A USER PERSPECTIVE ON THE ROVINA PROJECT

Vittorio Amos ZIPARO (*), Daniele CALISI (*), Giorgio GRISETTI (**), Jacopo SARAFIN (**),
Marc PROSMANS (***), Luc VAN GOOL (***), Bastian LEIBE (****),
Maurizio Di Stefano (*****), Luigi PETTI (*****), Wolfram BURGARD (******),
Fabrizio NENCI (*******), Igor BOGOSLAVSKYI (*******), Olga VYSOTSKA (*******),
Maren BENNEWITZ (*******), Cyrill STACHNISS (*******)

(*) Algorithmica Srl, Largo Alfredo Oriani 12, 00152, Rome, IT, +39 06 58303095, *ziparo@algorithmica.it*
(**) La Sapienza University of Rome, Via Ariosto 25, 00185 Rome, IT, +39 06 77274121, *grisetti@dis.uniroma1.it*
(***) KU Leuven, Kasteelpark Arenberg 10 bus 02441, 3001 Heverlee, BE, +32 163217 05, luc.vangool@esat.kuleuven.be
(****) RWTH Aachen, UMIC Research Centre, 52056 Aachen, +49 241 8020762, *leibe@umic.rwth-aachen.de*
(*****) ICOMOS Italia, Via di San Michele 13, 00153 Rome, icomos@icomositalia.com
(******) University of Freiburg, Georges-Köhler-Allee 79, DE, +49 761 2038026, *burgard@informatik.uni-freiburg.de*
(*******) University of Bonn, Nussallee 15, 53115 Bonn, DE, +49 228 732713, cyrill.stachniss@igg.uni-bonn.de

Abstract

ROVINA is a research project funded by the EC within FP7. ROVINA will provide tools for mapping and digitizing archeological sites - especially for difficult to access sites - to improve the preservation and dissemination of cultural heritage. Current systems often rely on static 3D lidar, traditional photogrammetry techniques, and are manually operated. This is expensive, time consuming, and can be even dangerous for the operators. ROVINA exploits the strong progress in robotics to efficiently survey hazardous areas and aims at making further progress in the reliability, accuracy, and autonomy of such systems.

Keywords (max 5): Robotics, Artificial Intelligence, Digitalization, and Archaeological Site

Introduction

Europe has a wealth of cultural heritage sites and Italy is the country that has the largest number of them in Europe[1]. The conservation of such sites is a challenging task, as it requires periodical surveys for which teams of experts have to carry heavy equipment in the field. Surveys require a substantial amount of time and manual labor by experts, thus rendering this task expensive. Furthermore, the process is slow and sometimes prone to error. In addition to that, many sites are dangerous to access for humans and pose serious risks to the surveyors in the field.

In this paper, we present the ROVINA project (http://rovina-project.eu). ROVINA is a three and a half-year research project that is co-funded by the European Commission in the frame of the 7th Framework Programmed (FP7-ICT-600890). The ROVINA consortium is composed of the University of Bonn (GE), RWTH Aachen University (GE), the University of Freiburg (GE), the University of Leuven (BE), Sapienza University of Rome (IT), Algorithmica Srl (IT), and the Italian Committee of the International Council of Monuments and Sites (IT). ROVINA aims at making surveying of cultural heritage sites faster, cheaper, and safer through the use of autonomous robots, which will

---

[1] http://whc.unesco.org

enable 3D reconstructions of hazardous sites at a new scale and quality. The ROVINA robots are meant to autonomously explore archeological sites. The data collected during the exploration is processed and stored in the cloud and will deliver advanced analysis services for structural engineers, historians, and preservation experts. As the models will be accurate and visually appealing, ROVINA will also feature a browser-based online museum for the general public.

