

Low-cost 3D scanning systems for cultural heritage documentation

3D scanning
systems for
cultural
heritage

Quentin Kevin Gautier

*Department of Computer Science and Engineering,
University of California San Diego, La Jolla, California, USA*

Thomas G. Garrison

*Department of Geography and the Environment, University of Texas at Austin,
Austin, Texas, USA*

Ferrill Rushton

*Department of Electrical and Computer Engineering,
University of California San Diego,
La Jolla, California, USA*

Nicholas Bouck, Eric Lo and Peter Tueller

*Department of Computer Science and Engineering,
University of California San Diego, La Jolla, California, USA*

Curt Schurgers

*Department of Electrical and Computer Engineering,
University of California San Diego,
La Jolla, California, USA, and*

Ryan Kastner

*Department of Computer Science and Engineering, University of California San Diego,
La Jolla, California, USA*

Received 1 March 2020
Revised 13 April 2020
Accepted 13 April 2020

Abstract

Purpose – Digital documentation techniques of tunneling excavations at archaeological sites are becoming more common. These methods, such as photogrammetry and LiDAR (Light Detection and Ranging), are able to create precise three-dimensional models of excavations to complement traditional forms of documentation with millimeter to centimeter accuracy. However, these techniques require either expensive pieces of equipment or a long processing time that can be prohibitive during short field seasons in remote areas. This article aims to determine the effectiveness of various low-cost sensors and real-time algorithms to create digital scans of archaeological excavations.

The authors would like to thank the following persons for their help and contribution to this work: Kenny Criddle, Connie Du, Andres Gomez, Proud Heng, Nathan Hui, Tim Jiang, Waseem Khan, Etsu Nakahara, Giovanni Vindiola, Danbing Zhu. This work was funded in part by the REU Site Engineers for Exploration, supported by the NSF under Grant No. 1852403. Additional support was provided by the UCSD ECE and CSE departments and the UCSD Qualcomm Institute. The Guatemala portion of this research was conducted under permits from Instituto de Antropología e Historia (IDAEH) and the Ministerio de Cultura y Deportes. The Centro de Estudios Conservacionistas of the Universidad de San Carlos provide access to the field camp. Tunnel excavations in El Zotz Structure M7-1 as well as field infrastructure for the present study were supported by grants to Garrison from the Fundación Pacunam from 2012 to 2019. Conservation of stucco masks in M7-1 was funded by a National Geographic Society Conservation Trust Grant (C233-13) and the Fundación Pacunam. Garrison thanks Rony Piedrasanta (2015-present) and Lic. Jose Luis Garrido (2012) for field assistance with the M7-1 dig, and Dr. Edwin Román (2012-2015) and Licda. Yeny Gutiérrez (2016-present) for logistical support as project co-directors.



Journal of Cultural Heritage
Management and Sustainable
Development
© Emerald Publishing Limited
2044-1266
DOI 10.1108/JCHMSD-03-2020-0032

Design/methodology/approach – The authors used a class of algorithms called SLAM (Simultaneous Localization and Mapping) along with depth-sensing cameras. While these algorithms have largely improved over recent years, the accuracy of the results still depends on the scanning conditions. The authors developed a prototype of a scanning device and collected 3D data at a Maya archaeological site and refined the instrument in a system of natural caves. This article presents an analysis of the resulting 3D models to determine the effectiveness of the various sensors and algorithms employed.

Findings – While not as accurate as commercial LiDAR systems, the prototype presented, employing a time-of-flight depth sensor and using a feature-based SLAM algorithm, is a rapid and effective way to document archaeological contexts at a fraction of the cost.

Practical implications – The proposed system is easy to deploy, provides real-time results and would be particularly useful in salvage operations as well as in high-risk areas where cultural heritage is threatened.

Originality/value – This article compares many different low-cost scanning solutions for underground excavations, along with presenting a prototype that can be easily replicated for documentation purposes.

Keywords Archaeology, Cultural heritage, Documentation, Surveying and recording, Mapping

Paper type Research paper

1. Introduction

Digital metric documentation of cultural heritage is becoming commonplace at archaeological sites where high-resolution, remotely sensed data provide a more accurate record than traditional analog recording of excavation contexts, architecture and ancient monuments (e.g. [Garrison et al., 2016](#)). The collection of this data can come at great financial cost, as with terrestrial LiDAR (light detection and ranging) systems, or be time-consuming during postprocessing, as with photogrammetry. In some archaeological contexts, these are not effective solutions for documenting cultural heritage. Landscapes with extensive damage from looting and other destructive activities require too much field time for sensitive, expensive instruments, and a lack of real-time results can leave doubt about whether the sensors effectively documented the targets.

Large-scale aerial LiDAR acquisitions in Belize ([Chase et al., 2014](#)) and Guatemala ([Canuto et al., 2018](#)) reveal thousands of previously undetected structures created by the ancient Maya civilization (1000 BCE–1519 CE), but also illuminate the extent to which unchecked, illicit looting has damaged this landscape. Ground truthing of aerial LiDAR data would benefit from the ability to simultaneously acquire 3D documentation of looted contexts to more accurately quantify damage in cultural heritage management assessments. The recent wanton destruction of archaeological sites, such as Palmyra, by the Islamic State of Iraq and Syria (ISIS) presents an extreme case of a high-risk zone where recovery efforts could have benefited from a real-time, rapid documentation system to provide damage assessment estimates. Unfortunately, such incidents are not unique, and archaeology would benefit from a rapid 3D scanning sensor to record damage and plan recovery.

This article presents the details of prototype development and testing for just such a system. While this new system is not as spatially accurate as more expensive or time-consuming methods, we believe that there are sufficient examples of extensively looted archaeological landscapes and zones with high risk for cultural heritage destruction that a rapid, cost-effective 3D documentation system such as the one presented here would be an asset in cultural heritage management. The system was tested in ancient Maya archaeological contexts and refined in mud caves near the University of California, San Diego, where we developed the prototype.

2. Background

Our prototype is designed to efficiently document underground or other restricted spaces such as those presented by damaged ancient architecture. There exist multiple ways of collecting three-dimensional (3D) data to reconstruct a large underground or indoor

environment, which we can group into three main categories: (1) scans using terrestrial LiDAR, (2) scans using photogrammetry and (3) scans based on Simultaneous Localization and Mapping (SLAM) algorithms with lower-cost sensors.

2.1 LiDAR

Terrestrial LiDAR scanners are high-precision devices that can capture the geometry of their surrounding environment. LiDARs emit lasers in many directions and calculate the distance to the device by measuring the time delay to the reflected signal. These devices are often used for cultural heritage documentation due to their high accuracy. This type of documentation is popular in Maya archaeology (Garrison *et al.*, 2016; Cosme *et al.*, 2014), which means that there is good, high-resolution baseline 3D data to compare against other sensors and methods. Individual LiDAR scans can generally achieve millimeter accuracy and a few centimeters accuracy when combined (Grant *et al.*, 2012); however, the process of collecting the data over extensive excavation sections is very long. Furthermore, the final result is obtained after a lengthy postprocessing step that cannot be performed directly in the field, preventing a quick visualization of the current scan. Finally, the cost of high-quality laser scanners is often a barrier to large-scale deployment or deployment in difficult environments (e.g. narrow, dusty tunnels) where the equipment is at risk of being damaged.

