Robots for Exploration, Digital Preservation and Visualization of Archeological Sites

J. Serafin*, M. Di Cicco*, T. M. Bonanni*,
G. Grisetti*, L. Iocchi* and D. Nardi*
Sapienza University of Rome

C. Stachniss*
University of Bonn

V. A. Ziparo*
Algorithmica S.r.l.

Abstract—**Monitoring and conservation of archaeological sites are important activities necessary to prevent damage or to perform restoration on cultural heritage. Standard techniques, like mapping and digitizing, are typically used to document the status of such sites. While these task are normally accomplished manually by humans, this is not possible when dealing with hard-to-access areas. For example, due to the possibility of structural collapses, underground tunnels like catacombs are considered highly unstable environments. Moreover, they are full of radioactive gas radon that limits the presence of people only for few minutes. The progress recently made in the artificial intelligence and robotics field opened new possibilities for mobile robots to be used in locations where humans are not allowed to enter. The ROVINA project aims at developing autonomous mobile robots to make faster, cheaper and safer the monitoring of archaeological sites. ROVINA will be evaluated on the catacombs of Priscilla (in Rome) and S. Gennaro (in Naples).**

I. INTRODUCTION

The protection and the preservation of cultural heritage is of fundamental importance. Europe has a big wealth of archaeological sites and, according to UNESCO[1], Italy is the country that counts the largest number in Europe, and more in general in the world. In this perspective, mapping and digitizing are standard techniques used for monitoring and conserving such sites. Unfortunately these tasks are very challenging and expensive. Indeed, systems used for digitizing cultural heritage are typically built around static 3D laser scanning technology, that needs to be brought periodically into the field by teams of experts. Additionally, these surveys also require a lot of manual work making the whole task slow and prone to errors.

* While this is acceptable for accessible sites, many other archaeological locations are dangerous and they prevents the digitization due to the serious risks posed on the surveyors working in it. As an example, consider the case of catacombs. Catacombs are underground tunnels, located in different cities in Europe, that can extend for several kilometers and multiple floors. Due to the high risks in exploring such environments, most of these tunnel networks are still partially unexplored. Indeed, catacombs are really dangerous for their structural instability and the high probability of collapses. Additionally, since in most of them there is no ventilation, they can contain high concentrations of radioactive radon gas. This limits the presence of humans to 15-30 minutes at most.

Fig. 1. Pictures of the Priscilla catacombs in Rome.

Recently, in the field of artificial intelligence and robotics, there has been a deep progress in the development of autonomous robots that offer various services to their users. For this reason, robots are one of the possible answers to the problem of preserving cultural heritage in areas where humans would be in serious danger. The task of autonomous exploration on such unstructured environments, however, is still challenging. For example, in the case of three-dimensional underground sites such as catacombs, multiple issues arise. This kind of areas pose enormous difficulties to autonomous robots because of the problematic perception (no/poor lighting, unpredictable situations, difficult scene interpretation, complex traversability analysis, etc...). Moreover, the communication capabilities are limited or missing, restricting the continuous supervision and teleoperation by human users.

In this paper we describe the ROVINA project[2]. ROVINA is a three-year and a half research project that is co-funded by the European Commission in the frame of the FP7 program. The University of Bonn (GE), University Freiburg (GE), University Aachen (GE), University Leuven (BE), Sapienza University of Rome (IT), Algorithmica S.r.l. (IT) and the International

---

[1] http://whc.unesco.org/en/list/stat

* All authors equally contributed to this paper.

[2] This work has partly been supported by the European Commission under contract number FP7-600890-ROVINA. http://www.rovina-project.eu/

Council of Monuments and Sites (IT) compose the ROVINA consortium. The project aims at using autonomous mobile robots to make faster, cheaper and safer the surveying of archaeological sites. The main goals of ROVINA are:

- to provide novel technology that supports the preservation of cultural heritage by allowing the acquisition of digital models in hard-to-access environments;
- to extend the technology of autonomous navigation for robots designed to explore unknown underground environments such as caves and catacombs;
- to develop novel techniques to construct large 3D textured models of these poorly structured environments;
- to offer a cost-effective support for monitoring such sites and to enable comparative analysis that will allow to devise better preservation plans.

In Italy there are approximately 100 catacombs that are distributed across 29 cities. As one may predict, most of them are in Rome that has 48 catacombs, but there are also 11 in Naples. These kind of archaeological sites are a very interesting case of study and, for this reason, the ROVINA project will be evaluated in the catacombs of Priscilla (in Rome) and S. Gennaro (in Naples). Consider that the catacombs of Priscilla (see Fig. 1) extend for about 13 km over 3 floors.

The remainder of this paper is organized as follows: in Sec. II we give an overview of the state-of-the-art for cultural heritage preservation, in Sec. III we present the achievements of the ROVINA project and Sec. IV concludes the paper.

II. STATE OF THE ART

As stated in the previous section, the ROVINA project aims at building robots that are able to perform automatically the existing practices used in cultural heritage. In this section we give an overview of the main activities involved in the preservation and documentation of archaeological sites.