The project will be evaluated through two case studies in the catacombs of Rome and Naples. Indeed catacombs are a very interesting case study and are found in several cities. There are approximately 100 catacombs in Italy, scattered through 29 cities. As one may expect, most of the catacombs are in Rome[2], that counts 48 of them, but there are another 11 concentrated in Naples[3]. Catacombs are not limited to Italy, and there are others in Europe and around the world[4], for example in Paris and others in England, Ukraine, Malta, Egypt, and the Czech Republic. Catacombs often extend for several kilometers and at multiple depth levels. For example, the Roman catacomb of S. Priscilla extend for 13km over multiple floors and this catacomb has been selected as ROVINA's primary test site. Most catacombs are partially unexplored due to the high risk of entering them. On the one hand, catacombs are unstable and there is a high risk of collapse. On the other hand, several of the (non-ventilated) catacombs, for example those located in Rome, yield a high concentration of radioactive radon gas limiting human intervention to time periods of at most 15-30 min.

State of the art

The ROVINA project aims at automating existing practices, thus it is conceived to seamlessly integrate with current activities in the realm of cultural heritage. In the following of this section, we provide a brief overview of the core activities related to the preservation and documentation of archaeological sites.

Measuring is the key building block of any surveying activity. There are two types of measurements: "direct" and "indirect". Direct measurements are those that the operator performs directly (and manually) on the artifact with classic tools such as Charles metric, the plumb line, water systems, etc. In these cases, operators either build polygons, possibly with the help of topography, to give "robustness" to measurements or, alternatively, they triangulate manually. Instead, indirect measurements are usually based on laser scanning and/or image analysis (using traditional photogrammetry techniques). Both, laser and image-based reconstructions allow morphometric surveys. Laser-based systems, and in general instruments based on "time-of-flight", are typically used for large sites or large artifacts. These sensors can be rather expensive but offer a high precision. The Leica Cyrax scanner, for example, has a scanning range of 2m - 150m with a resolution that is higher than 1cm and a full 3D scan takes a few minutes. Depending on the application at hand, the requirements on the precision of the measurements may vary. For documentation activities, a range resolution of 1cm is often sufficient. When the task at hand is related to diagnosis instead, the required precision may increase. Image-based systems have been shown to provide precisions similar to those of Lidar scanners, especially when sufficient images close to the surfaces can be taken. They have the advantage that image/color data are available, that are perfectly aligned with the 3D data. The captured data - as for lidars - produce point clouds that need further post-processing. Image-based systems usually employ high-resolution commercial cameras such as the 10MPixel Nikon D200 and

---

[2] http://it.wikipedia.org/wiki/Catacombe_di_Roma
[3] http://it.wikipedia.org/wiki/Catacombe_di_Napoli
[4] http://en.wikipedia.org/wiki/Category:Catacombs

commercial photogrammetry software or, more recently, self-calibrating structure-from-motion systems as ROVINA uses (Vergauwen and Van Gool, 2006; Moons et al., 2009) or multi-view stereo approaches.

Documentation is the core activity in surveys and aims at producing digital archives of the site under observation. In the realm of cultural heritage, documentation activities are performed by public bodies, which for the specific case of Italy usually are superintendencies and ministries. The digital archives can take the form of 3D models that can be either purely geometric or can also include textures from images. When the surveys have a purely documentary purpose, i.e., they are not for the purpose of pure measurement or diagnosis, 3D models can have a lower resolution and do not need to be extremely precise but they should rather be visually appealing. Their main goal is to disseminate cultural heritage to broad audiences. To this end, tools such as virtual museums are of paramount importance.

Classification activities are usually tied to documentation tasks and pertain to the categorization of elements of a site into taxonomies or ontologies with different degrees of complexity. For example, in an industrial setting you may want to categorize rooms of a plant - and equipment therein - based on their functional properties. In a cultural heritage site, you may want to classify architectural components and items based on a number of parameters such as period of construction, materials used, state of conservation ("stato di conservazione") and so on. Human users usually perform classification by manually tagging items and portions of the environment. Especially when data is collected on a geographical scale, the models are archived into Geographical Information Systems (GIS). Data can then be queried on both a geographical and qualitative level. For example, one may look for all the sites built before 1000 B.C. in Italy or all the pots made of ceramic from Germany.