2.2 Photogrammetry

“Structure-from-motion” photogrammetry is a general technique that uses computer vision algorithms to reconstruct a 3D shape from a set of 2D pictures (Luhmann *et al.*, 2013). It is a very accessible solution to scan different types of environments; the user only needs a digital camera and potentially a good light source. This technique also provides good results for cultural heritage documentation (Yastikli, 2007; Fassi *et al.*, 2013). The largest drawback is the computational time required to process the data. Due to this limitation, it can be difficult to evaluate the results in the field. Additionally, the computation time grows very rapidly with the number of photos, and that number grows quickly with the size of the area to scan, which renders the scan of an entire excavation very difficult. The sometimes-cramped conditions of archaeological contexts with fragile cultural heritage monuments can also be a deterrent to photogrammetric methods that require the photographer to be in close proximity to sensitive features for extended periods of time.

2.3 Simultaneous localization and mapping

SLAM algorithms are a class of algorithms – first defined in 1986 (Smith and Cheeseman, 1986) – focused on localizing an agent within an unknown environment while simultaneously building a map of this environment (Durrant-Whyte and Bailey, 2006). The goal of SLAM is to provide an update of the position and/or a map in real time. The definition of real time varies with the application, but in our case, we target an update around the same speed as camera sensors, that is, 30 Hz. We are particularly interested in the use of visual SLAM algorithms, utilizing visual sensors, but SLAM works with any number and combinations of sensors (Kohlbrecher *et al.*, 2011; Meyer *et al.*, 2011; Li *et al.*, 2014; Forster *et al.*, 2017; Ibragimov and Afanasyev, 2017; Mur-Artal and Tardós, 2017).

In parallel with the development of SLAM algorithms, many low-cost visual and depth sensors have been commercialized in the past few years. The Microsoft Kinect and the Intel RealSense are two examples of such sensors, which combine an RGB camera with a depth sensor that provides geometrical information about the scene.

The combination of SLAM algorithms and low-cost sensors yields a good solution to the problem of obtaining quick 3D scans of a scene. However, the quality of results is unknown and can be difficult to assess for multiple reasons. First, there is a wide variety of SLAM algorithms available, all using different methods to achieve the best quality of results.

Second, each sensor has its own specificity, and specifically all sensors present a certain degree of noise that needs to be mitigated by the algorithms. Finally, many algorithms are developed in a similar, well-lit environment, but few are tested in the field where the conditions are often less than ideal to maintain a good quality of tracking. We tested our prototype in Guatemala where archaeological excavations at a Classic Maya (250–1000 CE) site present a challenging environment due to the lack of good lighting and narrow tunnels without connecting loops.

Several studies actually compare and evaluate SLAM algorithms. Zennaro and colleagues (Zennaro *et al.*, 2015) compared two low-cost depth cameras in terms of noise, and Kraft’s team (Kraft *et al.*, 2017) evaluated these cameras on trajectory estimation. Huletski and colleagues (Huletski *et al.*, 2015) compared multiple SLAM algorithms on a standard data set. The common point of these studies is the use of an indoor, office-like environment. Though there has also been extensive evaluation of SLAM in outdoor environments for autonomous driving (Bresson *et al.*, 2017). Most of the evaluations are driven by the availability of data, typically through standard benchmarks with provided ground truth (e.g. (Sturm *et al.*, 2012; Geiger *et al.*, 2012)).

Other works study the usage of various handheld or robot-mounted sensors dedicated to the collection of 3D data in cultural heritage sites. Various studies present the evaluation of a handheld LiDAR scanning system in different field environments ((Zlot *et al.*, 2014; Farella, 2016; Di Filippo *et al.*, 2018; Nocerino *et al.*, 2017)). This system constitutes an improvement over larger terrestrial LiDARs in terms of usability, but remains a costly piece of equipment along with closed-source, commercial software. (Nocerino *et al.*, 2017) and (Masiero *et al.*, 2017) both evaluated a commercial backpack mapping system, providing a good mobile sensing solution, but also containing costly sensors. We present a study of multiple SLAM algorithms with multiple consumer-grade depth cameras, (see (Leingartner *et al.*, 2016) for a similar study), in a more constrained environment, and compare more recent and complex SLAM algorithms.

2.3.1 Loop closure. All SLAM algorithms are subject to drift. Drift is introduced by the accumulation of small errors during the reconstruction process and is usually the main reason of low accuracy in the resulting 3D models. For this reason, tracking is an important step for SLAM algorithms, along with loop closure. Loop closure enables a global optimization of the map when the user creates a loop in the scanning trajectory and revisits a place previously scanned. Our methodology takes loop closure into account and our results present the difference between performing a loop closure and not performing it.

3. Methodology overview

Our goal was to evaluate the digital reconstruction accuracy of SLAM algorithms, using multiple combinations of sensor hardware and application software. In order to achieve this goal, we chose data collection sites with conditions representative of the typical sites that our system is targeting (underground environments). We created 3D models of these sites by collecting data with 1) multiple SLAM-based scanning techniques and 2) LiDAR scans. The LiDAR scans were collected to provide a reference for the various 3D models. These scans were taken with ample overlap (around 1m between scans) to allow for a precise point-based registration, thus providing us with ground-truthed 3D models, which can be compared against the results of the real-time SLAM-based 3D scanning. Section 4 describes our data acquisition setup and methodology, along with the chosen sites, and Section 5 describes the data processing methods that we applied to compare the SLAM results to the LiDAR results.

4. Data acquisition

Our goal was to measure the performance of SLAM-based scanning techniques in a realistic scenario. We selected the hardware and software based on a set of requirements aimed at providing archaeologists with a usable system in the field. These requirements consist of designing a system that is lightweight, simple, handled by a single user and is easily

manipulated inside a tight and dark tunnel environment. In addition, the system must provide its own power, be self-contained (no external connection) and can be transported to and operated in a difficult environment (hot, humid, dusty, etc.). We also wanted to build a device based on readily available and low-cost components and materials and design software tools that are open source. Additionally, we wanted to test this device in field conditions, identical or similar to the conditions found in typical archaeological excavations.

4.1 Hardware setup

Our prototype consists of a backpack containing a laptop and an external tablet with a light panel and 3D sensors, as shown in [Plate 1](#). The laptop is the most expensive component, in order to run the real-time software needed for the scanning.

The first part of this prototype consists of a backpack modified with a solid frame that can handle a large laptop while keeping a good airflow to prevent overheating. In the same backpack, we keep several lithium polymer (LiPo) batteries to let the computer run for several hours and provide power to the sensors and lights. The second part of the device is a custom frame supporting a tablet facing the user, a light panel facing forward and a mount for a 3D sensor facing forward. The frame possesses handles for better manipulation in difficult areas and to give the user flexibility to scan at different angles. The tablet provides feedback from the laptop and can display the status of the scanning operation. The light panel is designed to provide enough brightness in a completely dark tunnel area, while limiting the creation of hard shadows that can cause issues with the SLAM algorithms. It consists of several strips of small LEDs that emit a bright white and diffuse light, sufficient to cover the field of view of our cameras. The sensor mount is designed to easily swap the sensor in use. Both parts of our hardware design are described in more details in our repository [\[1\]](#).

SLAM algorithms are optimized for a certain set of sensors. A very popular solution is to use RGB cameras as they are generally low-cost, light and small. However, while simple RGB cameras are good for tracking the position of a device, for example, a drone or robot, for 3D



Plate 1.
Scanning backpack
prototype. On top is the
handheld device,
containing a light
panel, a sensor mount
and a tablet for
visualization (photos
by Q. Gautier, 2019). On
the bottom left is the
backpack containing a
laptop and a metal
frame for better cooling
(photo by Q. Gautier,
2019). The bottom right
picture illustrates the
use of the backpack
inside the excavation

reconstruction purpose, we chose to use RGB-D cameras. RGB-D cameras are active or passive cameras that also generate 3D depth information alongside the regular RGB feed. To this purpose, they employ different technologies: structured light, time-of-flight or stereo vision. In our case, we tested RGB-D cameras covering these three types of technologies, all based on a near-infrared spectrum of light. The cameras that we use are summarized in Table 1. Two of these cameras can potentially operate in daylight, while the others only work in an indoor or underground environment only.