One of the main activities connected to a cultural heritage survey is measuring, which can be divided in two different categories: “direct” and “indirect”. The first one is related to manual measurements taken directly on the object by the operators. To this end, traditional tools like the plumb line, Charles metric, water systems and others instruments can be used. Usually, in these cases, operators increase the robustness of the measurements by either building polygons (if possible with the help of topography) or, alternatively, by making manual triangulation. Indirect measurements, instead, are mainly based on laser scanning and/or image analysis (by means of classic photogrammetry approaches), since they both allow morphometric inspections. Laser based systems, but more in general equipments based on time-of-flight measuring, are usually used when dealing with large scale sites or large artifacts. Even if expensive, these sensors are capable to collect large amount of data with high precision. As an example, the Leica Cyrax scanner has a range of 2m-150m, with a resolution that is higher than 1cm, and a full 3D scan (with millions of points) needs a few minutes to be acquired. Clearly, the requirements on the precision of the measurements may vary depending on the application. For documentation purposes, a range resolution of 1cm is usually enough. Image based systems are often used when an higher accuracy is required, like in the case of tasks related to diagnosis. If a sufficient number of images close to the surfaces can be acquired, this kind of applications have demonstrated to provide precision comparable with those of Lidar scanners. The main advantage is that texture information (color) is available and perfectly aligned with 3D data, hence no calibration between the two is needed. Image-based systems typically rely on high-resolution commercial cameras such as the 10MPixel Nikon D200. The acquired images are processed by using commercial photogrammetry software, multi-view stereo approaches or, more recently, self-calibrating structure-from-motion systems as ROVINA does ([1][2]).

An other core activity in surveys is the documentation. Documentation is fundamental since it is necessary to produce digital archives of the observing site. In the field of cultural heritage, this kind of activities are usually performed by public entities, as superintendencies and ministries in Italy. A digital archive can be represented by geometrical 3D models that could contain, or not, texture information. If the survey has diagnosis or measurements purposes, the 3D models should be very precise and with a high resolution. Conversely, if the goal is to disseminate cultural heritage to a broad audiences, than 3D models can have a lower resolution but they should be visually appealing. Under this point of view, tools such as virtual museums are of utmost importance.

Classification activities are usually connected to documentation tasks. They are related to the categorization of the elements of an archaeological site into taxonomies or ontologies with different degrees of complexity. For example, in a cultural heritage area, it could be of interest to classify architectural components and objects based on a number of parameters such as materials used, period of construction, preservation status and so on. Usually this classification is performed by humans that manually tag the items and parts of the environment. If the data is collected on a geographical scale, than models are archived into so called Geographical Information Systems (GIS). In this case, the archive can be interrogated to retrieve information on both a geographical and qualitative level. For example, one may look for all the buildings dated before 1000 B.C. in Italy or all the wooden vessels made in England.

Finally, diagnostic is an activity whose goal is to gather data about the preservation status of the surveyed areas, this is necessary to prevent further damage or perform restoration of the artifacts on the site. The result of this activity are the so called Table of Deterioration and Table of Materials. The first one shows on a map the locations with possible damages such as molds and cracks, while the other maps the site based on the materials by which each area or object is composed. At a European level, the standards required for the Table of Deterioration are established by the European Committee for standardization (CEN) (i.e., WS Construction, WS Measurement and WS Material). In Italy the standards are defined by the UNI-NorMal commission that is composed by the Ente Nazionale Italiano di Unificazione (UNI) and

Fig. 2. The ROVINA robot in the catacombs of Priscilla in Rome.

the Commissione NORmalizzazione MAteriali Lapidei (NorMaL).

III. ROVINA

In this section we give a detailed explanation of what have been developed and achieved so far in the ROVINA project. More specifically, in Sec. III-A we introduce the robotic platform we use in the field; 3D data processing, comprising world reconstruction and semantic analysis, is presented in Sec. III-B and Sec. III-C; the details of the navigation system are shown in Sec. III-D; finally Sec. III-E and Sec. III-F conclude this description of the ROVINA project showing respectively the user interface available during a mission, and the software that can be used to query digital archives.

A. The ROVINA Robot

At the core of the ROVINA project there are three robot prototypes (see Fig. 2) that we use to explore and collect data in archaeological sites, in our case catacombs. The initial analysis of the catacombs highlighted a harsh environments resembling the one encountered by rescue robotics. In particular, the terrain resulted to be rough with steep slopes, holes and challenging obstacles like drifts, debris or even fragile artifacts. For all these reasons, the base platform chosen for ROVINA was the Mesa Element. Thanks to the caterpillars, it can move easily in underground environments like catacombs or caves and, at the same time, it is big enough to provide housing and power to all the equipments required for the digitization of the sites.

To create a digital archive, the robot has to perform multiple tasks as 3D world reconstruction, navigation and semantic 3D analysis. Starting from the Mesa Element base, the robot was modified to make it suitable to carry all sensors (and the energy to power them) needed for these tasks. At current time, the platform provides an autonomy of ∼6 hours.