Diagnostics is typically associated with the goal to collect data on the state of the surveyed areas in order to prevent damage or perform restoration. From a practical perspective, the diagnostic activities have the purpose of generating specific deliverables that in the context of cultural heritage are the Table of Deterioration ("Tavola del Degrado") and the Table of Materials ("Tavola dei Materiali"). The Table of Deterioration is a map showing possible deteriorations such as cracks and molds. The table follows the standards dictated by the UNI-NorMal commission that is composed by the Ente Nazionale Italiano di Unificazione1 (UNI) and the Commissione NORmalizzazione MAteriali Lapidei (NorMaL). At a European level, the standards are dictated by CEN (i.e., WS Construction, WS Measurement, WS Material). The Table of Materials instead maps the areas of the site under survey to the different materials of which they are composed.

The ROVINA Robot

The core idea behind the ROVINA project is that one can improve measurement activities by replacing standard tools, which must be carried by operators into the site, with autonomous or semi-autonomous robots that can be remotely supervised. Thus, one of the goals of ROVINA is to assess the technological and commercial feasibility of building robots capable of surveying cultural heritage sites. For this reason, the ROVINA robot, rather than being a production-ready system, is a prototype that has been designed having in mind flexibility and reduced cost. For example, the robot mounts the Ocular RE05 laser that has a maximum range of 30m and offers a lower resolution compared to the above mentioned laser scanners. Compared to the Faro Focus3D X 330, for example, that has a range of 330m and a resolution of 2mm, the RE05 is inferior. Nevertheless, the RE05 costs much less and is very flexible as one can dynamically control the scanning field (full, bounded elevation, and region). If

the ROVINA system succeeds and operates with the RE05, it will be easy to upgrade the robotic system using a Faro laser scanner, as the algorithms are independent of the specific laser. Better performance of the laser will simply mean better performance of the entire system. Similar considerations apply also to other components. For example, we have opted for laptops rather than for embedded computers. This makes the life easier for developers and allows us to easily upgrade the hardware if needed. The price we pay is an increased volume of the robot and thus reduced mobility. At the same time, there is very little dependency on one producer or retailer when upgrading or replacing parts of the system.

Currently, we have built three robot prototypes. Figure 1 depicts one of these prototypes and Figure 2 shows a subset of the sensors of the ROVINA robot. The robot was designed based on a number of environmental and software requirements that have been gathered in a number of inspections of Catacombs (ROVINA Consortium, 2013).

The first requirement was the mobility. The robot must be able to navigate rough terrain and tackle small obstacles such as stairs or pallets, while still being able to go through the small passages of a catacomb that are roughly 80cm wide. Figure 3 depicts typical environments and it illustrates some of the mobility challenges that we encountered during our preliminary inspections of the catacombs. These include debris, stairs, and holes in the ground. Given the peculiar structure of the environment the best option would have been to build a custom tracked robot with flippers. Nevertheless, for the sake of budget reduction and for providing a working prototype as soon as possible we decided to customize an existing, off-the-shelf, tele-operated robot. We surveyed a number of existing solutions and we finally identified the Mesa Robotics Element (rightmost element of Figure 2) as the platform with the best performance/cost ratio. This robotic platform is agile enough to traverse most of the obstacles encountered in the catacombs and is small enough to navigate through narrow corridors.

The Mesa Element is a remote controlled platform that we had to extend for the purposes of the project. We had to equip the platform with two laptops and a sensor suite (Figure 2 and Figure 5) composed of a Lidar, two RGB-D cameras, an array of RGB cameras, an inertial measurement unit (IMU), battery status monitors and thermal/humidity sensors. Regarding the array of RGB cameras, we were not able to identify an off the shelf solution to allow sufficient amount of control in the given environmental circumstances. Several iterations have been executed testing different camera configurations, including firewire and Ethernet based communication. We finally converged to custom arc setup of 7 Ethernet based cameras with a 2MPix resolution each, and having lenses of 60 degree opening angle. Allowing an omnidirectional view on the catacomb's walls, outside the view range of the scanner or the other devices. The camera acquisition is designed to run on a separate second computer platform, which is independent from the platform needed for navigation and scanning.