4.2 Software setup

For the 3D reconstruction in real time, we need software that can process the data from the RGB-D camera to create a 3D map of the environment, while providing visual feedback to the user. The visual feedback is important to track the status of the algorithms running, for example, whether the algorithm has lost track of its previous trajectory, in which case the user needs to come back to a previous position to restart the scanning. It is also useful to determine how much area has been scanned and take a decision on the next area to process.

The research literature is abundant on the topic of camera-based SLAM algorithms, but not all provide an end-to-end application that meets our requirements. We chose a few popular open-source systems that can easily be set up and used with the hardware described earlier. We chose the following algorithms:

- (1) RTAB-Map (Labbe and Michaud, 2011, 2013, 2014). RTAB-Map is a SLAM framework that provides a dense 3D reconstruction of the environment from different possible types of sensors. The framework presents many features and parameters to adjust the speed and accuracy of the scan. The total number of parameters is very high, and testing all the possible options is extremely time-consuming and not feasible on the field with limited time and resources. The software comes with a good default set of parameters that we use in our experiments. This algorithm possesses a loop-closure feature to improve accuracy.
- (2) ORB-SLAM (Mur-Artal and Tardós, 2017). ORB-SLAM is another popular open-source SLAM algorithm. The open-source version of ORB-SLAM 2 provides a real-time tracking based on a camera input with optional depth sensing. The application

Table 1.
RGB-D cameras used
in our experiments

	Technology	Notes
Microsoft Kinect v1	Structured light	
Microsoft Kinect v2	Time-of-flight	Global shutter
Intel RealSense ZR300	Active stereo	Works outdoors
Intel RealSense D435	Active stereo	Works outdoors

Table 2.
Summary of the
hardware/software
combinations used in
our SLAM experiments
at different sites

	ORB-SLAM	RTAB-map	InfiniTAM v2/v3
Kinect v1	Mud caves	Mud caves	Mud caves
	El Zotz	El Zotz	El Zotz
Kinect v2	Mud caves	Mud caves	Mud caves
	El Zotz	El Zotz	El Zotz
RealSense D435	Mud caves	Mud caves	Mud caves
RealSense ZR300	El Zotz	El Zotz	El Zotz
Two Kinect v1		Mud caves	
Note(s): Note that while we performed experiments with the RealSense ZR300, the results were too noisy to be analyzed properly			

has very little parameters and focuses on delivering an out-of-the-box functional tracking solution. The drawback of this algorithm is that it only saves a sparse map of the environment. In order to create a full 3D reconstruction, we need to apply some postprocessing to the output data. It does, however, perform a loop-closure optimization during the scanning procedure.

- (3) InfiniTAM v2 and InfiniTAM v3 (Kahler *et al.*, 2015; Kähler *et al.*, 2016). InfiniTAM is an application that focuses on RGB-D-based 3D reconstruction. It creates a dense map of the environment in real time by leveraging the GPU compute capabilities of the computer to accelerate the scanning. The difference between the two versions is mainly a loop-closure feature introduced in v3 that provides a global consistency to the map, but sometimes at the cost of stability in our experiments.

Our repository contains the scripts used to run these algorithms with their default parameters and process the output data in the case of ORB-SLAM.

4.3 Data collection sites

We evaluated our prototype inside active excavation tunnels at the Maya archaeological site of El Zotz, Guatemala (Plate 2). At this site, we collected data in the Pyramid of the Wooden Lintel (Structure M7-1, see Figure 1), an approximately 20-m tall structure that was extensively looted in the 1970s before archaeologists began extending the existing tunnels in 2012 to salvage information from the structure (Garrison *et al.*, 2012; Houston *et al.*, 2015). This excavation is large enough to create a good test environment and has been thoroughly scanned by LiDAR. Furthermore, the looted section of the tunnel can be considered representative of the types of contexts where our prototype scanner would be most useful.

We performed a second set of experiments in the Arroyo Tapiado Mud Caves system (Holler, 2019) located within the Anza-Borrego State Park near San Diego, California. These mud caves provided a controlled environment in a local context where we could run multiple similar scans without the constraints of an active excavation or the expense of working through developmental trial and error in an international field context where some resources are limited. While these natural caves are generally larger than a typical archaeological tunnel, they still present a similar setup in terms of scanning challenges. We particularly chose to operate our system in locations with sharp turns to emulate the shape of excavations.

4.4 Acquisition methodology

In all our experiments, we used the backpack system described in Section 4.1 and collected data with all the cameras available at the time of the experiment. For each of the experiments, we ran one of the SLAM algorithms described in Section 4.2, but we also recorded all the

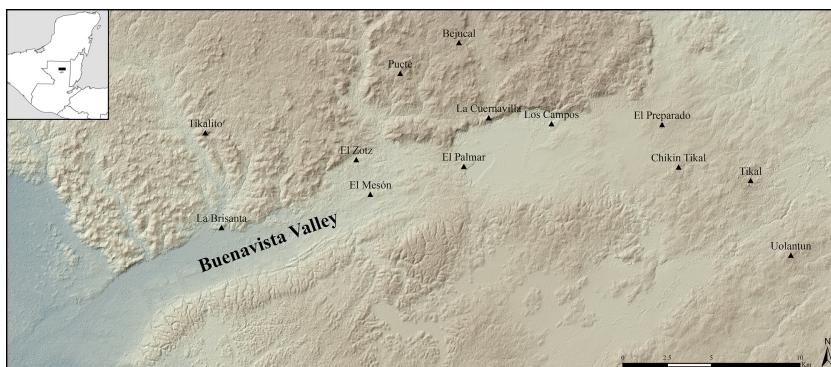
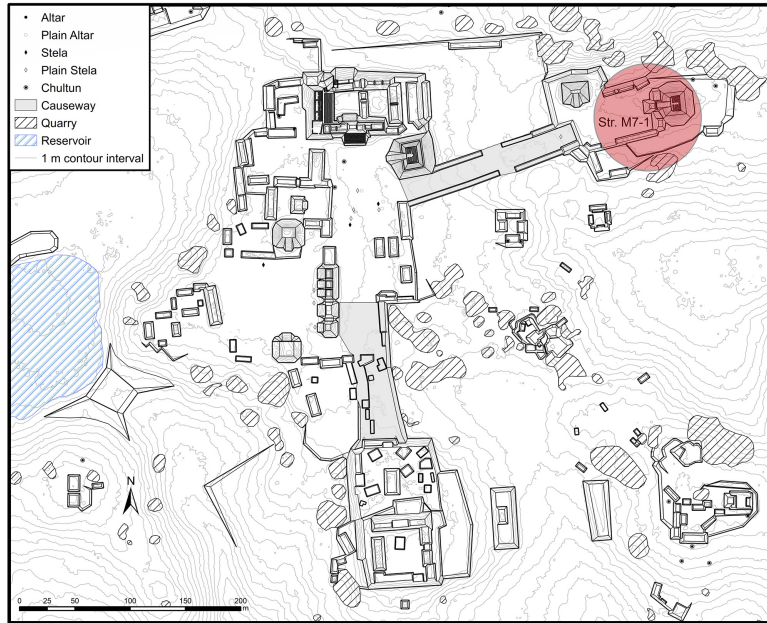


Plate 2.
Map of the area around
the Maya
archaeological site of
El Zotz, Guatemala
(map by
T. Garrison, 2019)

Figure 1.
The pyramid of the wooden lintel (structure M7-1), highlighted within the archaeological site of El Zotz (map by T. Garrison, 2019)



sensor data on the computer drive. This setup allowed us to replay the collected data through other algorithms in a very similar condition as the field experiment.

