

# 3-D Reconstruction and Measurement System Based on Multimobile Robot Machine Vision

Moemen Y. Moemen<sup>ID</sup>, Haidy Elghamrawy<sup>ID</sup>, Sidney N. Givigi<sup>ID</sup>, *Senior Member, IEEE*,  
and Aboelmagd Noureldin<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—Visualization and measurement for civil structure have been a constant field of research. Typically, 3-D measurement systems are static instruments that offer low degrees of flexible motion and require a significant amount of time for successful 3-D measurement. In this article, a machine vision system is developed to facilitate the 3-D measurements and reconstruction process. The proposed method depends on multimobile robots to achieve a portable 3-D measurement and reconstruction system. It consists of a 3-D reconstruction and measurement pipeline, which deploys multiple vision sensors mounted on teleoperated mobile to acquire the 3-D scan and 2-D image data. It can operate in both indoor and outdoor environments as for each structure a sequence of images and LiDAR scans are processed. The outcomes of the proposed method are a 3-D visualization for the structure along with its dimensional and volumetric measurements. The performance of the proposed system is evaluated using indoor and outdoor experiments to validate its application for measuring real structures. The results obtained showed the robustness and mobility of the proposed method.

**Index Terms**—3-D measurement, 3-D vision, image analysis, remote sensing, structural engineering.

## I. INTRODUCTION

M EASUREMENT and reconstruction systems are an essential tool to assess civil structure deformations. Civil structures represent a principal asset for any country or organization, and due to several factors, these buildings can deform over time. These deformations can affect the visual appearance of the structure in the short term and affect its

Manuscript received June 8, 2020; accepted September 10, 2020. Date of publication September 25, 2020; date of current version November 25, 2020. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2020-03900 and Grant STPGP 521432. The Associate Editor coordinating the review process was Dr. Jing Lei. (*Corresponding author: Haidy Elghamrawy*.)

Moemen Y. Moemen is with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: m.moemen@queensu.ca).

Haidy Elghamrawy is with the Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON K7K 7B4, Canada, and also with the Department of Engineering Mathematics and Physics, Faculty of Engineering, Cairo University, Cairo 12613, Egypt (e-mail: elghamra@rmc.ca).

Sidney N. Givigi is with the School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: sidney.givigi@queensu.ca).

Aboelmagd Noureldin is with the Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON K7K 7B4, Canada, and also with the School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: aboelmagd.noureldin@rmc.ca).

Digital Object Identifier 10.1109/TIM.2020.3026719

structural stability in the long term. The ability to measure and visualize these structures is crucial to perform early preventive maintenance. The importance of having a 3-D model for a given structure was proved to be highly significant in the catastrophic Notre-Dame de Paris fire (France) in 2019. The presence of a reconstructed 3-D model aided the architects and civil engineers to have a visualized reference for the cathedral before the fire.

Modern reconstruction methods utilize 3-D laser scanners (LiDARs) and image-based reconstruction. However, current methods suffer from being static or limited to specific environmental scenarios [1]. Moreover, a lot of the existing techniques require laborious efforts to carry the reconstruction system around for measurement acquisition. Most of the methods reported in the literature depend on a single technology for the reconstruction process, which can make it prone to error and environmental limitations [2]. In this article, the goal is to create a 3-D reconstruction and measurement system that is mobile and able to maneuver in different environments with minimal direct human intervention. The proposed approach combines various vision sensors to provide multiple computational 3-D models with dimensional and volumetric measurements. The models are generated using multiple vision sensors mounted on a mobile wheeled robot, which is wirelessly controlled.

The main contribution and novelty of the proposed approach is a new design for a multiple sensor reconstruction and measurement system that is mobile and expandable. Moreover, the system is implemented on a mobile unmanned ground vehicle (UGV) for increased mobility. For the experiments described in this article, 64-bit complex instruction set computing was used. However, the algorithms presented in this work can be ported to a single-board computer as NVIDIA Jetson. The system realization on single-board computers can introduce possible online reconstruction.

Several experiments were conducted in a controlled and uncontrolled environment to assess the performance of the proposed system. For each experiment, multiple 3-D models are generated and compared against each other to perform a quantitative comparison for each vision sensor utilized. Moreover, experiments on real civil structures were conducted to validate the performance of the method. Indoor experiments were conducted in a controlled environment to assess the proposed system results qualitatively and quantitatively. It was possible to do this comprehensive assessment due to the

technical feasibility of obtaining ground truth dimensional measurements for indoor structures. However, for outdoor experiments, it was only feasible to assess the visualization quality of the 3-D model. Inability to obtain ground truth measurements or CAD models for the outdoor buildings made the quantitative assessment for outdoor buildings technically not feasible.

The remainder of this article is structured as follows. Section II provides a literature review of 3-D measurement and reconstruction techniques. The architecture of the proposed mobile reconstruction system is detailed in Section III. Section IV explains the proposed 3-D reconstruction and measurement methods. In Section V, the experimental setup is illustrated, and the reconstruction of indoor and outdoor object results are presented and discussed. Finally, Section VI gives the conclusion and future work.

## II. LITERATURE REVIEW

The 3-D measurement and reconstruction systems have been widely addressed in the literature. These systems can be categorized according to the vision sensor and reconstruction method utilized in each. Reconstruction and measurement using LiDAR data have been deployed in different applications and research work, such as Stanford's "Digital Michelangelo Project" [3] and building facades modeling [4], to name a few. Measurement systems based on LiDARs have been presented in previous works [5] to provide the 3-D point cloud-based measurement systems. LiDAR-based systems in the literature perform the 3-D measurements with varying accuracies according to the application. For example, Cabaleiro *et al.* [6] presented a deformation measurement method based on the 3-D LiDAR data to assess structures and buildings deformation. Another different approach [7] deployed 3-D point cloud data for measuring mobile robot location in an indoor facility.

Image-based 3-D reconstruction and measurement systems were another approach developed, which is also known as photogrammetry [8]. Photogrammetry has been an active area of research for decades. Contrary to the expensive nature of 3-D LiDARs, the low cost and accessibility of cameras [9] motivated the research community to investigate this problem. Multiple image-based approaches utilizing unorganized photograph groups were introduced [10]. Some of the methods depend primarily on the structure from motion (SFM) algorithm for the 3-D reconstruction of the scenes. A study called desktop versus cloud computing software for 3-D measurement of building façades has been conducted for 3-D image-based measurement. In this study, SFM-based reconstruction technologies, such as Pix4D and Autodesk Remake, were compared and assessed. The conclusion was that image-based reconstruction methods could provide accurate results with high potential to acquire measurements reliably and simply [11].