In order to avoid jeopardizing the autonomy of the Mesa Element by using its batteries to power the extra sensors and computational units, we designed an additional power system. Experimental evidence shows that the additional power system allows our robot to run at full power for ~6 hours. Considering a target speed of 30cm/sec we estimate that the robot can survey approximately an area corresponding to a path 6.5km long.

To house these components, we have built a case of aluminum and polycarbonate on top of the Element. The housing has been made shockproof where necessary and has been designed for maximizing ease of use. Figure 4 shows three iterations of the robot case, while Figure 1 shows the final version mounted on the Mesa Element. The robot must be able to operate in the target

environments without damaging on-board devices. This means that the housing must be robust to water dripping from the ceiling, some dust, and high levels of humidity (~95%). To deal with these challenges, we provided the robot with an ingress protection to dust and water equivalent to IP52. The robot also has a cooling system capable of controlling both temperature and humidity inside the case.

Mission Control

The surveyor deploys the robot at the entrance of an archeological site and remotely operates it through the mission control interface. The interface is very similar to the one of a 3D videogame. The surveyor has a bird's eye view of the robot and looks at a local 3D reconstruction while tele-operating it with a joystick. Automatic navigation systems avoid the robot causing damage due to human error, for example stopping the robot when it gets too close to an obstacle. Additionally, video streams can be projected on the 3D reconstruction and a 2D map is shown in order to provide a global view. This interface configuration of the mission control interface is called *multi-modal*. During the exploration the operator can select regions of the environment and annotate them for further analysis and classification after the completion of the mission. The robot itself can also highlight interesting areas (such as ones containing pots, frescos or bones) for further inspection.

The multi-modal configuration requires a lot of bandwidth that is provided by 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. To this end, as soon as the connection quality reaches a given threshold the robot starts releasing Zigbee devices (see Figure 5 right). These devices act as repeaters and allow for maintaining connectivity at the price of a lower bandwidth. These devices can create a low-cost wireless mesh network that can last years thanks to its low-power characteristics and that can act as a sensor network and continuously stream a wide set of data such as temperature, humidity and pressure. When the robot starts deploying zigbee nodes, the interfaces switches to a so-called *Supervisory Mode* in order to adapt to the new bandwidth limitations. An example interface is shown in Figure 6 and 7. The map provides a more abstract 2D representation of the environment called traversability map (Bogoslavskyi et al., 2013). Colors in the map provide qualitative information on the terrain: black denotes unexplored areas, green safe to traverse areas, yellow denotes the areas of uncertainty, where there is not enough information to make a safe decision and red denotes dangerous areas that the robot cannot traverse. The user can select a target location by clicking an area in the map. The robot will automatically compute the safest path and follow it to the target location without the need for direct tele-operation. While traversing unexplored areas, the robot will update the 2D map accordingly to the new information it has perceived through its sensors. This is achieved through exploration techniques that consider the expected gain of novel information (Stachniss and Burgard, 2012; Stachniss and Burgard, 2003). In addition to that, the robot can recognize and signal interesting features and objects in the environment. For example, the robot can communicate to the surveyor that it has identified a manufact made of ceramic in a given location. The surveyor can then request a picture (as shown on the top left mockup of the mission control interface in supervisory mode) to further assess the discovery that then can be annotated in the map.

Although the zigbees can greatly increase the time that the robot is connected, the robot will eventually run out of devices to deploy. When this happens, the mission control interface will enter a so-called *autonomy mode*. In this mode the user pre-plans short missions during which it will lose contact with the robot. For example, the user may ask the robot to explore a given region for 30 min and to report on the traversability of the terrain and on interesting items it may encounter. During the

mission the robot will not be in contact with the operator. When time is over, or when the entire area is explored, the robot will return where the mission had started and report through the supervisory interface the traversability map of the explored area annotated with the interesting objects it has found. On request, the robot may transmit additional data such as pictures. Based on this information, the operator can plan successive mission as, for example, to further explore the area or to go elsewhere.