Due to the catacombs air conditions, the robot must be able to operate without damaging on-board devices, this means that it must be robust to water dripping from the ceiling, some dust and high levels of humidity (∼95%). To overcome these issues, temperature and humidity sensors were mounted on the platform so that they are monitored continuously. Moreover, a battery status sensor has been used to keep track of the remaining power of the robot. This is fundamental because, in case the robot is running low on battery, the autonomous navigation system has to interrupt the mission and bring the platform back to the base.

For mapping and navigation purposes, we decided to mount three RGB-D cameras (Asus Xtion Pro Live) in front of the robot. This kind of sensor generates depth and RGB images at high frame rates from which we extract, in the form of point clouds, 3D local models of the environment. These models are crucial for our processing pipeline since they are used to build a map of the environment that the robot uses to navigate. Moreover, always in terms of navigation, we exploit the point clouds to compute a traversability map that is useful to discriminate when the robot can, or not, traverse a certain patch of terrain. To increase the robustness of the mapping system, the platform is equipped with an Inertial Measurement Unit (XSens Mti-300-AHRS IMU) that helps providing accurate measures of the movements of the robot.

The 3D dense reconstruction of the environment is performed by using a self-calibrating structure-from-motion algorithm ([1][2]), which is feeded by an arc of 7 RGB cameras and lights placed on top of the ROVINA robot. Each camera is able to generate images at 2MPix and it mounts lenses allowing a 60 degree opening angle. All images are streamed via ethernet cables, that are connected to a central switch communicating with the on-board computers.

Finally, a single ray 3D laser scanner (Ocular Robotics RE05) is used for mapping or scanning purposes during a mission. Other features include a maximum range of 30 meters, a 30kHz sample rate and 0.01 degree angular resolution. This laser is able to generates point clouds that cover a range of 360° in the azimuth and 70° in the elevation. Moreover, it is possible to concentrate the scanning on a specific part of the sorrounding environment generating an high density point cloud of the region of interest. These characteristics make this sensor suitable for different goals ranging from mapping, in particular in large scale environments, to documentation or diagnosis purposes.

B. World Modeling and 3D Reconstruction

The mapping process is fundamental in order for the robot to be able to autonomously navigate the site, and to give the humans operators a better understanding of the environment. To this end, we extended state of the art approaches for 3D Simultaneous Localization and Mapping (SLAM). One of the main issues of this task is that, because of the high complexity, the computational load is huge and thus, hard to be performed online. Despite this, we wrote a Graph SLAM variant based on condensed measurements that can run online during the mission [3]. The overall SLAM method is composed by an optimization back-end [4] and a set of heterogeneous front-ends that exploits the available sensors on the robot (i.e., RGB-D cameras, the 3D laser scanner and the array of RGB cameras).

Our main front-end is based on a novel camera tracking system ([5]), which relies on depth images acquired with RGB-D cameras. While the robot is moving, the measurements

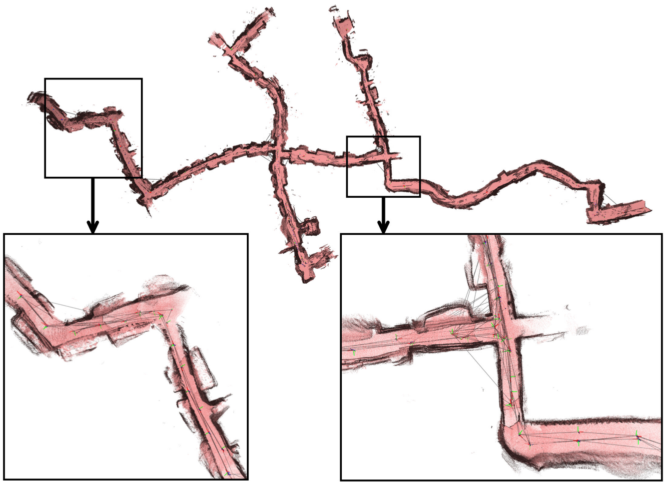


Fig. 3. Top view with particulars of the 3D map constructed on-line by the ROVINA robot in the catacombs of Priscilla in Rome.

generated by these sensors are aligned and accumulated together to form aggregated point clouds denoted as local maps. We start a new local map each time we detect an inconsistency in the point cloud alignment, or when the robot has moved of a certain amount (both in terms of translation or rotation). By comparing the internal trajectories of the local maps, we search for possible loop closure candidates. A loop closure candidate between two local maps is validated and accepted by performing a geometrical consistency check of the resulting alignment. Fig. 3 shows an example map of the catacombs of Priscilla in Rome.

The multicue front-end part of this task is heavily integrated with the RGB camera arc. This arc is designed to perform omni-directional captures, and each camera view overlaps for about half of the image size with the next viewpoint. Taking into account the relative positions of the cameras, and the motion of the cameras orthogonal to their viewpoints, the associated point clouds are generated. We developed a novel approach ([6]) that, instead of performing SfM (Structure from Motion) on a traditional image sequence, it analyzes the incoming data on a grid, where one axis corresponds to the several neighboring/overlapping cameras, and another axis to the changes in time. Each image on this grid is evaluated for finding matches in a local neighborhood, which are used further in the process for camera bundling. Moreover, whereas traditional SfM assumes uniform lighting conditions, the mobile lighting of the robot causes the brightness of the images to change in each point in the different viewpoints. Therefore the feature extraction and matching strategies have to be adjusted to avoid matching problems. The correspondence search across the images is implemented by using SURF-like features, while the local intensity profile on the image regions has been gradient-corrected to guarantee a better resemblance across the imagery.