Reconstruction and measurement systems can be additionally categorized according to the locomotion capability into fixed and mobile measurement systems. Fixed reconstruction systems have been deployed in previous works, depending on monocular vision [12] and stereovision [13] for 3-D reconstruction and measurement. Mobile measurement systems

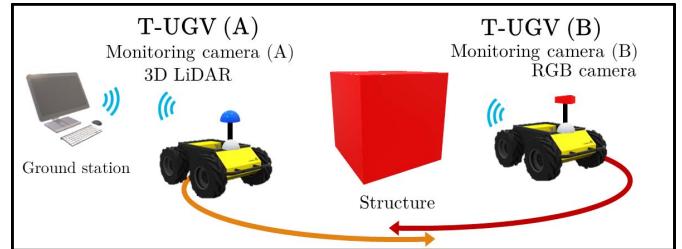


Fig. 1. High-level illustration for the proposed system composed of two T-UGVs equipped with vision sensors.

emerged from the need to reconstruct large stationary objects or buildings, which can even exist in an uncontrolled indoor environment. Autonomous and mobile robots are deployed in different research works to present a flexible and mobile measurement solution. O'Rourke [31] investigated the problem of finding minimal volume boxes circumscribing a given set of 3-D points; it was proved that it is unnecessary for a minimum volume box to have any sides flush with a face of the convex hull. Mobile 3-D measurement evolved in terms of complexity reduction and versatility of deployment scenarios using laser scanners [32]. In addition, camera-based mobile measurement systems were studied and deployed in several research works [33], [34]. Other studies have been conducted using vision-based multiple collaborative robotic platforms that were used to measure and locate 3-D objects [14]. Similar research work was oriented toward crack measurement detection and analysis using mobile measurement systems [15].

## III. SYSTEM ARCHITECTURE

### A. System Overview

The proposed system converts a physical 3-D structure into a computational 3-D model along with its dimensional and volumetric measurements. As shown in Fig. 1, the system consists of two 4-wheel drive teleoperated UGVs; each vehicle acts as a mobile reconstruction and measurement system. Each T-UGV is equipped with a vision sensor, a monitoring camera, and a laptop/smart device. The vision sensor on each T-UGV represents one of the main inputs to the system. The sensors preserve a 2-D RGB image and a 3-D laser scan for the required structure, which will be utilized in the proceeding stages.

The monitoring camera is used for real-time video streaming to the system operator to visualize the surrounding environment for safer teleoperation. During a typical reconstruction task, both T-UGVs must rotate around the desired structure to acquire several images/scans. The generated 3-D model should be able to reflect the physical and dimensional properties of the original structure. Therefore, these acquisitions should represent different perspectives for the object. Finally, the 3-D model can be used for 3-D visualization, meteorological dimensional, and volumetric measurements.

### B. System Pipeline

The proposed system is composed of a multistage pipeline that is divided into startup and configuration, data acquisition,

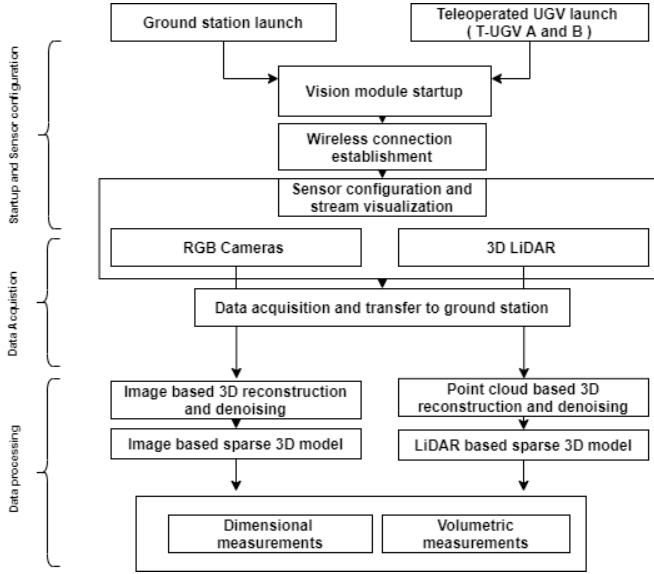


Fig. 2. System pipeline for obtaining a colored 3-D model with dimensional measurements.

and data processing stages. As shown in Fig. 2, the proposed pipeline consists of the following:

- 1) ground station and T-UGVs launch along with machine vision module startup;
- 2) wireless connection establishment between the ground station and the T-UGVs;
- 3) machine vision sensor configuration for data acquisition and remote visualization over a wireless network;
- 4) data acquisition and transference to the ground station for processing;
- 5) 3-D image-based and laser-based model reconstruction;
- 6) 3-D model computation with dimensional and volumetric measurements.

### C. Wireless Communication Paradigm

Several wireless communication channels are utilized in the proposed system for data transmission, visualization, and robot control purposes. As shown in Fig. 3, four Wi-Fi networks and two Bluetooth connections are utilized for two mobile robot systems. Each robot is connected to a dedicated network for vision sensors data visualization and transmission. The second dedicated Wi-Fi network is utilized to transmit the monitoring camera stream to be used by the system operator to guarantee safe navigation. Each mobile robot is teleoperated using a Bluetooth wireless controller.

## IV. 3-D MODEL RECONSTRUCTION AND MEASUREMENT

The 3-D LiDAR and RGB cameras are utilized in the proposed system as primary vision data acquisition sensors. Multivision sensors deployment increases the robustness of the reconstruction and measurement system. The inputs to the 3-D reconstruction module are sensor data, and the outputs are 3-D scaled models with dimensional and volumetric measurements.