Cloud Services

Once the measurement activities (i.e. the surveying mission) are over, the robot will have gathered an enormous amount of data from its sensors that include laser scans and pictures. These data are uploaded to a cloud computing facility in order to provide services for documentation, classification and diagnosis.

At the core of all the services there is a high-resolution textured 3D model of the archeological site. Note that, although the robot builds a 3D model during the mission (Kümmerle et al., 2011), due to real-time requirements, the model is approximate (Grisetti et al., 2012). However, this model is just accurate enough to enable autonomous navigation and operator awareness. Our focus is set on making sure that the robot does not get lost rather than on the accuracy of the resulting models. To this end, we have devoted a considerable effort in making the approach more robust to outliers (Agarwal et al., 2014), to assessing the degree of consistency of maps (Mazuran et al., 2014) and to automatically calibrating the sensors (Basso et al., 2014; Tedaldi et al., 2014). Despite being approximate, the 3D model that has been reconstructed online can be fed as an initial guess to more accurate, yet time-consuming, techniques (Vergauwen and Van Gool, 2006; Moons et al., 2009) which are then used to compute - off-line in the cloud - a highly accurate model based on the large amount of pictures that have been collected by the camera array. Figures 8, 9, and 10 show some examples of portions of the reconstructions.

The cloud-based services provide tools for *classification* activities. In particular, there is a web tool (see Figure 11) that allows annotating areas of some of the pictures taken by the robot based on the composing materials and archeological taxonomies. These manual annotations are then fed to machine learning algorithms that generalize the examples, and use them both for identifying interesting objects/areas during the mission and for annotating the accurate 3D model. Our approach to semantic mapping (Hermans et al., 2014) won the IEEE ICRA'14 Best Vision Paper Award. Once the 3D models are semantically annotated, users can pose complex queries. For example, one can look for all the niches made out of tuff in catacombs in the region of Lazio.

ROVINA also offers a number of services related to *Diagnostics*. Indeed, classified areas can also be used for automatically generating reports such as tables of materials and tables of decay. This can be simply achieved by generating sections or projections of the 3D models along with their annotations. Another use of the models in the realm of diagnostic activities is the possibility to compare 3D models of the same site at different points in time. Indeed, such a tool can greatly help engineers performing structural analysis, for example highlighting enlargement of cracks or collapses in the structure.

Finally, as the 3D models generated by ROVINA are both accurate and visually appealing they are also used for *Documentation* purposes. The ROVINA services include a browser-based 3D virtual site viewer that allows virtual tourists to visit high quality 3D textured reconstructions of the sites, which are made interactive thanks to the semantic annotations and to additional information (e.g. videos, text etc.).

Conclusions

The ROVINA project, despite being at an early stage, has already achieved a number of important goals. Three prototypes of the robot are now available and are recording data. The obtained datasets show that our 3D reconstruction and semantic segmentation algorithms are capable of generating small-scale models that have many of the features of the final system. We are currently improving and integrating our developed techniques, while exploring new challenges related to autonomous navigation and user interfaces.

Bibliography

P. Agarwal, G. Grisetti, G. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss. 2014. *Experimental analysis of dynamic covariance scaling for robust map optimization under bad initial estimates*. In Proc. of the Int. Conf. on Robotics & Automation , 2014.

F. Basso, A. Pretto, and E. Menegatti. *Unsupervised Intrinsic and Extrinsic Calibration of a Camera-Depth Sensor Couple*. .In Proc. of the Int. Conf. on Robotics & Automation, 2014.

I. Bogoslavskyi, O. Vysotska, J. Serafin, G. Grisetti, and C. Stachniss. 2013. *Efficient traversability analysis for mobile robots using the kinect sensor.* In Proc. of the European Conference on Mobile Robots, 2013.