As said before, due to the rough terrain, holes and obstacles, catacombs are really challenging environments. Under these conditions, a SLAM algorithm can easily fail resulting for the robot to be lost during the exploration. To minimize the possibility for the ROVINA platform to fall in these kind of situations, we developed multiple methods to increase the robustness of the whole system. More specifically, we improved the outliers rejection ([7][8]), the map consistency checking ([9]), and we developed novel systems ([10][11]) to automatically calibrate the sensors.

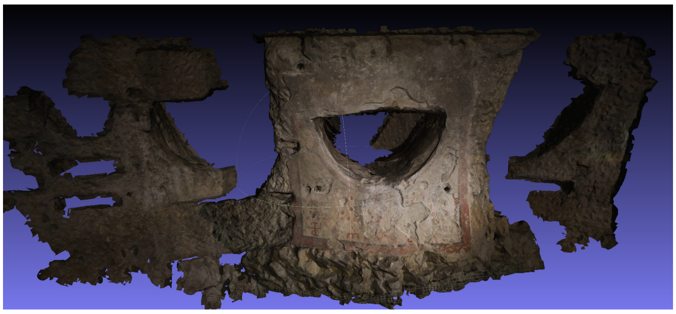


Fig. 4. Example of dense 3D reconstruction using the Structure From Motion pipeline.

The camera arc is also the fundamental tool used for 3D dense reconstruction. As in the case of the SLAM task, the mobile lighting causes the brightness and color of all visible surfaces to be affected by their positions, distances, and angles to the active light sources on the robot. It is important to compensate for those effects in order to create a reconstruction that is visually pleasing and photometrically reliable, not in the least for further study, analysis, and classification. We introduced a novel methodology to jointly exploit photometric stereo (PS) and normals characteristics to refine the shape and determine a reflection model close to a real Bidirectional Reflectance Distribution Function (BRDF). Starting from a low-resolution mesh generated with the SfM approach, we apply photometric stereo by sampling the BRDF space to detect base materials and jointly refine the geometry and reflectance model. We validated the approach by reconstructing several challenging objects consisting of multiple non-Lambertian materials, and this both geometrically and photometrically. An example is shown in Fig. 4.

C. Semantic 3D Analysis

The goal of this task is to develop methods for interpreting the 2D (image) and 3D (point cloud) sensor data, and for classifying it into semantically meaningful regions. Since ROVINA test case are archaeological sites, we concentrate on a range of different semantic aspects like geometric properties, functional classes, shape and material characteristics, and semantic labels relevant for archaeological study and preservation.

In order to solve this problem we introduced a new method ([12]) which is integrated with the underlying mapping system. A subset of the images from a camera stream is processed by the semantic segmentation module, resulting in automatically generated pixel-wise semantic labels. When a local map from

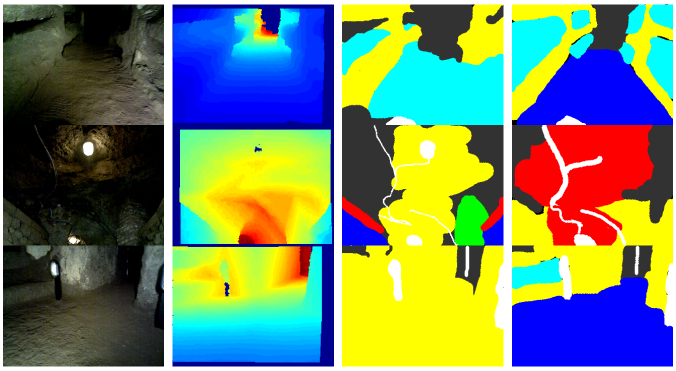


Fig. 5. Samples from the Priscilla catacombs annotations. From left to right, color image, depth image, material annotation and object class annotation are shown. In both annotated images, gray represents the *unknown* class. The color legend for the labels is shown in Table I.

the SLAM module is completed, the semantic information is integrated for each point in this map. This yields a spatial and temporal integration of semantic information, giving a consistent semantic labeling for each local map. In order to adapt and improve the system performances on data coming from the catacombs, we manually annotated part of the images acquired with Xtion RGB-D sensors in the catacombs of Priscilla in Rome. Based on this information, we trained and validated the semantic segmentation process.

Analyzing the data highlighted some of the challenges of our target scenario. Even for a human annotator, it is not possible to identify roughly 30% of the image content, since some regions are simply too dark or too far away to be recognized. But also for the other classes it was often hard to make out the exact class, as the intra-class similarity is very high. Examples of these annotations and an overview over the label distribution can be seen in Fig. 5 and Table I. In order to deal with those problems, we mark uncertain regions as unknown in the ground-truth annotation. We try to detect such regions automatically in the RGB-D input data and use this knowledge in order to not force our classifiers to make uncertain decisions. To deal with the problem of similar classes, we model the class confusion probabilities and integrate them into the semantic segmentation process.

