simultaneously take pictures of the viewed scenes.

Figure 9. Wiring diagram of the whole system. The figure illustrates the wiring of the whole system with the grid with four HC-SR04 ultrasonic sensors with the Logitech C920 Pro Camera (A), the STM32F407VG discovery board (B), the Proto Shield for development with the Multiplexer 74hc4051 (C), the Arduino Mega 250 for connection between the IMU sensors and computer (D), as well as the MPU9250 Breakout board IMU Motion sensors (E).

Figure 10 below shows a depiction of the actual system with all the actual components used for its implementation.

5.1.2. 3D Reflection Points

In this research, the process to generate 3D point clouds involves several steps among which are accurate ultrasonic measurements, drift-free mobile platform position estimations and a modified iterative closest point algorithm for a more accurate depth component of the 3D point cloud.

Through the fusion of the ultrasonic sensors measurements (in the Z axis) and the position estimations (X and Y axes), it was possible to determine the reflection point of the entire system to any given surface. This is the obstacle sensing ability that can be recorded with the help of the 3D point clouds shown in Figure 11 which depicts a section of the laboratory where the research was undertaken, which displays a section of the indoor environment where some obstacles such as furniture, electronic materials can be sensed through the 3D point cloud generation, as shown in Figure 12.

After the camera calibration, a sequence of 2D images was input before being converted to grayscale. While undistorted images are acquired, the corner features are extracted where matching points are identified. Just as the extraction of the feature points and the identification of the match points, the object location is acquired using both the scene's matching points and the depth data analysis from the ultrasonic sensors and inertial navigation system.

5.2. 3D Dense Reconstruction

5.2.1. Extrinsic Calibration

The MATLAB camera calibration application is used to calibrate this webcam. For this session, the calibration target was a checkerboard pattern with a checkerboard square of 25 mm. First, the application aimed to detect the checkerboard pattern in each of the subsequent added pictures, and secondly to display
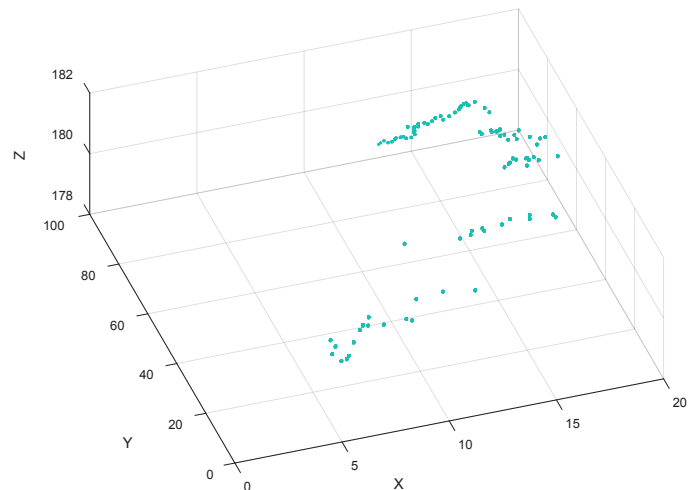

Figure 10. Actual setup illustration.

Figure 11. 3D point clouds reflect the shape of obstacles in the indoor environment.

Figure 12. The spatial environment to be modelled.

the accepted images with the green circles showing the detected checkerboard corners. **Figure 13** below is the webcam's centric view of the accepted checkerboard patterns.

Upon completing this session, the projection matrix of the Logitech webcam C920 Pro specific to this study was computed.

5.2.2. 3D Reconstruction

The dense reconstruction was two folds since first the Sfm reconstruction first took place followed by the correction from the ultrasonic and IMU generated point cloud using its Z component. In the ICP algorithm, a step did implement the replacement of the Z component.

For the Sfm, the image pictures displayed below in **Figure 14**, which has 2 views, shows a view of an indoor environment. These two adjacent pictures were turned into 3D views following the methodology presented in this work.
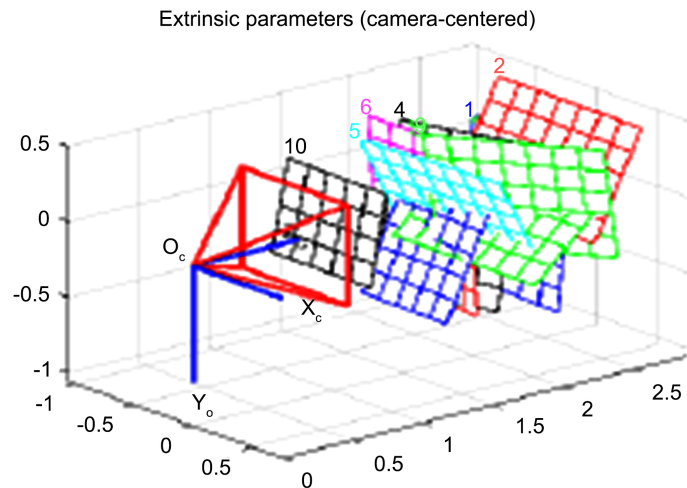
Figure 13. The Logitech webcam C920 Pro centric view of the accepted checkerboard patterns.
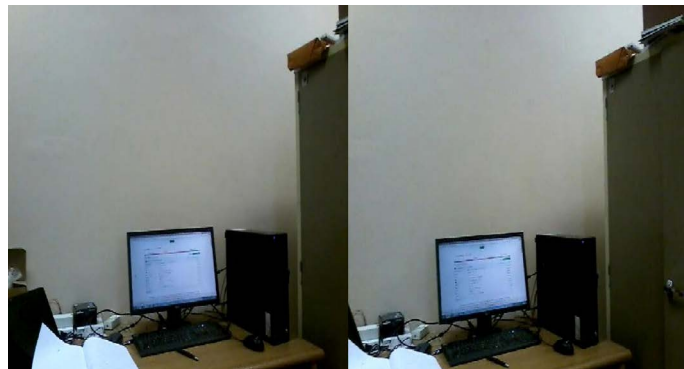