The first data collection at the El Zotz archaeological site happened in 2014, when we extensively scanned multiple excavations using a FARO Focus3D 120 LiDAR scanner (Garrison *et al.*, 2016). We use these LiDAR scans as ground truth reference models. In 2016, we started to run our SLAM experiments with various combinations of sensors and algorithms. The main results that we analyze in Section 6 were generated from large-scale scans covering about 45 m of the excavation tunnel in the Pyramid of the Wooden Lintel (Str. M7-1). We used the Kinect v1, Kinect v2 and ZR300 cameras and ran the ORB-SLAM algorithm. Most of these results start from the entrance of the tunnel, proceed to the end of the active excavation, turn around to reach the entrance again and turn around one last time to force the algorithm to perform a loop closure. This path goes through a series of steep architectural tiers from one of the earlier pyramidal substructures, including small ladders to transverse them. This renders the scans difficult in that area, but it is representative of the difficult field conditions that may be encountered during cultural heritage documentation.

Outside of the field seasons, from 2016 to 2018, we chose the Arroyo Tapiado Mud Caves site as a control site to refine our prototype. We selected a section of the *Carrey* cave, presenting several sharp turns. We scanned this whole section with multiple hardware/software setups, by using a very similar scanning walking pattern each time. We experimented with both Kinect cameras, the ZR300 and D435 cameras, and used ORB-SLAM and RTAB-Map. We also provided an experimental setup with two Kinect cameras simultaneously scanning the environment with the RTAB-Map framework. On each experiment, we also recorded the sensor data to possibly be replayed later through other algorithms.

5. Data processing

Our goal was to measure the quality of results of various combinations of sensors and algorithms. In this context, the quality of the result corresponds to the difference between

each sensor and algorithm combination as compared against a reference ground truth model, which is considered more accurate. In this case, we consider the LiDAR models to be sufficiently accurate for digital metric surveys, as they consist of dozens of 360 degrees scans carefully registered together (Garrison *et al.*, 2016). However, quantifying the difference between our experiment models and the ground truth reference can be challenging due to multiple factors affecting the output: sensor noise/accuracy, algorithm noise/accuracy, resolution of the output and so on. We chose to measure model quality with two different methods: one comparing the 3D models directly and the other comparing the trajectory path of our data collection sensors.

5.1 Measuring point cloud difference

The most obvious method to evaluate 3D scanning output against an available reference model is to overlay them together and observe the difference. However, a manual overlay can be difficult and introduce further error that is not related to the scanning method. We solve this problem by aligning the point clouds through rigid registration, solving for translation and rotation only (all our models are at scale). We use the iterative closest point (ICP) (Besl and McKay, 1992) method from the CloudCompare (Girardeau-Montaut, 2011) software to find the best alignment between each SLAM result and the ground truth LiDAR scan. From these aligned models, we can compute a point-to-point difference (Cignoni *et al.*, 1998) and use it to colorize the points. This colored model highlights the area with high difference compared to the reference. While the colored model can provide valuable information, it requires a visual inspection. We also summarize all the point-to-point differences by calculating the root mean squared error (RMSE), which provides the average residual error between points. The RMSE can be used as a metric for a quick comparison between results.

This method suffers from one major issue: all the types of error are averaged. First, the registration provides the best alignment that minimized the sum of squared errors over all points. This means that the location of areas with high error is not always accurate (larger error tends to appear on edges of the model). The statistics (min, max, RMSE, etc.) are still meaningful; however, they include multiple types of error, including sensor errors and software errors. This is why a visual inspection and a second quality metric are needed.

5.2 Measuring trajectory difference

When evaluating and comparing SLAM algorithms, most of the literature focuses on comparing the trajectories (Sturm *et al.*, 2012; Duarte *et al.*, 2016; Geiger *et al.*, 2012; Burgard *et al.*, 2009). A trajectory is defined as the position and rotation of the camera center at each scan (or at each *key* scan, depending on the algorithm) in the 3D space. The trajectory is generally considered as the most important output of the SLAM algorithm. If this output is accurate, then sensor 3D data can be joined together into a coherent model. Sensor noise can be reduced or averaged by using techniques that rely on an accurate estimate of the camera positioning. Several metrics exist to evaluate the quality of the trajectory with respect to a ground truth.

In our experiments, however, we do not possess a ground-truth trajectory. Instead, we recover a reference trajectory by using the LiDAR data. We incrementally register each camera scan to the reference LiDAR model and consider the resulting trajectory as a ground truth for comparison purposes. The process starts by a manual alignment of the first scan, which is followed by a fine-tuned alignment from an ICP algorithm. Each following scan is then aligned using ICP as well. We reduce the amount of sensor error by removing data past 3 m of the camera (as the error generally increases with the distance). We monitor most of the process visually and by reporting the RMSE of the registration at each step. The average RMSE of all the registered scans is generally on the order of 1 cm. We also reconstruct a 3D

model from the recovered ground-truth trajectory and the camera scans to verify that the output is similar to the LiDAR model.

From each pair of SLAM trajectory and recovered ground-truth trajectory, we provide statistics on the absolute trajectory error (ATE), which measures the difference between each SLAM pose and its corresponding reference pose, and the relative pose error (RPE) that quantifies the error on pose deltas (the error over a certain distance; in this case, per one meter).

6. Results

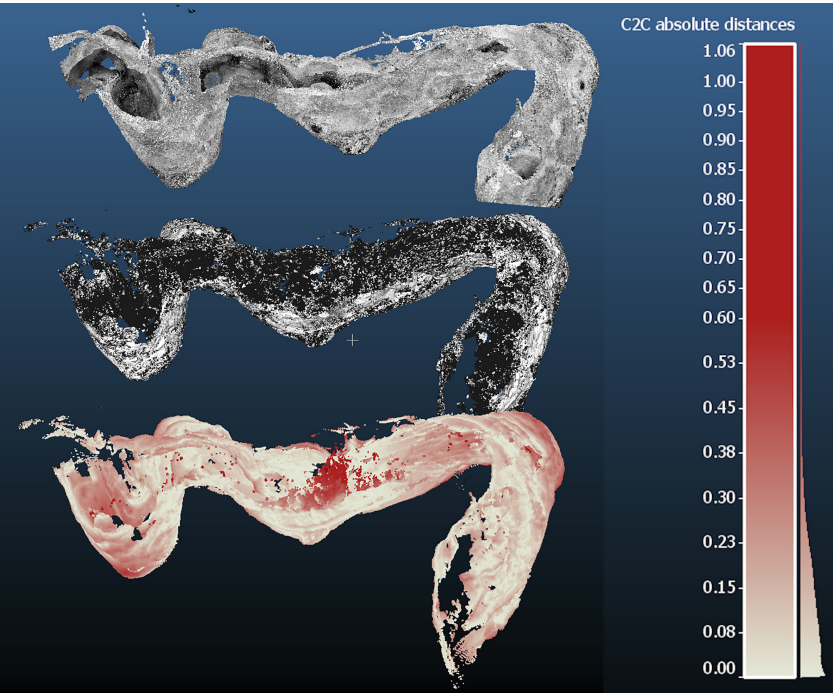
6.1 Arroyo Tapiado Mud Caves

We performed a controlled set of experiments at the Arroyo Tapiado Mud Caves site near the University of California, San Diego. This location was the best option for replicating the archaeological conditions in Guatemala in a local setting where all of the laboratory resources would be available during prototype development. First, we defined an area of approximately 50 m in length and scanned this area with the same FARO LiDAR system that was used to scan the El Zotz excavations. Then, we ran a series of scans using our SLAM scanning solution with three different sensors: Kinect v1, Kinect v2, RealSense D435 and two different algorithms: ORB-SLAM, RTAB-MAP. All the scans were performed by the same operator and followed a similar trajectory and scanning pattern. Additionally, we recorded all sensor data and replayed them through the InfiniTAM v2 algorithm afterward.