The semantic annotation task is heavily related to the 3D dense reconstruction. The reconstruction/texturing pipeline can only make weak assumptions on geometry or material composition and starts from the conjecture that a material can be typically described as a linear combination of base materials for which a (close-to-) BRDF representation can be computed. By observing the reflectances of each point in each of the camera viewpoints, an initial BRDF representation is built for each point. Using ACLS, a generalized form of Non-Negative Matrix Factorization (NNMF), the observed BRDFs are clustered into a limited number of base materials. In order to determine the total number of materials, we run ACLS with an increasing number of base materials k and we stop when the newly generated BRDFs are too close to the ones already created in the previous step. We assign each point to the base material that gave the higher mixing weight from ACLS. The process has been further elaborated by introducing a mix PCA (Principal Component analysis) with NNMF, since pure ACLS does not necessarily give a optimal basis from a visual point of view. The process also takes into account the resolution and scale of the surface, since these affect the normal descriptions, and thus the BRDF sampling. Further, as soon as the weighting is fixed, a virtual albedo color texture can be generated which compensate the non-uniform lighting from the recording (i.e. the camera arc in particular). Another aspect of semantic segmentation is the final representation of the 3D model. The possible amount of data coming out of the SfM based pipeline is by definition too large to hold in a single geometry. Therefore, we provided a multi-resolution representation for the detail, and a sectioning or partitioning scheme to have local representations (i.e. chunks) that seamlessly have to fit together. These chunks are to be fed into the visualization module. The current approach involves a Poisson-based reconstruction scheme at different depths around locally neighboring camera positions.

D. Navigation

In order to allow autonomous navigation, a robot has to be able to understand when it can traverse, or not, a certain area. To this end, we developed an accurate, fast to compute, and comparably easy to implement traversability analysis approach ([13]) for mobile robots. Our system operates on depth images acquired with a Microsoft Kinect or an Asus Xtion Pro Live camera, and analyzes the visible area in front of the robot at a rate ranging between 10 and 25 fps on a notebook without using the graphics processing unit (GPU). Not relying on GPUs has the advantage of requiring less energy, which is a relevant issue for small-scale robots. Our approach has been implemented and evaluated in several sites including the catacombs of Priscilla in Rome, see Fig. 6.

In addition to this, we developed an exploration system that drives the robot with the goal to minimize both the uncertainty about the map appearance and the robot position inside the map. Uncertainty-reducing exploration is achieved by following the next frontiers while, at the same time,

Material classes		Structural classes	
Label	Occurrence (%)	Label	Occurrence (%)
Tufa	43.73	Wall	40.47
Regular tufa block	8.93	Floor	22.30
Rubble	6.34	Niche	6.00
Irregular tufa block	4.26	Ceiling	1.50
Modern brick	2.16	Other	1.01
Plaster	1.42	Pillar	0.87
Other	0.97	Fresco	0.55
Marble	0.26	Epigraph	0.36
		Arch	0.11
Unknown	31.65	Unknown	26.57

TABLE I

LABEL DISTRIBUTION IN THE DATASET OF THE CATACOMBS OF PRISCILLA IN ROME. FOR BOTH THE MATERIAL AND STRUCTURAL CLASSES, THIS TABLE SHOWS THE PERCENTAGE OF PIXELS LABELED WITH A CERTAIN CLASS.

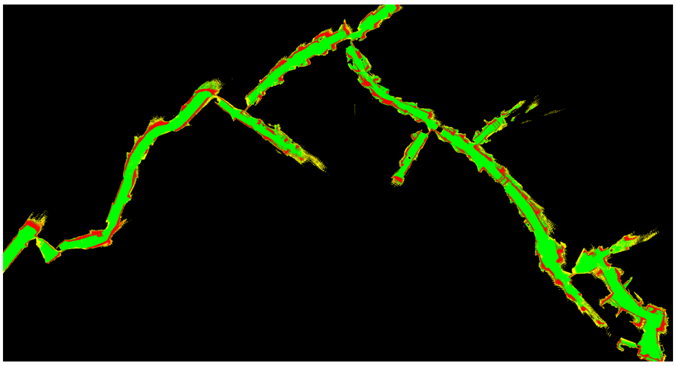


Fig. 6. Traversability map of a fraction of the explored space in the catacombs of Priscilla in Rome. Green, red, yellow and black patches denotes respectively traversable, not traversable, hard to traverse and unexplored areas.

trying to lead the robot towards loop closures. In order to choose frontiers with this criteria, it is important to have a guess of what does the map look like beyond said frontiers, which means predicting what will the system perceive in the unexplored areas of the map.

To improve the robot localization, we assume that the map we are exploring has some structure resemblance either within itself or with previously explored maps. While exploring, we then populate a database of local maps, that we will use as reference (we use FabMap2). When considering the possible appearance of the map beyond a given frontier, we query the database looking for maps that are similar to the local map surrounding the frontier. The OpenFabMap2 query returns a set of maps and probabilities associated to them, which indicate a confidence level on the similarity of the maps. We then consider each of the returned maps and, exploiting the Voronoi diagrams, we overlap them with the map we used as a query. Once the overlapping is performed, we use the overlapped map to predict which areas of the environment will be explored in the case that it actually looks as the overlapped map suggests.