Figure 14. System input samples.

The features of two images were detected and associated, as shown in **Figure 14**, **Figure 15**, and **Figure 16**.

The corner points in the scene were considered as the features (**Figure 15** and **Figure 16**) for further processing.

After that, the images with features extracted were associated by matching the feature points as shown in **Figure 17**.

Just as the extraction of the feature points and the identification of the match points, the object location was acquired using both the scene's matching points and the depth data analysis from the ultrasonic sensors and inertial navigation system.

After using the Z co-ordinate from the ultrasonic and IMU generated 3D point cloud (**Figure 18**) to correct for depth estimation, the 3D model was reconstructed in the form of dense reconstruction (**Figure 19**).

6. Conclusions

This article proposed a method for the 3D reconstruction of indoor environments using sensor fusion. This technique helps in assistive living environment
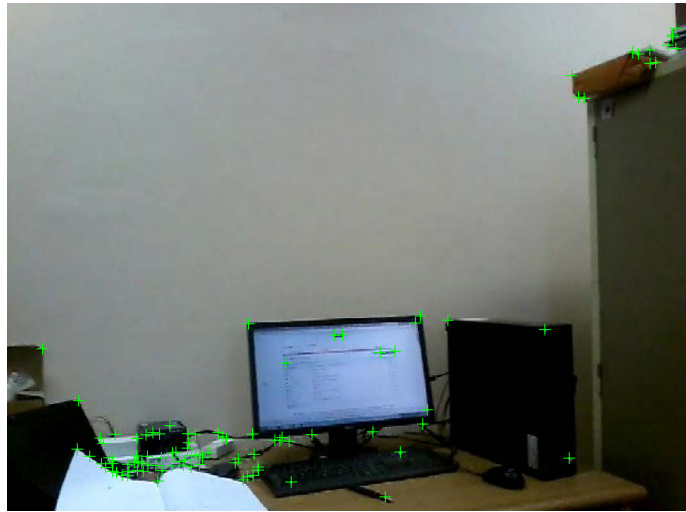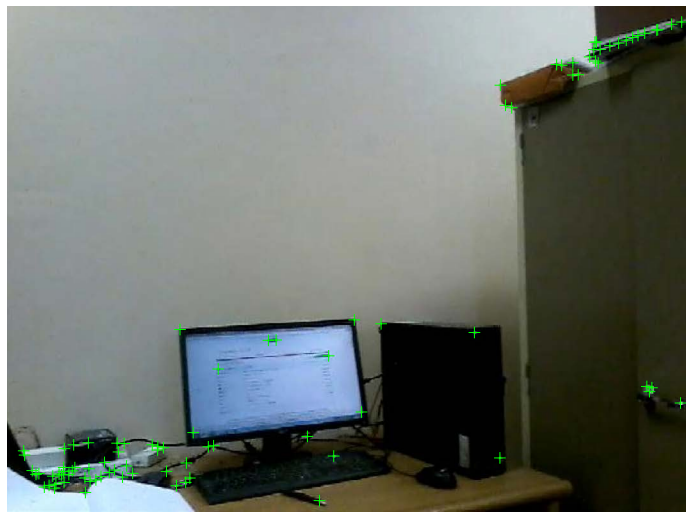
Figure 15. Corner points detected in sample 1.
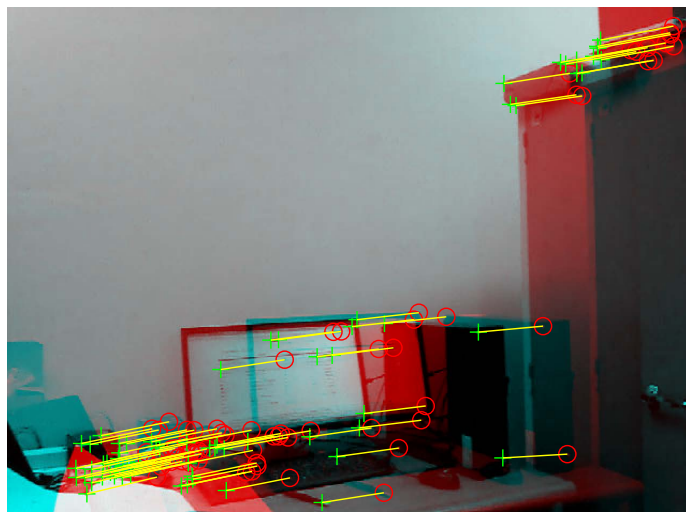