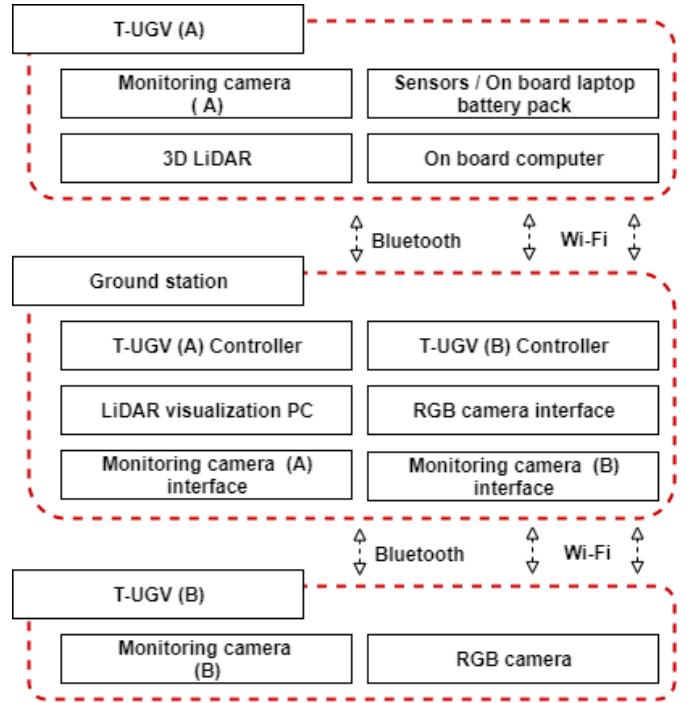


Fig. 3. System communication diagram.

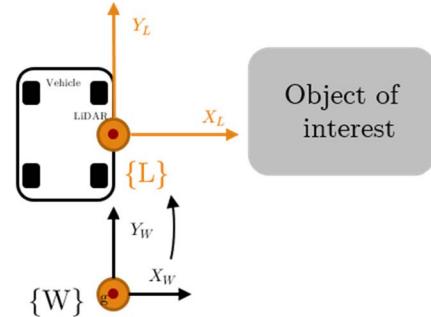


Fig. 4. LiDAR-based reconstruction mobile system coordinate.

### A. LiDAR-Based Point Cloud Model Generation

Point cloud registration and processing is the primary step to acquire a sparse 3-D model from 3-D LiDAR data. In this section, the LiDAR coordinate system will be defined, and the point cloud registration process will be discussed along with the applied data postprocessing operations.

As shown in Fig. 4, LiDAR coordinate system  $\{L\}$  is a 3-D Cartesian coordinate system whose origin is the geometric center of the sensor. LiDARs produce  $N$  number of  $P_c$  scans/second, where  $c$  is the scan number of an arbitrary point cloud where  $c \in \mathbb{Z}^+$ . Each scan  $P_c$  contains a set of 3-D points where each point  $r_i \in P_c$ , and its coordinate in  $\{L\}$  is denoted as  $H_{(c,r_i)}^L$ , as  $i$  is the point index,  $i \in \mathbb{Z}^+$ .

The point cloud registration problem can be defined as having a sequence of point clouds  $P$ , where  $P_c$  is an arbitrary point cloud in this sequence. The objective is to minimize  $\text{dist}(T_{c+1}^L(P_{c+1}), P_c)$ , where  $T_{c+1}^L$  is the six degrees of freedom spatial rigid transformation in  $\text{SE}(3)$  between  $P_c$  and  $P_{c+1}$ , and

also,  $T_{c+1}^L(P_{c+1})$  can be denoted as and  $P'_{c+1}$ , while dist is a distance metric. LiDAR Odometry and Mapping (LOAM) [16] feature-based point registration with nonlinear optimization was deployed in the proposed system. This approach enhances the registration process by depending on point cloud features as lines and planes in LiDAR data. Then, these features are saved to the map for edge-line and plane-surface matching. LOAM method achieves accurate registration with low-drift and low computational complexity utilizing feature-based registration described in [16] as

$$e = \frac{1}{|S| \cdot H_{(c,r)}^L} \left\| \sum_{j \in S, j \neq r} (H_{(c,r)}^L - H_{(c,j)}^L) \right\| \quad (1)$$

where subset point cloud  $S \subset P_{c+1}$  and  $j$  is the closest neighbor for point  $r$  in  $P_c$ . A threshold value of  $e$  is predetermined according to the sensor type; if  $r_i > e$ , then  $r_i \in$  edge or line, and if  $r_i < e$ , then  $r_i \in$  plane. The outcome of the registration process is a set  $\bar{P}$  where

$$\bar{P} = \{\{P'_1\}, \{P'_2\}, \{P'_3\} \dots \{P'_N\}\} \quad (2)$$

$P'_1 \dots P'_N$  are  $N$  point cloud subsets of  $\bar{P}$ , in which each point cloud is registered and aligned with the rest of the point clouds using the LOAM algorithm. Registered point clouds should pass through a series of postprocessing steps after a successful registration. These steps apply different point cloud processing techniques to denoise and merge the registered point clouds to obtain a subset of points representing the object of interest. Since LiDAR point clouds can suffer from noisy measurements due to environmental factors as environments with high reflectivity [17], denoising operations are required to increase the quality of measurements for the 3-D model by rejecting outliers. The denoising procedure can be described as follows:

$$E = \bar{P} - M \quad (3)$$

where  $M$  is a subset representing point cloud outliers as  $\subset \bar{P}$ . The outcome of the denoising process is point cloud  $E$ , where several outliers are eliminated. Obtaining outliers  $M$  is through applying statistical outlier removal (SOR) filter for each point cloud subset  $\bar{P}$ . The SOR filter is part of PCL library [18] for point cloud processing where it is defined as

$$T = \mu_d + \alpha \cdot \sigma_d \quad (4)$$

where  $\mu_d$  is the mean of distances  $d_i$ , which is the average distance between the base point  $r_i$  and  $k$  nearest neighboring points,  $\sigma_d$  is the standard deviation of distances  $d_i$ , and  $\alpha$  is a standard deviation multiplier. The point cloud merging process is required to obtain a single point cloud that has all the acquired measurements. The merging process is defined as

$$E' = \sum_{k=1}^{k=N} E_k. \quad (5)$$

This process merges all subset point clouds in  $E$  to obtain a single point cloud  $E'$ . The process is performed by a box-grid filter that divides the 3-D space into voxels. The points in each voxel are combined into a single output point by averaging