In addition to that, we worked on a new approach for exploring a known 3D geometry during exploration. Most existing approaches to mobile robot exploration assume zero prior knowledge about the world. They mostly assume that the robot starts with an empty map and seeks to find a sequence of motion commands so that the robot covers the full environment with its sensors. In case a rough map of the space is available beforehand, we can speed up the exploration by planning optimal routes. Our work relies on a simplified Voronoi-style graph structure of the environment and seeks to find an efficient exploration strategy to cover the scene with the robots proximity sensor as fast as possible. Such a topo-metric graph can be provided by humans or can be automatically derived from floor plans or previously built maps. By knowing the topology of the environment including metric distances, the robot can generate more effective exploration trajectories, which find dead ends, small loops, and similar structures first so that the platform will not have to return to these locations later during the mission. We formulate the exploration problem as a Traveling Salesman Problem and we use its solution as a guideline for exploring the environment.

The last main aspect, but not less important, of the navigation task is the so called robust homing. This problem is related to the ability of a robot to navigate back to its start location at the end of the mission. From our point of view, there are two different situations that one should distinguish. First, the map is consistent. In this case, the robot can navigate back on the shortest path with a standard planning approach. Second, the map is inconsistent and as a result of that, the robot cannot plan a path based on its map. In this second case, the robot has to "rewind" its trajectory by incremental matching its current observations to the ones recorded when entering the scene. This rewinding can be implemented robustly but leads to long trajectories and also, in order to implement such a system, we need the possibility to evaluate if the map is consistent or not.

To overcome this problem, we developed a novel statistical approach to evaluate the consistency of maps without requiring ground truth information. In particular, we formulated a new measure taking into account the discrepancies between overlapping laser scans, with a method that is loosely based on the negative information approach proposed in [14]. We modeled the statistical properties of the consistency measure and computed a cascaded hypothesis test, which allows to discern when a map is consistent within a set confidence value. We also addressed the issue of rewinding a trajectory in order to bring the robot home in a safe manner if the map is inconsistent. If the odometry was perfect, i.e. it would provide the correct robot pose in the world frame, we would be able to simply invert the motion commands performed by the robot and it would safely reach the starting location. However, since this is not the case, we employed the iterative closest point (ICP) matching algorithm to correct the robot position after carrying out each motion. Addionally, in order to correct the position of the robot while rewinding the trajectory, we store the pointclouds as well as their positions in the odometry frame in a persistent storage while the robot explores the catacombs. As soon as the robot understands that he is "lost", he can rewind the trajectory saved up to the current moment.

E. User Interface and Mission Control

The target users of the robot are cultural heritage experts, called surveyors, that wants to explore and build 3D models of the archaeological site at hand. Surveyors deploy the robot at the entrance of an archaeological site and they monitor it through a Mission Control Interface (MCI). The interface looks like to the one of a modern 3D video-game. The surveyor is able to see a birds eye view of the platform moving inside a local 3D reconstruction of the robot surrounding environment. In addition to this, other video streams can be projected on the 3D reconstruction and a 2D mini-map can be visualized providing a global view. This particular interface configuration is called Multi-Modal (see Fig. 7). During the exploration the operator can select regions of the environment and annotate them for offline analysis and classification after the end of the

Fig. 7. Mission Control Interface (MCI) in Multi-Modal mode in the catacombs of Priscilla in Rome.

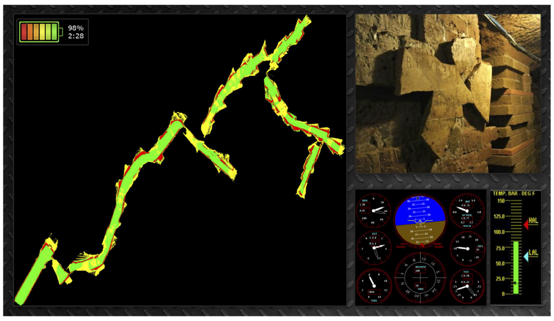


Fig. 8. Mission Control Interface in Supervisory Mode in the catacombs of Priscilla in Rome.

mission. Moreover, the robot itself is able to autonomously highlight interesting areas (such as ones containing pots, frescoes or bones) for further inspection.

The multi-modal configuration requires a large bandwidth that is provided through a dedicated point-to-point Wi-Fi connection. While this type of connection is suitable for many indoor environments, it has a limited range in a catacomb and its performance will quickly degrade until the connection is lost. When the network performance starts to decrease, the interface switches to a Supervisory Mode that is able to handle a lower bandwidth. A snapshot of the interface is shown in Fig. 8. The interface shows a light 2D representation of the environment through a traversability map ([13]). Colors, in this representation, provide qualitative information about the terrain: black indicates unexplored areas, green safe to traverse areas, yellow hard to traverse areas and red denotes dangerous areas that the robot cannot overcome. The user has the possibility to select a target location by clicking on an area on the map. The robot will automatically compute the shortest and safest path to the target location without the need for direct teleoperation. While moving through unexplored areas, the robot will update the 2D map based on the new information perceived with its sensors. Moreover, it can recognize and highlights interesting features and objects in the environment. For example, the platform can report that he has found an artifact made of ceramic in a certain location. The surveyor can at this point request a picture, as shown on the top right view of the MCI in supervisory mode, to visualize the discovery that can be eventually annotated in the map.