Here we present the results of comparing the output data from the SLAM algorithms to the LiDAR 3D model.

6.1.1 Cloud-to-cloud difference. Plate 3 shows an example of the cloud-to-cloud difference between a SLAM scan and a LiDAR scan. On top is the LiDAR scan of the mud caves site

Plate 3.
From top to bottom:
The LiDAR scan of the mud cave, the SLAM scan using RTAB-MAP and Kinect v2 and the same SLAM scan with point-to-point error with the LiDAR scan highlighted (darker shade of red = larger error). The scale indicates the error in meters, and the relative distribution is presented on the right side of the scale. Scans created by Q. Gautier



where we have focused our experiments. In the middle is the point cloud obtained from scanning the cave with the Kinect v2 and the RTAB-MAP algorithm. First, we aligned these models globally using ICP. Then, for each point in the SLAM point cloud, we computed the distance to the nearest neighbor in the reference LiDAR scan. The bottom model shows these point-to-point distances colored on a color scale from white (small distance) to red (large distance). In general, we notice more errors accumulated where the cave presents a turn. This is mainly due to two reasons. First, more drift is accumulated in areas with less visibility because the sensor has a more limited view of environment features. The other reason is that turns are places where sensor noise tends to accumulate. The depth quality of RGB-D cameras decreases with the distance; therefore, distant obstacles can accumulate the noise from multiple scans.

We summarize the cloud-to-cloud differences for each combination of algorithm and sensor by computing the RMSE over the entire point cloud. The results are presented in Figure 2. We can observe a similar error in multiple models using Kinect v1 and v2 on ORB-SLAM and RTAB-MAP. InfiniTAM v2 tends to give a larger error, as expected from an algorithm without loop closure detection. The RealSense camera tends to show a less consistent, larger error. This is due to the technology employed to generate depth data. Stereo matching can be noisier, especially as distance from the sensor increases. Intel provides several software filtering options for this camera, and we can expect better results when fine-tuning the process for this specific application. We also compare the results to a setup using two Kinect v1 cameras with RTAB-MAP. The model in this case presents a better RMSE as the software can utilize more sensor data to refine its trajectory calculation.

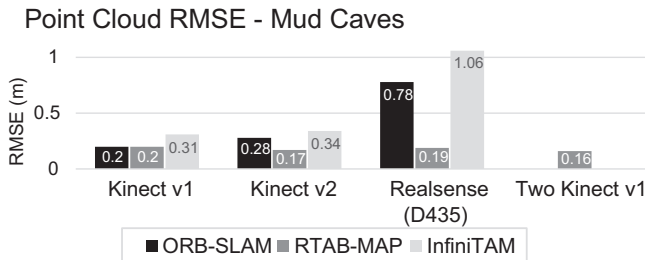


Figure 2. RMSE of the cloud-to-cloud difference between SLAM algorithm results and the LiDAR cloud, for the mud caves site

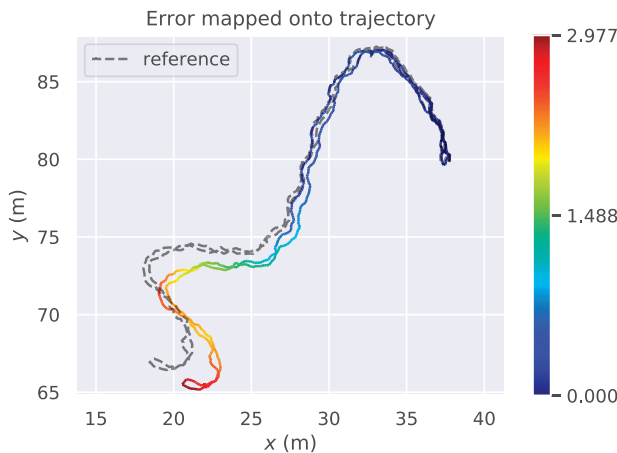


Figure 3. Trajectory comparison between RTAB-MAP with Kinect v1 and the recovered ground truth. The position error is plotted on the SLAM trajectory in meters

6.1.2 *Trajectory difference.* For each trajectory output from the SLAM algorithms, we recovered the corresponding ground-truth trajectory using the LiDAR data. We measured the ATE between the two outputs in two different ways: (1) by aligning the first pose of each trajectory – this gives an idea of how much error the algorithm accumulates with respect to the starting position, and (2) by aligning the whole trajectory to minimize the average error – this informs about the average error over the entire scan.

Figure 3 shows an example of the ATE between a trajectory from RTAB-MAP with Kinect v1 and the reference output, with respect to the starting position. We summarize the results in Table 3. The InfiniTAM results with the RealSense camera are not included, as they are too noisy to provide a good reconstruction. Generally, the results from the Kinect cameras are similar. Despite possessing a global shutter camera, the Kinect v2 does not show a consistent improvement; however, we believe that it is possible to improve the results by using a higher resolution than the default used in most algorithms. The RealSense camera creates more noise, and this translates to a less accurate trajectory. The InfiniTAM algorithm accumulates a lot of error over time as its tracking is based on depth data only. This is clearly reflected by a

Table 3. Absolute trajectory error (ATE) and relative position error (RPE) on different combinations of sensors and algorithms

	ORB-SLAM ATE (aligned)	RPE	RTAB-Map ATE (aligned)	RPE	InfiniTAM v2 ATE (aligned)	RPE
Kinect v1	0.64 (0.26)	0.04	0.98 (0.33)	0.05	2.10 (0.59)	0.05
Kinect v2	1.08 (0.36)	0.08	0.70 (0.31)	0.06	1.21 (0.56)	0.10
RealSense (D435)	5.24 (2.20)	0.25	1.49 (0.32)	0.11	–	–
Two Kinect v1	–	–	0.87 (0.25)	0.07	–	–

Note(s): ATE compares the trajectories aligned on the first pose only. ATE (aligned) compares trajectories globally aligned. RPE measures the relative error per 1m segments. All values are expressed in meters and averaged over the entire trajectory

Figure 4. Comparison of trajectories with and without loop closure using RTAB-Map with Kinect v1. The trajectory is plotted in meters for each axis (the z axis is aligned to gravity) against time, represented by numbered camera frames (index)



much worse ATE (large global error), but still keeping a similar RPE as other algorithms (good local consistency).

6.1.3 Loop closure analysis. While loop closure is a crucial step of most SLAM algorithms in order to increase the accuracy and reduce the drift, loops are very rare in tunnels and excavations. In this type of environment, one possible way to trigger a loop closure is to turn around, travel back to a previous point in the trajectory and turn around again. This last turn around allows the algorithm to recognize a visited place by matching similar visual features. We analyze the difference between closing the loop and not closing it. We run the sensor data through the ORB-SLAM algorithm, once keeping the loop closure and once stopping before backtracking. [Figure 4](#) shows the profiles of the trajectories with and without loop closure and the recovered ground truth. Both SLAM trajectories present a small drift over time with respect to the ground truth, but the drift is clearly higher without loop closure.