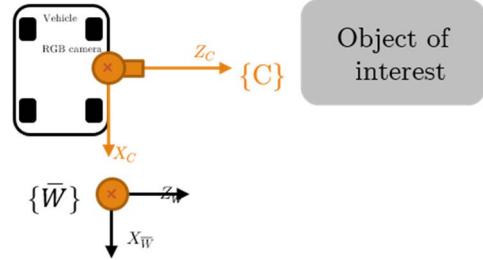


Fig. 5. Image-based reconstruction mobile system coordinate.

the  $X$ ,  $Y$ , and  $Z$  coordinates [19]. The object of interest selection is the final procedure to obtain point cloud  $Q$ , where  $Q \subset E'$ . The point cloud should contain all the 3-D points representing the object of interest. The selection process is performed through manual region selection for points that belongs to the object. These points are excluded from  $E'$  to obtain point cloud  $Q$ , which is the point cloud that will be utilized in visualization and measurements.

### B. Image-Based Point Cloud Model Generation

Image-based model generation is a 2-D to 3-D transformation problem, and SFM is the reconstruction technique deployed in the proposed system. In this section, the camera coordinate system is defined, and the SFM algorithm is explained along with the applied data postprocessing operations. As shown in Fig. 5, the RGB camera coordinate system  $C$  is a 3-D Cartesian coordinate system whose origin is the geometric center of the camera imaging sensor. The SFM image-based reconstruction problem can be defined as having a set of RGB images  $I$  for an object of interest  $D$ , where  $I_c$  is an arbitrary image in set  $I$ ,  $c \in \mathbb{Z}^+$ . The objective is to detect the visual features in each image in  $I$  and match extracted features between images in group  $I$ . Then, the previously matched features are used to extract the 3-D location of a 2-D matched point using multiple-view geometry.

For each generated RGB point cloud  $Y_C$  that consists of a set of points where each point  $X_i \in P_c$  and its coordinate in  $\{C\}$  is denoted as  $H_{(c,X_i)}^C$ , as  $i$  is the point order,  $i \in \mathbb{Z}^+$ . Several SFM algorithms have been developed [20]; however, VisualSfM is the SFM implementation deployed in the proposed system due to the following considerations [21].

- 1) The algorithm makes a good balance between speed and accuracy.
- 2) Scale-invariant feature transform (SIFT)-GPU [22] enhances the speed of the feature extraction process.
- 3) Preemptive feature matching enhances the speed of the matching process.
- 4) BA algorithm reduces the time complexity.

The image-based reconstruction process is divided into three procedures, namely, feature extraction, feature matching, and geometry estimation. Feature extraction is based on extracting visual information through describing image patches; each image patch is mapped into a feature vector, which is called “descriptor” [23]. In the proposed system, the SIFT descriptor was utilized using SIFT GPU implementation. Conventional

CPU-based SIFT implementations have a drawback of relatively high features extraction processing time, especially for large image sets. This drawback was overcome using the GPU-based SIFT implementation.

The feature matching process is required to find the corresponding features between different images. Preemptive feature matching [21] was deployed in the proposed system. The method can reduce the pairs of images matching up to 95% while keeping enough good matches for 3-D reconstruction. The matching process can be further enhanced by suitable vehicle navigation around the object of interest during the data acquisition stage.

The geometry estimation stage aims to compute essential matrix  $\mathbf{E}$ , which defines the relationship between the corresponding 2-D points  $(x, x')$  in different images using a pinhole camera model [24]. The epipolar constraint is defined as

$$0 = (x')^T \mathbf{E} x. \quad (6)$$

The bundle adjustment algorithm is used to refine the reconstructed 3-D points and camera parameters. Multicore bundle adjustment [25] is utilized in our proposed system to enhance the speed of the nonlinear optimization process.

The 3-D image-based model will require sequential post-processing operations composed of point cloud-scale correction, point cloud denoising, and object of interest selection stage. The proposed system depends on a monocular camera in 3-D image-based reconstruction. It is known that images taken by monocular cameras suffer from scale ambiguity [26]; this problem leads to the recovering of 3-D points up to a scale. This problem was solved by using a reference object of known dimensions that might be present in the environment of a reconstruction or can be induced into it. The reference object is utilized to obtain a scale correction factor  $\alpha$ , where the scale correction can be defined as

$$\bar{Y} = \alpha Y \quad (7)$$

where  $Y$  is the unscaled point cloud and  $\bar{Y}$  is the scale corrected point cloud. The scale correction factor  $\alpha$  is defined as the ratio between the reference object pre-known measurement and the measurement obtained in the reconstructed 3-D point cloud. Similar to the 3-D LiDAR case, the latter post-processing operations are reused to denoise and crop the object of interest from the main point cloud. The techniques and formulation described in Section IV-A can be reused successfully by utilizing SOR filter and manual region selection. Point cloud  $Q'$  is the final output of the image-based reconstruction module, where all its points are utilized in the measurement and visualization of the object of interest.

### C. Dimensional and Volumetric Measurements

The proposed system provides the 3-D visualization for a physical object of interest along with dimensional and volumetric measurements. Point clouds  $Q'$  and  $Q$  are the outputs of each 3-D reconstruction method deployed in the proposed multiple vision sensor system. For each point cloud, two sets of measurements are automatically calculated. The measurements are divided into bounding box dimensions and

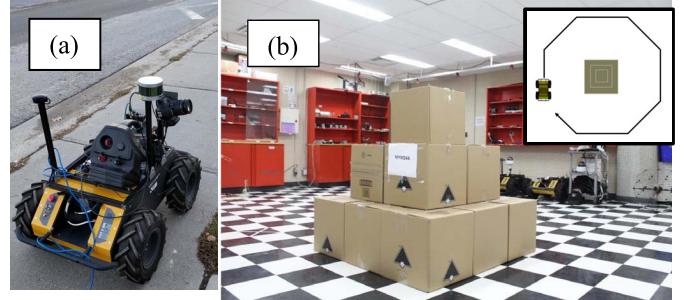


Fig. 6. (a) Clearpath Husky UGV with vision sensors mounted. (b) Pyramid structure as an object of interest along with the motion path illustrated.

convex hull volume measurements. A bounding box for a set of points is a closed volume that completely contains the union of these points [27]. The bounding box measurements are dimensional measurements divided into length, width, and height of the box.