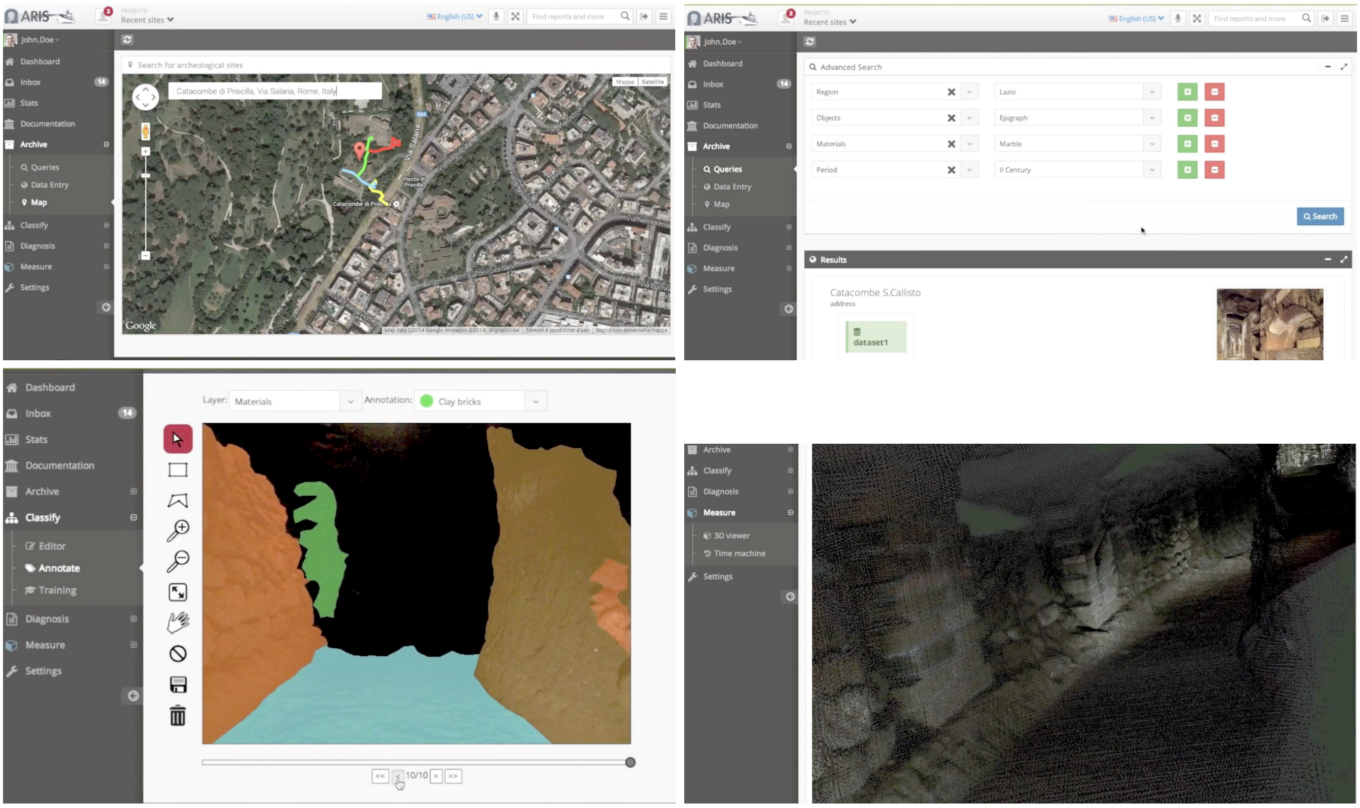
There are cases where the robot is unable to maintain the connection. When this happens, the MCI enter in autonomy mode. This mode allow the user to set out short missions during which it will loose contact with the robot. As an example, the user could ask to the robot to explore a certain region for 30 min and to report on the traversability of the terrain and on items of interest it may find. During the mission the robot will not be in contact with the operator. When the mission time is over, or the entire area is explored, the robot will return where the sub-mission started reporting by means of the supervisory interface. On request, the robot can send additional data such as pictures. Based on this information, the operator can plan other sub-mission as to go elsewhere or explore more accurately the same area.

F. Digital Archive

The tool for the visualization of archaeological sites, called ARIS (Archaeological Information System), is a cloud based application that allows to archive and analyze the data collected by the robot during the missions. ARIS is developed as a Software as a Service (SaaS) and, as such, it is available through a web interface. ARIS offers a number of services related to cultural heritage sites, including: documentation, archival, classification, diagnosis and measure.