6.2 El Zotz archaeological site

We tested our scanning setup in a cultural heritage context at the archaeological site of El Zotz, Guatemala, and report our results here. Due to various constraints of the field, not all our experiments were executed in the exact same conditions. For this reason, we replayed the saved camera data through the different algorithms to increase the consistency. We collected data with the Kinect v1 and v2 cameras. Due to algorithm failure, the Kinect v2 data do not perform any loop closure. We also collected data with a RealSense ZR300 camera; however, the results are too noisy to be properly analyzed here.

We also compared the SLAM models to a LiDAR scan taken at a previous time, and therefore, the models can differ at certain specific areas since excavation conditions changed as archaeologists expanded new tunnels and backfilled other branches. The cloud-to-cloud difference is higher for this reason and can only be used as a comparison metric between SLAM algorithms and sensors. The trajectory reconstruction has been manually adjusted to ignore the areas with differences.

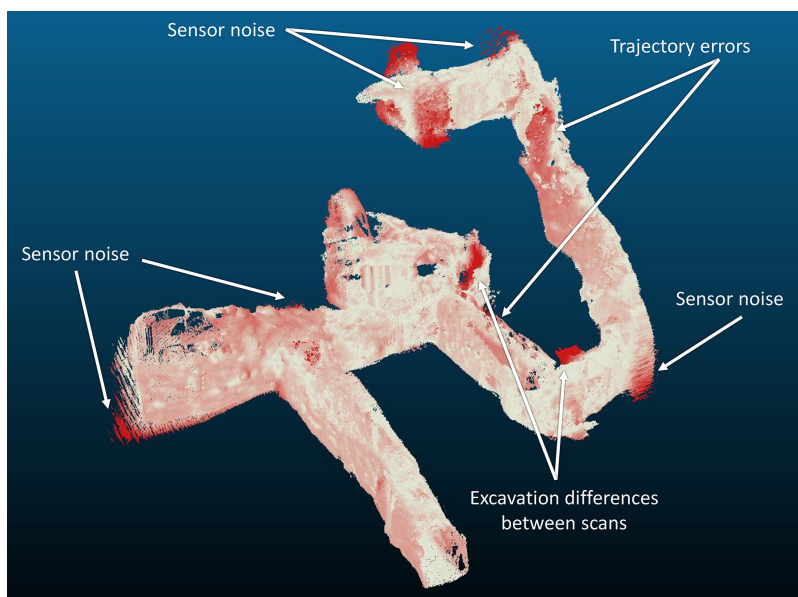


Plate 4.
Point-to-point difference between the point cloud from ORB-SLAM with Kinect v1 and the LiDAR point cloud. We show the SLAM cloud with the difference colored on each point. Darker red corresponds to a larger error, which happens in areas with concentrated sensor noise and areas where the two models differ in geometry. Scan created by Q. Gautier

6.2.1 Cloud-to-cloud difference. Plate 4 shows an error plot of a SLAM model against the LiDAR model on which we performed a visual inspection. We observed areas with high error and manually inspected the models to analyze the source. These areas mostly correspond to the differences between the scans. Otherwise we notice a similar trend with the mud caves models, where certain walls accumulate noise from sensor. This is more noticeable with ORB-SLAM as the sensor data are simply merged together without any kind of filtering.

Figure 5 shows the summary of RMSE for the two cameras. ORB-SLAM with Kinect v2 fails to reconstruct, but otherwise presents better results for Kinect v1. In this case, the Kinect v2 gives better results even though there is no loop closure. This environment is more difficult, and scans are very shaky, which can be handled better by the global shutter camera.

6.2.2 Trajectory difference. We recovered a reference trajectory based on the LiDAR data for our camera scans. Figure 6 shows the trajectory errors from ORB-SLAM and RTAB-MAP with Kinect v1. In this case, the RTAB-MAP algorithm has accumulated more drift than ORB-SLAM. We summarize the trajectory metrics in Table 4. The results are similar to what we measured in the mud caves environment, with InfiniTAM results being worse in global consistency compared to the image-based tracking algorithms.

7. Analysis and discussion

In this section, we discuss high-level trends and directions to take to maximize the accuracy of SLAM-based scanning, based on the results presented in Section 6 and our observations in the field.

7.1 Sensors

The RGB-D cameras that we used are broadly divided into two categories: active depth sensing (Kinect v1 and v2) and passive depth sensing (RealSense). Both our results and our observations confirm that active sensing produces the best results in terms of 3D scans. The difference in technology for active sensing (structured light with Kinect v1 and time-of-flight with Kinect v2) has a less significant impact on the result, although there tends to be a slight global improvement with the time-of-flight sensor. Stereo-based sensors produce depth maps that tend to be spatially and temporally noisy, which increases the drift in SLAM algorithms and increases the frequency of tracking loss during data collection. The ZR300 camera produced results too noisy to be analyzed in Section 6.2. However, it is important to note that the active sensors can only operate indoors and would not be suitable to scan the exteriors of sites. Additionally, the stereo sensors can be improved in software through the fine-tuning of filters, although this increases the complexity of the scanning system.

7.2 Algorithms

The algorithms can also be divided in two categories: feature-based tracking (ORB-SLAM and RTAB-MAP) and registration-based tracking (InfiniTAM). Feature-based algorithms mostly rely on image features and therefore, can more easily recover from tracking failure.

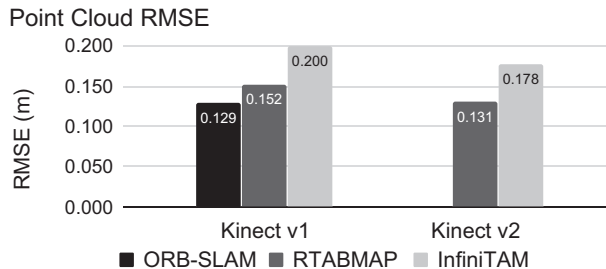


Figure 5. RMSE of the cloud-to-cloud difference between SLAM algorithm results and the LiDAR cloud, for the M7-1 excavation site

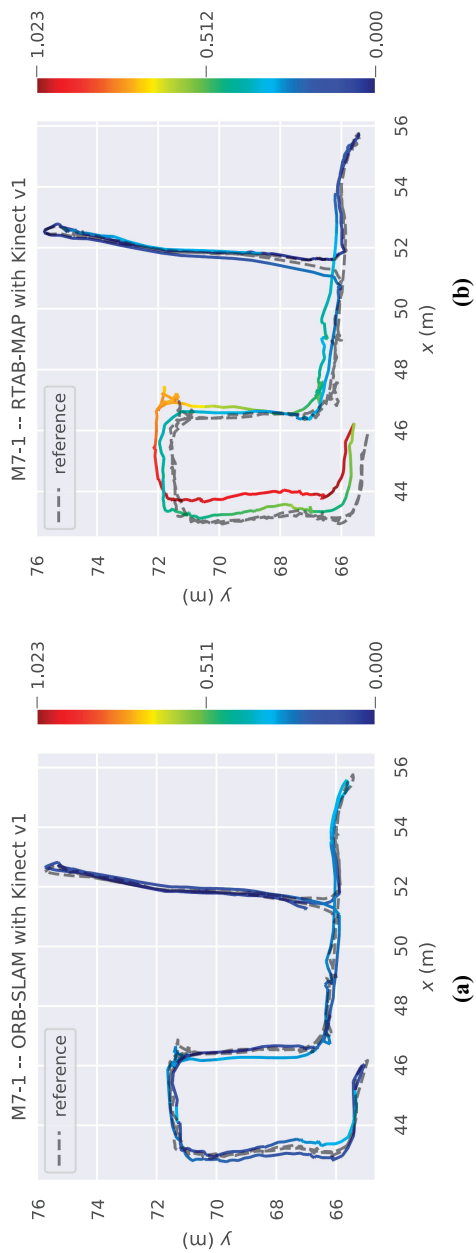


Figure 6. Top view of the trajectory comparison between SLAM results and the recovered ground truth. The position error is plotted on the SLAM trajectory in meters. (a and b) Trajectory error using ORB-SLAM with Kinect v1

However, these algorithms cannot recognize the same tunnel viewed from a different point of view (e.g. after turning around) and must rely on loop closure to create a consistent model in that case. Registration-based algorithms rely on depth data only and as such can recognize an environment based on its geometry only. These algorithms tend to be less stable in terms of tracking as they rely primarily on frame-to-frame alignment. InfiniTAM v3 is a good compromise, but our experiments failed to properly scan on long distances. While InfiniTAM v2 can be used for scanning without much tweaking, InfiniTAM v3 would require more tuning to be useable in these conditions. Feature-based algorithms worked better in our experiments, but they did require a particular attention to the lighting of the scene. Our scanning system is equipped with a light panel emitting diffuse light, which minimizes the negative effects on the SLAM algorithm; however, a strong, static light is recommended whenever possible.