The usage of bounding boxes simplifies the measurement process for a 3-D point cloud. The method is simple to implement; however, it can suffer from considerable inaccuracy when used within noncube or cuboid-like structures. Convex hull volume measurement is the second measurement obtained from the proposed measurement system. A convex hull of a shape can be defined as the smallest convex set that contains it [28]. The quickhull computation algorithm proposed by Wu *et al.* [29] has been deployed in the proposed measurement system. Convex hull volumetric computations are usually computationally expensive; however, the convex hull can deform according to the geometry of the point cloud. The obtained measurements aim to present a physical quantitative measure for the 3-D reconstructed models.

## V. EXPERIMENTS

### A. Experimental Setup

The experimental setup consists of two teleoperated Clearpath Husky UGV robots with vision sensors mounted on each, as shown in Fig. 6(a). Both robots are connected to the ground station using the Wi-Fi and Bluetooth communication channels. The vision sensors equipment and communication paradigm were detailed in Section III. Data acquisition scalability is one of the features in the proposed reconstruction and measurement system. Data acquisition time for images and 3-D scans can be reduced using multirobot systems, especially when scanning relatively large buildings. While the scanning time and area increase proportionally with the size of the structure, deploying more than one T-UGV reduces the data acquisition time and effort.

The main objective of the experiments is to reconstruct and measure a 3-D structure to evaluate the visualization and measurement capabilities of the proposed measurement system. The indoor setup was made to verify the measurement and system outputs in a controlled environment. The pyramid structure that is shown in Fig. 6(b) was used in the experiments, where the first two levels in the pyramid are used for measurement analysis and visualization.

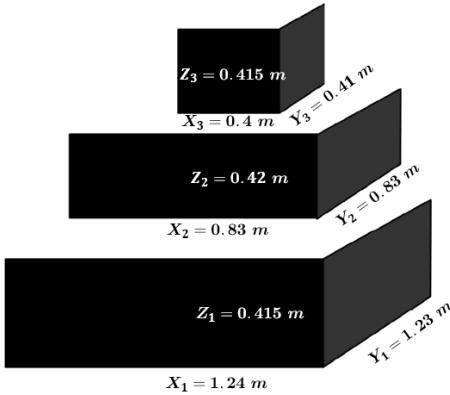


Fig. 7. Pyramid structure with respective dimensions in meters.

The third-level cube is used as a reference object for scale correction of the image-based reconstruction. Fig. 7 shows the pyramid model with its respective dimensions. The dimensional reference values are measured using a tape measure, which has an uncertainty of  $\pm 0.05$  cm. Reference volumetric measurements are computed using the dimensional reference values. The experiment required two operators to maneuver each robot around the object of interest. During maneuvers, each robot is acquiring images and/or 3-D scans; both robots rotated around the object in circular trajectories. The data collected from both robots is wirelessly transferred to the ground station to be processed and analyzed.

### B. Experimental Environment

The pyramid structure experiments were indoor experiments conducted in a controlled environment. The lighting conditions were constant, and no external sunlight was introduced into these experiments.

These experiments' objective was to provide a suitable environment that can allow quantitative analysis for the system outcomes. Another outdoor experiment was performed to assess the visualization quality of the proposed method. This experiment was conducted in an uncontrolled environment and had different light variations. The experiments were performed to evaluate and validate the system capabilities in both indoor and outdoor scenarios.

### C. Application of 3-D Reconstruction and Measurement

In order to assess the quality of the 3-D reconstruction and evaluate the error in the measurements, ten data acquisitions had been processed using the proposed system. During each data acquisition, both UGVs acquired images using Canon T6 camera and 3-D scans using a Velodyne VLP 16 at the same time. The robots were controlled by a human operator using a wireless controller, while each operator was watching the robot's surrounding environment through the video feed transmitted from each monitoring camera. For a single experiment, the average number of LiDAR point scans acquired was 1000 scans, whereas the average number of RGB images acquired was 350 images per experiment.

The LiDAR point clouds and RGB images are used as inputs to the 3-D reconstruction and measurement module to compute the 3-D models. Then, the developed system outputs the 3-D volumetric and dimensional measurements. The dimensional measurements were calculated using the bounding box method, where the relative dimensional measurement error is defined as

$$d_e^{U_i} = \frac{U'_i - U_i}{U_i} \cdot 100 \quad (8)$$

where  $d_e^{U_i}$  is the relative measurement error for specific  $U$  side in the pyramid structure in a given level  $i$ ,  $U = \{X, Y, Z\}$ ,  $i = \{1, 2\}$ ,  $U_i$  is the absolute value, and  $U'_i$  is the measured value. The volumetric measurements were calculated using the convex hull volume computation, where the relative volumetric error is defined as

$$v_e^i = \frac{V'_i - V_i}{V_i} \cdot 100 \quad (9)$$

where  $v_e^i$  is the relative volumetric error for each level  $i$  in the pyramid structure. The measurement errors for LiDAR and image-based techniques are shown in Tables I–III. The LiDAR related measurements have a superscript  $L$  and image or camera-related measurements have a superscript  $C$ , whereas error measurements are defined as in (8) and (9). It can be noted from Table I that the average errors for  $Z$  dimension of ground level in both the LiDAR and image-based models are significantly higher than the ones of the other dimensions. Raw LiDAR and image-based point clouds consist of 3-D points representing the whole environment captured with vision sensors. The captured data include the object of interest, multilevel pyramid in our work, and other different objects in the environment. Raw 3-D point clouds usually need to be filtered and segmented to obtain a subset of 3-D points representing the object of interest in the environment. In our proposed system, the segmentation process is done by selecting a subset (voxel) from the whole point cloud, and the voxel should represent the object of interest. The procedure includes separating ground points from the desired 3-D model points. The process is subjected to a margin of error due to the segmentation error when selecting the voxel. This error directly affects the accuracy of measurement for scanned 3-D structures specially in the lower portions of the structure which are nearer to the ground level (Level 1).

### D. Analysis of the results

It can be observed that measurements obtained from both 3-D models have different degrees of uncertainty. Image-based models have uncertainty type A of  $\pm 3.76\%$  in dimensional measurements and  $\pm 6.04\%$  in volumetric measurements [30].