G. Grisetti, R. Kümmerle, and K. Ni. 2012. *Robust optimization of factor graphs by using condensed measurements*. In Proc. of the Int. Conf. on Intelligent Robots and Systems, 2012, pp. 581–588.

A. Hermans, G. Floros, and B. Leibe. 2014. *Dense 3D semantic mapping of indoor scenes from RGB-D images.*In Proc. of the Int. Conf. on Robotics & Automation, 2014.

R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. 2011. *g2o: A general framework for graph optimization*. In Proc. of the Int. Conf. on Robotics and Automation, 2011.

M. Mazuran, G. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss. 2014. *A statistical measure for map consistency in slam*. In Proc. of the Int. Conf. on Robotics & Automation, 2014.

T. Moons, L. Van Gool, and M. Vergauwen. 2009. *3D reconstruction from multiple images Part 1: principles*. Foundations and Trends in Computer Graphics and Vision, Vol.4, No.4, 2009.

ROVINA Consortium. 2013. *Rovina Deliverable 1.1: Environment analysis, hardware requirement analysis and data collection*. http://rovina-project.eu/del/del-1-1.pdf.

C Stachniss and W Burgard. 2003. *Mapping and exploration with mobile robots using coverage maps.* In Proc. of the Int. Conf. on Intelligent Robots and Systems, 2003.

C. Stachniss and W. Burgard. 2012. *Particle Filters for Robot Navigation.* Foundations and Trends in Robotics, Vol. 3, Number 4, pages 211-282, doi: 10.1561/2300000013, 2012

D. Tedaldi, A. Pretto, and E. Menegatti. 2014. *A Robust and Easy to Implement Method for IMU Calibration without External Equipments*. In Proc. of the Int. Conf. on Robotics & Automation, 2014.

M. Vergauwen and L. Van Gool. 2006.*Web-based 3D reconstruction service.* Machine Vision and Applications, Vol.17, pp. 411-426, 2006.

Figure 1 – The designed and constructed ROVINA robot (here depicted without the camera array).

Figure 2 – A subset of the sensors and the base of the ROVINA robot platform.

Figure 3 – Typical environments encountered by the ROVINA robot.

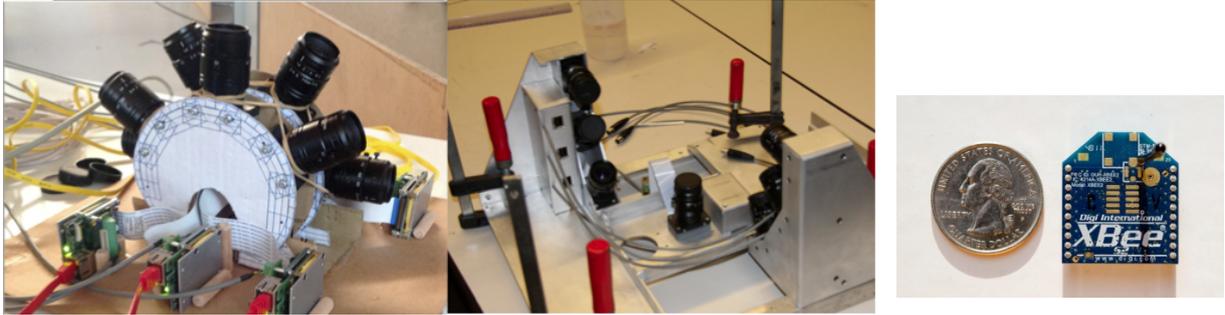*Figure 4 – Three iterations of the robot chassis design process.*

Figure 5 – Considered setups for image acquisition. Left: test setup, middle: final camera configuration, right: Zigbee node.