7.3 Scanning procedure

In [Section 6](#), we highlight the importance of creating a loop closure, even when the environment presents no obvious such loop. This procedure applies to feature-based algorithms, or algorithms capable of handling loop closure, but is still useful in all cases simply to add new points of view of the scene and increase the number of details. However, due to the nature of these real-time algorithms, finding a loop closure can sometimes lead to a tracking failure, which forces the user to start over. Techniques for recovering the tracking in real time are improving and the next generation of algorithms might solve this problem.

There are also several good practices that we have noted during our experiments. First, each depth sensor has its own specifications in terms of sensing range and field of view. These limitations are very important for the user to keep in mind, as they generally make the difference between a failed scan and a good scan. Typically, there should always be a recognizably distinctive portion of the scene (containing visual features and geometric features) within the nominal range of the sensor (especially above the minimum range). Second, the scanning speed is important. While generally faster than other techniques, SLAM-based scanning is sensitive to fast motion, which introduces errors from: motion blur, lens distortion from rolling shutter cameras (minimized by using a global shutter camera such as Kinect v2), more distance between frames (minimized by using higher-frequency cameras such as the RealSense D435) or simply breaking the small motion assumption made by registration-based algorithms.

8. Conclusion

We have analyzed multiple hardware and software solutions to help with the documentation of cultural heritage within restricted spaces. Through multiple experiments in the field, we have established an accuracy value for different combinations of sensors and algorithms in these low-light conditions. In terms of software, image-based tracking algorithms tend to perform better for large-scale area scanning but may offer a sparser 3D model without postprocessing. Dense SLAM algorithms relying on depth tracking offer a good solution for small areas but do not scale well, unless implemented with loop-closure solutions. In terms of sensors, active depth sensing solutions targeted at indoor use perform better than stereo-based cameras in an underground environment. We have found that on average, the best performing combination of sensor and algorithm for our prototype scanning system is a

Table 4.
ATE and RPE in
meters, on the M7-1
excavation

	ORB-SLAM ATE (aligned)	RPE	RTAB-MAP ATE (aligned)	RPE	InfiniTAM v2 ATE (aligned)	RPE
Kinect v1	0.21 (0.19)	3.32*	0.47 (0.29)	0.06	2.27 (2.49)	0.08
Kinect v2	–	–	0.93 (0.17)	0.08	1.10 (0.47)	0.10

Note(s). *The trajectory has a large discontinuity causing a very large value for RPE

Kinect v2 with RTAB-MAP. This result can be generalized to new generations of cameras with similar technology (global shutter with a time-of-flight depth sensor), which, when running a feature-based SLAM algorithm, are likely to perform better.

SLAM algorithms provide a low-cost alternative to LiDAR and a fast alternative to photogrammetry methods. They are generally limited in the quality of results: in the presented use case of documenting archaeological underground excavations, the best system can produce about 20 cm of error in average after alignment and up to 50 cm in some cases. The proposed system is nevertheless a good solution for the rapid documentation of areas with extensive damage from looting where having an expensive instrument in difficult field conditions is impractical. Likewise, such a system has cultural heritage documentation applications in high-risk areas where archaeological sites are damaged or threatened in order to make rapid assessments of impacts and resource requirements.

Note

1. <https://github.com/UCSD-E4E/maya-archaeology>

References

- Besl, P.J. and McKay, N.D. (1992), "A method for registration of 3-d shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14 No. 2, pp. 239-256.
- Bresson, G., Alsayed, Z., Yu, L. and Glaser, S. (2017), "Simultaneous localization and mapping: a survey of current trends in autonomous driving", *IEEE Transactions on Intelligent Vehicles*, Vol. 2 No. 3, pp. 194-220.
- Burgard, W., Stachniss, C., Grisetti, G., Steder, B., Kümmerle, R., Dornhege, C., Ruhnke, M., Kleiner, A. and Tardós, J.D. (2009), "A comparison of slam algorithms based on a graph of relations", *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2089-2095.
- Canuto, M.A., Estrada-Belli, F., Garrison, T.G., Houston, S.D., Acuña, M.J., Kovác, M., Marken, D., Nondédéo, P., Auld-Thomas, L., Castanet, C., Chatelain, D., Chiriboga, C.R., Drápela, T., Lieskovský, T., Tokovinine, A., Velasquez, A., Fernández-Díaz, J.C. and Shrestha, R. (2018), "Ancient lowland maya complexity as revealed by airborne laser scanning of northern Guatemala", *Science*, Vol. 361 No. 6409, available at: <https://science.sciencemag.org/content/361/6409/eaau0137>.
- Chase, A.F., Chase, D.Z., Awe, J.J., Weishampel, J.F., Iannone, G., Moyes, H., Yaeger, J. and Kathryn Brown, M. (2014), "The use of lidar in understanding the ancient maya landscape: Caracol and Western Belize", *Advances in Archaeological Practice*, Vol. 2 No. 3, pp. 208-221.
- Cignoni, P., Rocchini, C. and Scopigno, R. (1998), "Metro: measuring error on simplified surfaces", *Computer Graphics Forum*, Vol. 17, Wiley Online Library, Oxford, UK, pp. 167-174.
- Cosme, G.M., Lorenzo, C.V. and Merlo, A. (2014), "La acrópolis de chilonché (Guatemala): crónica de las investigaciones de un patrimonio en riesgo en el área maya", *Restauración Arqueológica*, Vol. 22 No. 2, pp. 99-115.
- Di Filippo, A., Sánchez-Aparicio, L.J., Barba, S., Martín-Jiménez, J.A., Mora, R. and González Aguilera, D. (2018), "Use of a wearable mobile laser system in seamless indoor 3d mapping of a complex historical site", *Remote Sensing*, Vol. 10 No. 12, available at: <http://www.mdpi.com/2072-4292/10/12/1897>.
- Duarte, A.C., Zaffari, G.B., da Rosa, R.T.S., Longaray, L.M., Drews, P. and Botelho, S.S.C. (2016), "Towards comparison of underwater slam methods: an open dataset collection", *OCEANS 2016 MTS/IEEE Monterey*, pp. 1-5.
- Durrant-Whyte, H. and Bailey, T. (2006), "Simultaneous localization and mapping: part i", *IEEE Robotics and Automation Magazine*, Vol. 13 No. 2, pp. 99-110.
- Farella, E. (2016), "3d mapping of underground environments with a hand-held laser scanner", *Proc. SIFET annual conference*.