ARIS core function is the data archive. Authorized users can upload the data to ARIS that then makes it available for controlled sharing and analysis. Data can include raw logs, point cloud maps, meshed maps, semantic point cloud maps and segmented images. Data can be queried both geographically through a GIS interface or semantically through a query interface. Specific interfaces are devoted to annotating images and inspecting the results of semantic classification. Similarly, we have an interactive 3D interface, based on WebGL technologies, that allows for inspecting 3D models. Fig. 9 shows some of the service interfaces implemented in ARIS.

Streaming 3D data can be very challenging due to its huge size. While the issue has been addressed in a straight forward way for point clouds by means of sampling, we found it very challenging to address streaming in the case of meshes. Indeed, we have devoted a considerable effort in understanding how to stream complex textured meshed 3D models in web browsers. In particular, we developed a prototype of a client-server architecture. The server maintains an efficient representation of the model's spatial structure. In this way, the client can request only the parts of the model that are relevant to the visitors current position, thus reducing loading time. The model parts are provided at multiple resolutions. Depending on the vicinity to the visitors position, the server provides links to corresponding model parts in appropriate resolutions. To exchange information between server and client, a communication protocol has been specified utilizing the REST paradigm and JSON objects. We rely on the scriptable MeshLab tool to generate different resolutions of the model parts. The web client is executed in a browser and communicates with the server providing the virtual position of the visitor, and requesting the URLs where



Catacombe di Priscilla, Via Salaria, Rome, Italy

Region: Lazio
Objects: Epigraph
Materials: Marble
Period: II Century

Catacombe S.Callisto
address
dataset1

Layer: Materials
Annotation: Clay bricks
10/10

Fig. 9. From left to right and from top to bottom: ARIS geographical information system, ARIS semantic query engine, ARIS semantic annotation tool and ARIS point cloud 3D viewer.

to download the relevant data.

IV. CONCLUSIONS

The ROVINA project has the goal to develop autonomous mobile robots to make faster, cheaper and safer the surveying of cultural heritage. Despite the project is still at a middle stage, it has already achieved a relevant number of important goals. Three prototypes of the robot are available and they can be used to record data during missions. In particular, the current SLAM, SfM and semantic segmentation algorithms have shown to be reliable tools to help surveyors in digitizing har-to-access archaeological sites. Moreover, the digital archive generated from the acquired data can be easily examined through user friendly interfaces developed within the project.

REFERENCES

[1] M. Vergauwen and L. Van Gool, “Web-based 3d reconstruction service,” *Machine vision and applications*, vol. 17, no. 6, pp. 411–426, 2006.

[2] T. Moons, L. Van Gool, and M. Vergauwen, *3D Reconstruction from Multiple Images: Part 1: Principles*. Now Publishers Inc, 2009.

[3] G. Grisetti, R. Kmmerle, and K. Ni, “Robust optimization of factor graphs by using condensed measurements.” in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2012, pp. 581–588.

[4] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, “g2o: A general framework for graph optimization,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Shanghai, China, May 2011.

[5] J. Serafin and G. Grisetti, “Using augmented measurements to improve the convergence of icp,” in *Proc. of the Int. Conf. on Simulation, Modeling and Programming for Autonomous Robots (SIMPAR)*. Bergamo, Italy: Springer, 2014, pp. 566–577.

[6] S. Georgoulis, M. Proesmans, and L. Van Gool, “Tackling shapes and brdfs head-on,” in *3D Vision (3DV), 2014 2nd International Conference on*, vol. 1. IEEE, 2014, pp. 267–274.

[7] M. Beinhofer, H. Kretzschmar, and W. Burgard, “Deploying artificial landmarks to foster data association in simultaneous localization and mapping.” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*. IEEE, 2013, pp. 5235–5240.

[8] P. Agarwal, G. Grisetti, G. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss, “Experimental analysis of dynamic covariance scaling for robust map optimization under bad initial estimates,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Hong Kong, China, 2014.

[9] M. Mazuran, G. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss, “A statistical measure for map consistency in slam,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Hong Kong, China, 2014, accepted for publication.

[10] F. Basso, A. Pretto, and E. Menegatti, “Unsupervised Intrinsic and Extrinsic Calibration of a Camera-Depth Sensor Couple,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2014.

[11] D. Tedaldi, A. Pretto, and E. Menegatti, “A Robust and Easy to Implement Method for IMU Calibration without External Equipments,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*. IEEE, May 2014.

[12] A. Hermans, G. Floros, and B. Leibe, “Dense 3d semantic mapping of indoor scenes from rgb-d images,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*. IEEE, 2014, pp. 2631–2638.

[13] I. Bogoslavskyi, O. Vysotska, J. Serafin, G. Grisetti, and C. Stachniss, “Efficient traversability analysis for mobile robots using the kinect sensor,” in *Proc. of the European Conference on Mobile Robots (ECMR)*, Barcelona, Spain, 2013.

[14] D. Hähnel, S. Thrun, B. Wegbreit, and W. Burgard, “Towards lazy data association in slam,” in *Robotics Research. The Eleventh International Symposium*. Springer, 2005, pp. 421–431.