Figure 16. Corner points detected in sample 2.

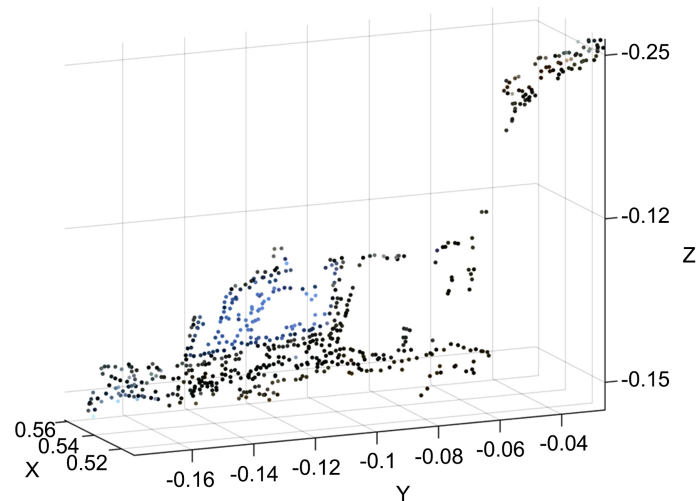**Figure 17.** Associated feature points.

Figure 18. The structure from motion generated 3D point cloud.

Figure 19. A view of the 3D model after reconstruction.

and aims to enhance the lives of both the handicapped individuals and/or those with impaired vision in their daily activities inside their indoor environments.

The existing algorithms generate the 3D images (point clouds) from 2D images, which offered high spatial resolution. Then the ultrasonic rangefinders were utilized to capture the spatial measurement, which was then used to correct the 3D images to obtain a higher spatial accuracy. Although not much novel research done on this point, inertial navigation system constitutes the basis of this research since without accurate positioning of the mobile platform (wheelchair), it is almost impossible to determine the various obstacles distance as well as the distance traveled. The experiment results show the platform achieved satisfactory performance for the requirements of mobile facility navigation and driving assistance.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Caruso, D., Engel, J. and Cremers, D. (2015) Large-Scale Direct SLAM for Omni-Directional Cameras. *Proceedings of the* 2015 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Hamburg, Germany, 28 September 2015-2 October 2015, 141-148. https://doi.org/10.1109/IROS.2015.7353366

[2] Longuet-Higgins, H.C. (1981) A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, **293**, 133-135.

[3] Tomasi, C. and Kanade, T. (1992) Shape and Motion from Image Streams under Orthography: A Factorization Method. *International Journal of Computer Vision*, **9**, 137-154. https://doi.org/10.1007/BF00129684

[4] Zhang, Q.L. and Pless, R. (2004) Extrinsic Calibration of a Camera and Laser Range Finder (Improves Camera Calibration). Department of Computer Science and Engineering, Washington University, St. Louis.

[5] Shi, X.Y., Peng, J.J., Li, J.P., Yan, P.T. and Gong, H.Y. (2019) The Iterative Closest Point Registration Algorithm Based on the Normal Distribution Transformation. *Procedia Computer Science*, **147**, 181-190.

[6] Han, J., Yin, P., He, Y. and Gu, F. (2016) Enhanced ICP for the Registration of Large-Scale 3D Environment Models: An Experimental Study. *Sensors*, **16**, 228. https://doi.org/10.3390/s16020228

[7] Gressin, A., Mallet, C., Demantké, J. and David, N. (2013) Towards 3D Lidar Point Cloud Registration Improvement Using Optimal Neighborhood Knowledge. *ISPRS Journal of Photogrammetry and Remote Sensing*, **79**, 240-251. https://doi.org/10.1016/j.isprsjprs.2013.02.019

[8] Abdullah, R.H. (2015) Design and Implementation of Ultrasonic Based Distance Measurement Embedded System with Temperature Compensation. *International Journal of Emerging Science and Engineering* (*IJESE*), **3**, No. 8.

[9] Rhee, J.H. and Seo, J. (2019) Low-Cost Curb Detection and Localization System Using Multiple Ultrasonic Sensors. *Sensors*, **19**, 1389. https://doi.org/10.3390/s19061389