Dimensional measurement uncertainty is caused by different factors, such as environment illumination, rolling shutter effect of the DSLR camera, and inaccuracies in camera matrix estimation in VisualSfM. Volumetric measurements were calculated using the convex hull volume method, as shown in Table III.

Measurements uncertainty occurs due to low point density portions in the 3-D model generated. This low model density can occur due to poor vision and texture features in

TABLE I  
DIMENSIONAL MEASUREMENT FOR LEVEL 1 FROM LiDAR AND IMAGE-BASED 3-D MODELS

Trial	$X_1^L$	$d_e^{X_1}(\%)$	$X_1^C$	$d_e^{X_1}(\%)$	$Y_1^L$	$d_e^{Y_1}(\%)$	$Y_1^C$	$d_e^{Y_1}(\%)$	$Z_1^L$	$d_e^{Z_1}(\%)$	$Z_1^C$	$d_e^{Z_1}(\%)$
True value (m)	1.24					1.23					0.41	
1	1.29	3.61	1.22	1.61	1.35	8.95	1.19	2.68	0.45	8.43	0.42	2.43
2	1.32	6.02	1.27	2.41	1.35	8.95	1.32	7.31	0.51	22.89	0.45	9.75
3	1.33	6.82	1.28	3.22	1.33	7.34	1.31	6.5	0.5	20.48	0.46	12.19
4	1.3	4.41	1.21	2.41	1.34	8.15	1.21	1.62	0.51	22.89	0.44	7.31
5	1.3	4.41	1.31	5.64	1.31	5.73	1.35	9.75	0.55	32.53	0.47	14.63
6	1.31	5.22	1.28	3.22	1.37	10.57	1.33	8.13	0.54	30.12	0.46	12.19
7	1.32	6.02	1.28	3.22	1.32	6.53	1.36	10.56	0.49	18.07	0.48	17.07
8	1.26	1.2	1.31	5.64	1.35	8.95	1.34	8.94	0.48	15.66	0.46	12.19
9	1.25	0.4	1.23	0.8	1.36	9.76	1.32	7.31	0.53	27.71	0.45	9.75
10	1.26	1.2	1.28	3.22	1.37	10.57	1.25	1.62	0.51	22.89	0.44	7.31
Avg. error	3.93%			3.13%	8.55%			6.44%	22.16%			10.48%

TABLE II

DIMENSIONAL MEASUREMENT AVERAGE ERRORS FOR LEVEL 2  
FROM LiDAR AND IMAGE-BASED 3-D MODELS

	LiDAR 3D Measurements average relative errors			Image-based 3D Measurements average relative errors		
	$d_e^{X_2}(\%)$	$d_e^{Y_2}(\%)$	$d_e^{Z_2}(\%)$	$d_e^{X_2}(\%)$	$d_e^{Y_2}(\%)$	$d_e^{Z_2}(\%)$
True value (m)	0.83	0.83	0.42	0.83	0.83	0.42
Avg.error	6.25%	11.8%	3.8%	3.97%	5.77%	4.52%

TABLE IV

STATISTICAL COMPARISON BETWEEN LiDAR AND IMAGE-BASED  
3-D MODEL MEASUREMENTS

	Dimensional measurement error			Volumetric measurement error (Convex hull)			Points/M odel
	Meth od	min	max	$\pm$	min	max	
Came ra	3.97 %	10.48 %	3.76 %	7.192 %	11.24 %	9.11 %	7k
LiDA R	4.36 %	22.89 %%	7.45 %	6.56 %	22.02 %	10.08 %	38k

TABLE III

VOLUMETRIC MEASUREMENTS FOR LEVELS 1 AND 2 USING  
LiDAR AND IMAGE-BASED 3-D MODELS

Trial	$V_1^L$	$v_e^1$	$V_1^C$	$v_e^1$	$V_2^L$	$v_e^2$	$V_2^C$	$v_e^2$		
True value	0.64 m <sup>3</sup>					0.28 m <sup>3</sup>				
1	0.66	3.12	0.55	14.06	0.31	7.26	0.23	20.41		
2	0.77	20.31	0.67	4.68	0.33	14.18	0.28	3.11		
3	0.77	20.31	0.71	10.93	0.32	10.72	0.3	3.8		
4	0.78	21.87	0.56	12.5	0.31	7.26	0.25	13.49		
5	0.85	32.81	0.76	18.75	0.29	0.34	0.3	3.8		
6	0.78	21.87	0.71	10.93	0.29	0.34	0.29	0.34		
7	0.8	25.0	0.74	15.62	0.31	7.26	0.28	3.11		
8	0.78	21.87	0.75	17.18	0.3	3.8	0.32	10.72		
9	0.83	29.68	0.68	6.25	0.3	3.8	0.27	6.57		
10	0.79	23.43	0.65	1.56	0.32	10.72	0.27	6.57		
Avg.	22.02%			11.24%	6.56%			7.19%		

TABLE V  
PERFORMANCE COMPARISON BETWEEN IMAGE-BASED  
AND LiDAR-BASED DEPLOYED TECHNIQUES

Operation	Image-based	LiDAR-based
- Setup preparation	20 minutes	20 minutes
- Experiment	5 minutes	10 minutes
- Feature extraction and matching	86 minutes	-
- Reconstruction/Registration	53 minutes	10 minutes
- Post processing	25 minutes	24 minutes
<b>Total time</b>	<b>189 minutes</b>	<b>64 minutes</b>

points/model metric, which can present a suitable visualization model.

On the other hand, the image-based model measurements' uncertainty is lower than LiDAR-based ones in most cases. In terms of performance, each method has different low-level procedures that contribute directly to the total time of 3-D model reconstruction. The internal steps deployed methods are shown in Table V. It can be observed that LiDAR-based reconstruction time in order of magnitude is lower than the image-based method. The main factor affecting the image-based techniques is the time required for feature extraction and matching process. This process solely takes

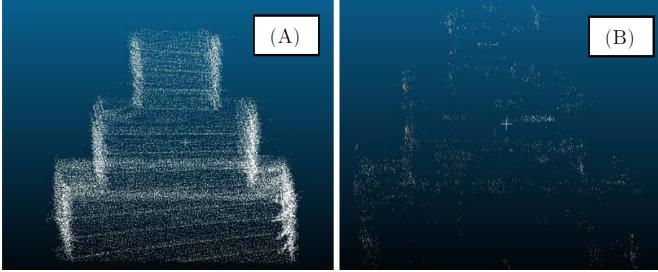


Fig. 8. (a) LiDAR-based 3-D model visualization. (b) Image-based 3-D model visualization.