- Fassi, F., Fregonese, L., Ackermann, S. and De Troia, V. (2013), "Comparison between laser scanning and automated 3d modelling techniques to reconstruct complex and extensive cultural heritage areas", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-5/W1, pp. 73-80.
- Forster, C., Zhang, Z., Gassner, M., Werlberger, M. and Scaramuzza, D. (2017), "Svo: semidirect visual odometry for monocular and multicamera systems", *IEEE Transactions on Robotics*, Vol. 33 No. 2, pp. 249-265.
- Garrison, T.G., López, J.L.G. and de Carteret, A. (2012), "Investigaciones en la estructura m7-1 (operación 21) (pirámide del dintel de madera)", Proyecto Arqueológico El Zotz Informe No. 7, Temporada 2012 pp. 59-97, available at: <http://www.mesoweb.com/zotz/El-Zotz-2012.pdf>.
- Garrison, T.G., Richmond, D., Naughton, P., Lo, E., Trinh, S., Barnes, Z., Lin, A., Schurgers, C., Kastner, R. and Newman, S.E. (2016), "Tunnel vision: documenting excavations in three dimensions with lidar technology", *Advances in Archaeological Practice*, Vol. 4 No. 2, pp. 192-204.
- Geiger, A., Lenz, P. and Urtasun, R. (2012), "Are we ready for autonomous driving? the kitti vision benchmark suite", *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354-3361.
- Girardeau-Montaut, D. (2011), "Cloudcompare (version 2.10) [GPL software]. (2019)", available at: <http://www.cloudcompare.org/>.
- Grant, D., Bethel, J. and Crawford, M. (2012), "Point-to-plane registration of terrestrial laser scans", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 72, pp. 16-26.
- Holler, C. (2019), "Chapter 101 – pseudokarst", in White, W.B., Culver, D.C. and Pipan, T. (Eds), *Encyclopedia of Caves*, 3rd ed., Academic Press, London, UK, pp. 836-849, available at: <http://www.sciencedirect.com/science/article/pii/B97801281412423001011>.
- Houston, S.D., Newman, S., Román, E. and Garrison, T. (2015), "A tomb and its setting", *Temple of the Night Sun: A Royal Tomb at El Diablo, Guatemala*, Precolumbia Mesoweb Press, San Francisco, CL, pp. 12-29.
- Huletski, A., Kartashov, D. and Krinkin, K. (2015), "Evaluation of the modern visual slam methods", *2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT)*, pp. 19-25.
- Ibragimov, I.Z. and Afanasyev, I.M. (2017), "Comparison of ros-based visual slam methods in homogeneous indoor environment", *2017 14th Workshop on Positioning, Navigation and Communications (WPNC)*, pp. 1-6.
- Kahler, O., Adrian Prisacariu, V., Yuheng Ren, C., Sun, X., Torr, P. and Murray, D. (2015), "Very high frame rate volumetric integration of depth images on mobile devices", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 21 No. 11, pp. 1241-1250, doi: [10.1109/TVCG.2015.2459891](https://doi.org/10.1109/TVCG.2015.2459891).
- Kähler, O., Prisacariu, V.A. and Murray, D.W. (2016), "Real-time large-scale dense 3d reconstruction with loop closure", in Leibe, B., Matas, J., Sebe, N. and Welling, M. (Eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, pp. 500- 516.
- Kohlbrecher, S., Von Stryk, O., Meyer, J. and Klingauf, U. (2011), "A flexible and scalable slam system with full 3d motion estimation", *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, IEEE, pp. 155-160.
- Kraft, M., Nowicki, M., Schmidt, A., Fularz, M. and Skrzypczynski, P. (2017), "Toward evaluation of visual navigation algorithms on rgb-d data from the first- and secondgeneration kinect", *Machine Vision and Applications*, Vol. 28 No. 1, pp. 61-74, doi: [10.1007/s00138-016-0802-6](https://doi.org/10.1007/s00138-016-0802-6).
- Labbé, M. and Michaud, F. (2011), "Memory management for real-time appearance-based loop closure detection", *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1271-1276.
- Labbé, M. and Michaud, F. (2013), "Appearance-based loop closure detection for online large-scale and long-term operation", *IEEE Transactions on Robotics*, Vol. 29 No. 3, pp. 734-745.

-
- Labbé, M. and Michaud, F. (2014), "Online global loop closure detection for large-scale multi-session graph-based slam", *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2661-2666.
- Leingartner, M., Maurer, J., Ferrein, A. and Steinbauer, G. (2016), "Evaluation of sensors and mapping approaches for disasters in tunnels", *Journal of Field Robotics*, Vol. 33 No. 8, pp. 1037-1057, doi: [10.1002/rob.21611](https://doi.org/10.1002/rob.21611).
- Li, R., Liu, J., Zhang, L. and Hang, Y. (2014), "Lidar/mems imu integrated navigation (slam) method for a small uav in indoor environments", *2014 DGON Inertial Sensors and Systems (ISS)*, pp. 1-15.
- Luhmann, T., Robson, S., Kyle, S. and Boehm, J. (2013), *Close-range Photogrammetry and 3D Imaging*, Walter de Gruyter, Berlin.
- Masiero, A., Fissore, F., Guarnieri, A., Piragnolo, M. and Vettore, A. (2017), "Comparison of low cost photogrammetric survey with tls and Leica Pegasus backpack 3d modelss", *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W8, pp. 147-153.
- Meyer, J., Schnitzspan, P., Kohlbrecher, S., Petersen, K., Andriluka, M., Schwahn, O., Klingauf, U., Roth, S., Schiele, B. and von Stryk, O. (2011), "A semantic world model for urban search and rescue based on heterogeneous sensors", in Ruiz-del Solar, J., Chown, E. and Plöger, P.G. (Eds), *RoboCup 2010: Robot Soccer World Cup XIV*, Springer, Berlin, Heidelberg, pp. 180-193.
- Mur-Artal, R. and Tardós, J.D. (2017), "Orb-slam2: an open-source slam system for monocular, stereo, and rgb-d cameras", *IEEE Transactions on Robotics*, Vol. 33 No. 5, pp. 1255-1262.
- Nocerino, E., Menna, F., Remondino, F., Toschi, I. and Rodríguez-Gonzálvez, P. (2017), "Investigation of indoor and outdoor performance of two portable mobile mapping systems", in Remondino, F. and Shortis, M.R. (Eds), *Videometrics, Range Imaging, and Applications XIV*, Vol. 10332, SPIE, pp. 10332 - 10332 - 15, doi: [10.1117/12.2270761](https://doi.org/10.1117/12.2270761).
- Smith, R.C. and Cheeseman, P. (1986), "On the representation and estimation of spatial uncertainty", *The International Journal of Robotics Research*, Vol. 5 No. 4, pp. 56-68, doi: [10.1177/027836498600500404](https://doi.org/10.1177/027836498600500404).
- Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D. (2012), "A benchmark for the evaluation of rgb-d slam systems", *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 573-580.
- Yastikli, N. (2007), "Documentation of cultural heritage using digital photogrammetry and laser scanning", *Journal of Cultural Heritage*, Vol. 8 No. 4, pp. 423-427, available at: <http://www.sciencedirect.com/science/article/pii/S1296207407001082>.
- Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S. and Menegatti, E. (2015), "Performance evaluation of the 1st and 2nd generation kinect for multimedia applications", *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6.
- Zlot, R., Bosse, M., Greenop, K., Jarzab, Z., Juckes, E. and Roberts, J. (2014), "Efficiently capturing large, complex cultural heritage sites with a handheld mobile 3d laser mapping system", *Journal of Cultural Heritage*, Vol. 15 No. 6, pp. 670-678, available at: <http://www.sciencedirect.com/science/article/pii/S1296207413002185>.

Corresponding author

Quentin Kevin Gautier can be contacted at: qkgautier@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com