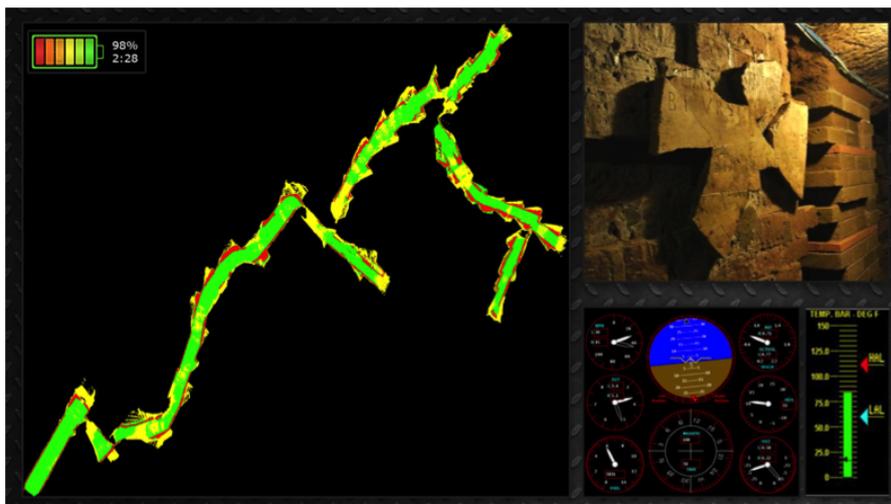*Figure 6 – Guardian angel view of the user interface.*

Figure 7 –Control interface showing the traversability analysis (green = traversable, red = non traversable, yellow = areas of high classification uncertainty).
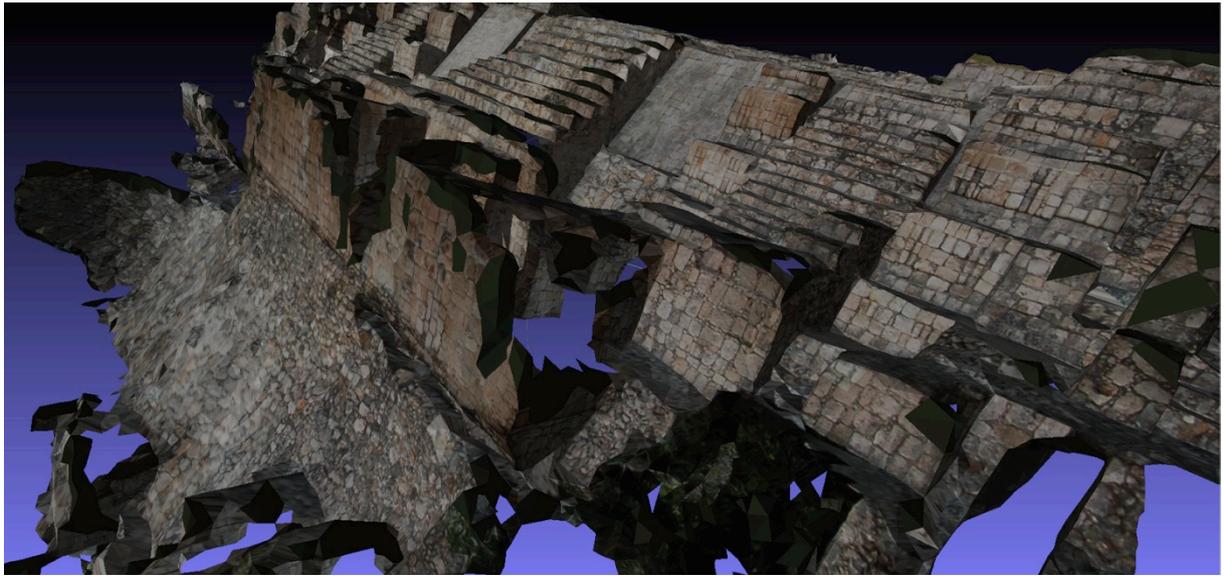
Figure 8 – High-resolution reconstruction obtained through SfM.

Figure 9 – High-resolution reconstruction of a portion of the catacomb obtained through SfM.