Fig. 9. Reconstruction and measurement system trajectory.

1.3 more time than the whole 3-D LiDAR model generation process.

Visualization of the 3-D model is a visual representation of the object of interest. In the previous context, each presented modeling method in the proposed system has its point of strength and weakness. On one side, image-based models generate RGB colored 3-D model, which enhances the visual experience; however, the downside of this method is that it is highly dependent on visual features. This can be observed in Fig. 8(b), where the only reconstructed regions are edges and sharp corners. Other regions of the pyramid structure did not have many visual features; hence, these regions are not reconstructed. On the contrary, the LiDAR-based model presents much denser 3-D models, which can help to visualize the whole structure without dependence on texture or visual features. The downside of this method is that it generates an uncolored 3-D model.

The performance of the developed system is evaluated in an uncontrolled open environment to reconstruct the façade of Mackenzie building in the Royal Military College of Canada (RMCC). This is a 93-m-wide façade of a four-story building. Fig. 9 shows the trajectory conducted to 3-D reconstruct this building. The LiDAR-based generated model for the experiment consisted of 3.5M uncolored points, whereas the image-based model consisted of 45k colored points. As shown in Fig. 10, each 3-D model represents different aspects of visualization. The image-based model represents a low-density colorful representation, whereas the LiDAR-based model represents a high-density uncolored representation. The outdoor experiment validates the applicability of the proposed system for measurement tasks with different scales.

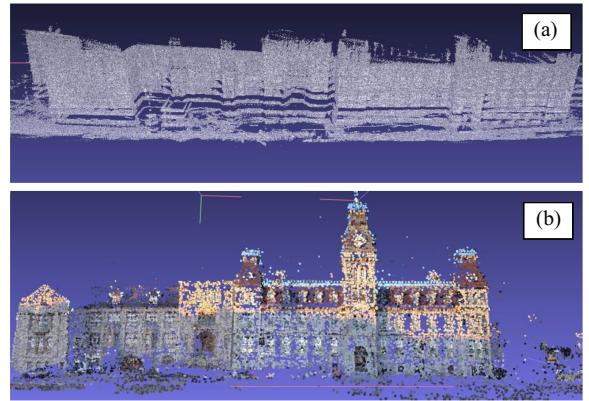


Fig. 10. (a) LiDAR-based 3-D model visualization. (b) Image-based 3-D model visualization.

## VI. CONCLUSION AND FUTURE WORK

This article presents a novel mobile 3-D measurement and reconstruction system using multiple vision sensors for both indoor and outdoor applications. The proposed system was assessed using numerous indoor and outdoor experiments.

According to the proposed method, a sequence of RGB images and LiDAR scans were processed to compute image- and LiDAR-based 3-D models for a measured object. The 3-D measurements were computed; moreover, the 3-D visualization of the measured object was generated. Multiple vision sensors deployment was able to offer a robust reconstruction and measurement system. The statistical analysis of the 3-D model error demonstrated the points of strength and weakness of each method deployed. It was concluded that LiDAR-based 3-D models were a better fit for visualization purposes while having a more significant error range due to point cloud misregistration. Image-based 3-D models had a lower error in both dimensional and volumetric measurements. However, the method requires a reference object for scale correction. Besides, it failed to visualize low textured objects.

A few enhancements can be proposed to improve measurement accuracy and visualization. First, a 3-D LiDAR with higher resolution can be used to improve accuracy. Second, a fusion algorithm can couple the acquired measurements from both vision sensors presenting a single 3-D model from both sensors. Third, 3-D information acquired from the vision sensors can be used for perception and path planning for automated vehicle movement instead of teleoperation. In addition, the image-based reconstruction pipeline can be optimized to decrease the total processing time. Moreover, the 3-D measurements from the proposed technique could be used to test structural deformations. Finally, the combination of UGVs and unmanned aerial vehicles (UAVs) may present better 3-D reconstruction results specially for high altitude points. UAVs with LiDARs and cameras are widely used in mapping applications from high altitudes. However, a UAV that can carry and power a 3-D LiDAR or a high-quality DSLR camera is relatively expensive compared to ground vehicles. In terms of deployment, UAVs can be very useful to scan the high points of buildings and rooftops, but it can be challenging to control the UAV at low altitudes to scan lower sections of buildings.

## REFERENCES

- [1] X. Brunetaud, L. D. Luca, S. Janvier-Badosa, K. Beck, and M. Al-Mukhtar, "Application of digital techniques in monument preservation," *Eur. J. Environ. Civil Eng.*, vol. 16, no. 5, pp. 543–556, May 2012.
- [2] J. Guo, J. Jiang, L. Wu, W. Zhou, and L. Wei, "3D modeling for mine roadway from laser scanning point cloud," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Beijing, China, Jul. 2016, pp. 4452–4455.
- [3] M. Levoy *et al.*, "The digital Michelangelo project: 3D scanning of large statues," in *Proc. 27th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, Chicago, IL, USA, 2000, pp. 131–144.
- [4] G. Vosselman and I. Suveg, "Map based building reconstruction from laser data and images," *Autom. Extraction Man-Made Objects Aerial Space Images (III)*, pp. 231–239, Jan. 2001.
- [5] D. Kong, L. Xu, X. Li, and S. Li, "K-plane-based classification of airborne LiDAR data for accurate building roof measurement," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1200–1214, May 2014.
- [6] M. Cabaleiro, B. Riveiro, B. Conde, and A. Sánchez-Rodríguez, "A case study of measurements of deformations due to different loads in pieces less than 1 m from lidar data," *Measurement*, vol. 151, Feb. 2020, Art. no. 107196.
- [7] Z. Huang, J. Zhu, L. Yang, B. Xue, J. Wu, and Z. Zhao, "Accurate 3-D position and orientation method for indoor mobile robot navigation based on photoelectric scanning," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 9, pp. 2518–2529, Sep. 2015.
- [8] E. Trucco, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [9] L. Ferrigno, A. Pietrosanto, and V. Paciello, "Low-cost visual sensor node for bluetooth-based measurement networks," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 2, pp. 521–527, Apr. 2006.
- [10] N. Snavely, S. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," *ACM Trans. Graph.*, vol. 25, pp. 835–846, Jul. 2006.
- [11] S. Peña-Villasenín, M. Gil-Docampo, and J. Ortiz-Sanz, "Desktop vs cloud computing software for 3D measurement of building façades: The monastery of San Martín Pinario," *Measurement*, vol. 149, Jan. 2020, Art. no. 106984.
- [12] Q. Wu, X. Wang, L. Hua, and G. Wei, "The real-time vision measurement of multi-information of the bridge crane's workspace and its application," *Measurement*, vol. 151, Feb. 2020, Art. no. 107207.
- [13] T. Xue, L. Qu, and B. Wu, "Matching and 3-D reconstruction of multibubbles based on virtual stereo vision," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1639–1647, Jun. 2014.
- [14] A. Nucher, H. Surmann, and J. Hertzberg, "Automatic model refinement for 3D reconstruction with mobile robots," in *Proc. 4th Int. Conf. 3-D Digit. Imag. Modeling*, Banff, AB, Canada, Oct. 2003, pp. 394–401.
- [15] D. Hähnel, W. Burgard, and S. Thrun, "Learning compact 3D models of indoor and outdoor environments with a mobile robot," *Robot. Auto. Syst.*, vol. 44, no. 1, pp. 15–27, Jul. 2003.
- [16] L. M. Paz, P. Pinies, J. D. Tardos, and J. Neira, "Large-scale 6-DOF SLAM with stereo-in-hand," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 946–957, Oct. 2008.
- [17] S. Hussmann, D. Schauer, and B. MacDonald, "Integration of a 3D-TOF camera into an autonomous, mobile robot system," in *Proc. IEEE Instrumentation Meas. Technol. Conf.*, Singapore, May 2009, pp. 547–552.
- [18] R. G. Lins, S. N. Givigi, and P. R. G. Kurka, "Vision-based measurement for localization of objects in 3-D for robotic applications," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 11, pp. 2950–2958, Nov. 2015.
- [19] R. G. Lins and S. N. Givigi, "Automatic crack detection and measurement based on image analysis," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 583–590, Mar. 2016.
- [20] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Proc. Robot., Sci. Syst. X*, Berkeley, CA, USA, Jul. 2014, pp. 55–63.
- [21] S.-W. Yang and C.-C. Wang, "On solving mirror reflection in LIDAR sensing," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 2, pp. 255–265, Apr. 2011.
- [22] (2017). *PCL Point Cloud Library*. Accessed: 12, 2019. [Online]. Available: <http://pointclouds.org/>
- [23] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets," *Auto. Robots*, vol. 34, no. 3, pp. 133–148, 2013.
- [24] X. Xu, M. Yasser, H. Ragab, Y. Gao, A. Noureldin, and K. He, "Visual structure from motion for UAV indoor localization," in *Proc. 32nd Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, Miami, FL, USA, Oct. 2019, pp. 416–425.
- [25] C. Wu, "Towards linear-time incremental structure from motion," in *Proc. Int. Conf. 3D Vis.*, Seattle, WA, USA, Jun. 2013, pp. 127–134.
- [26] C. Wu. *SiftGPU*. Accessed: Oct. 1, 2020. [Online]. Available: <http://wiki.ros.org/siftgpu>
- [27] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, Sep. 1999, pp. 1150–1157.
- [28] R. Hartley, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [29] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz, "Multicore bundle adjustment," in *Proc. CVPR*, Providence, RI, USA, Jun. 2011, pp. 3057–3064.
- [30] P. Gakne and K. O'Keefe, "Tackling the scale factor issue in a monocular visual odometry using a 3D city model," in *Proc. Int. Tech. Symp. Navigat. Timing*, Toulouse, France, Nov. 2018, pp. 1–11.
- [31] J. O'Rourke, "Finding minimal enclosing boxes," *Int. J. Comput. Inf. Sci.*, vol. 14, no. 3, pp. 183–199, Jun. 1985.
- [32] R. T. Rockafellar, *Convex Analysis* (Princeton Mathematical Series). Princeton, NJ, USA: Princeton Univ. Press, 1970.
- [33] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, Dec. 1996.
- [34] *Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement*, (GUM 1995 With Minor Corrections), Standard JCGM 100:20082008, Joint Committee for Guides in Metrology, 2008.



**Moemen Y. Moemen** received the M.Sc. degree in electrical and computer engineering from Queen's University, Kingston, ON, Canada, in 2019.

He was a member of the Navigation and Instrumentation (NavINST) Research Group, Kingston. He is currently a software and computer vision developer at IPG Photonics, Canada. His main research interests include computer vision, machine learning, and robotics.



**Haidy Elghamrawy** received the Ph.D. degree in electrical and computer engineering from Queen's University, Kingston, ON, Canada, in 2019.

She is currently a Post-Doctoral Fellow with the Royal Military College of Canada (RMCC). Her research interests include GNSS, antijamming techniques, computer vision, sensor fusion, and autonomous cars.



**Sidney N. Givigi** (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2009.

He is currently an Associate Professor with the School of Computing, Queen's University, Kingston, ON, Canada. His current research interests include autonomous systems and robotics.



**Aboelmagd Noureldin** (Senior Member, IEEE) is currently a Professor with the Department of Electrical and Computer Engineering, Royal Military College of Canada (RMCC), Kingston, ON, Canada, with Cross-Appointment at both the School of Computing and the Department of Electrical and Computer Engineering, Queen's University, Kingston. He is also the Founder and the Director of the Navigation and Instrumentation (NavINST) Research Group, RMCC. He has been a leader in the field of inertial navigation, global navigation satellite systems including GPS, wireless location, and navigation, intelligent multisensor systems with applications related to positioning and navigation of autonomous vehicles and mobile robot systems, machine learning-based human activity recognition and positioning, integrated wireless indoor navigation, intelligent transportation, road information services, crowd management, and Vehicular Internet of Things.