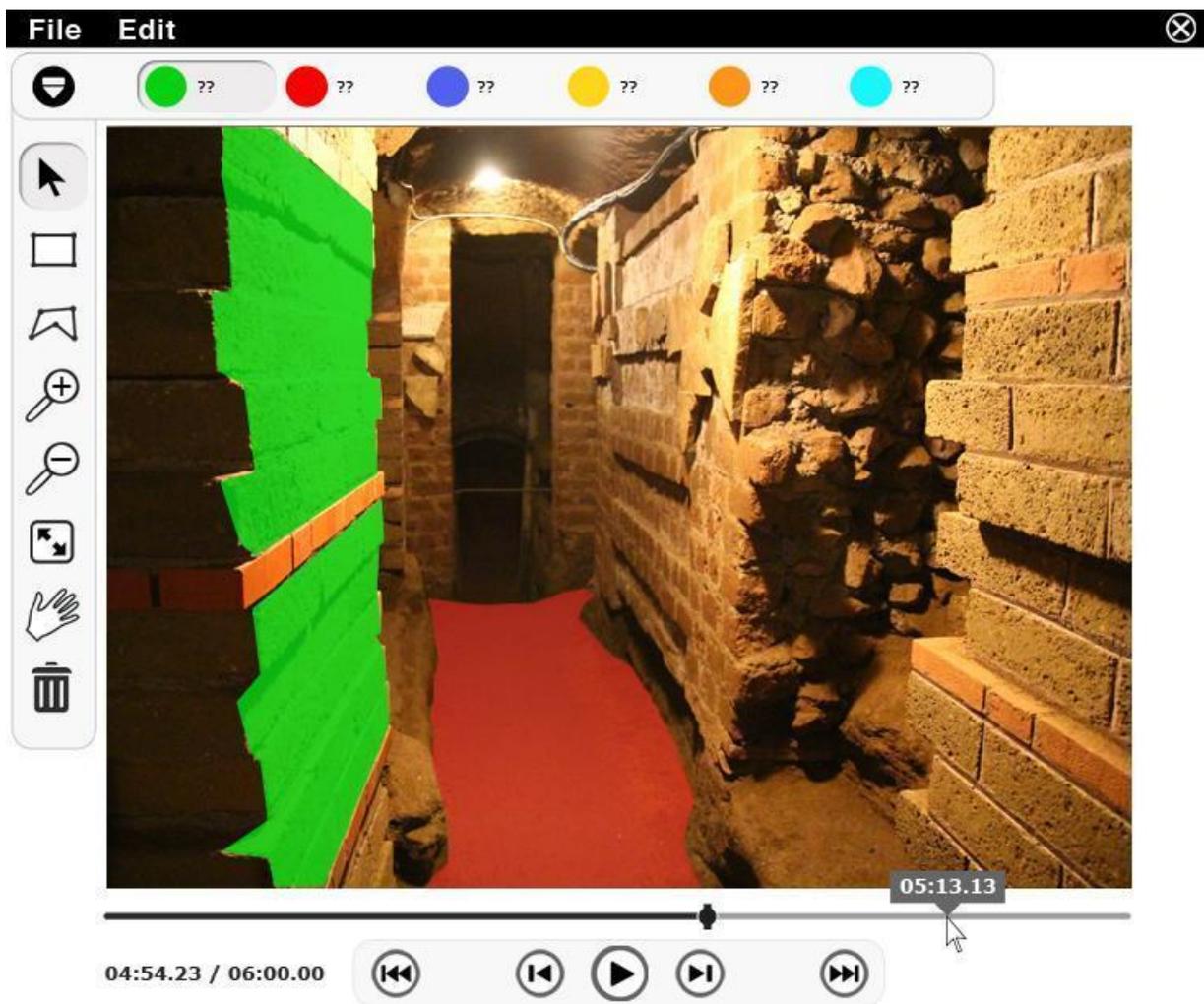*Figure 10 – Detail from the reconstruction in Figure 9.*

Figure 11 – Browser based semantic annotation tool. Environment structures can be colored and labeled through an intuitive user interface.