UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN
MATHEMATICS AND COMPUTER SCIENCE

Cycle XXXV

TITLE OF THE Ph.D. THESIS

Large Scale Acquisition of Complex Environments by Data Fusion from Mobile Visual and Depth Sensors

Scientific Disciplinary Sector(s)
INF/01 INFORMATICA

Ph.D. Student : Armando Arturo Sánchez Alcázar
Supervisor : Prof. Riccardo Scateni (UniCa)
Co-Supervisor : Prof. Giorgio P. M. Vassena (UniBS)

Final exam. Academic Year (2021-2022)
Thesis defence: February 2024 Session

Abstract

This thesis addresses the pressing need in industry and Architecture, Engineering, and Construction (AEC) for efficient scanning and modeling of large, structured environments. The primary objective is the development and application of a state-of-the-art wearable system integrating laser and visual scanning technology. This innovative mobile mapping system (MMS) is designed to capture complex man-made structures in diverse settings, ranging from indoor spaces to challenging outdoor environments and underground settings.

The MMS combines high-resolution photographic data with laser scanning and inertial inputs to generate detailed, dense point clouds, offering extensive coverage and depth. This approach addresses the limitations of geometric models by providing photorealistic representations essential for applications requiring accurate location recognition, object identification, and content creation.

Key contributions of this research include the development of the MMS prototype, capable of adapting to various data acquisition scenarios while ensuring scalability and minimal data redundancy. Additionally, the thesis explores the practical application of the MMS in a pilot AEC project, demonstrating its effectiveness in real-world scenarios for construction monitoring and integration with Building Information Modeling (BIM).

Through a combination of technical development, rigorous testing, and practical application, this thesis advances the field of mobile mapping. It opens new avenues for spatial data acquisition and modeling, particularly in environments where traditional mapping techniques fall short.

Keywords: Mobile Mapping Systems, Architecture Engineering Construction, 3D Reconstruction, Indoor Mapping, LiDAR Scanning, Panoramic Imaging, Spatial Data Acquisition, Wearable Technology, Point Clouds, Building Information Modeling.

Acknowledgments

Completing this PhD thesis marks both an end and a beginning, and I owe my gratitude to several individuals whose guidance and support have been invaluable.

I would like to express my sincere thanks to Prof. Giorgio Vassena and Matteo Sgrenzaroli at Gexcel for their guidance and expertise throughout this project. Their practical insights and assistance have been greatly appreciated and have significantly contributed to my research.

Equally, I am thankful to my academic supervisor, Prof. Riccardo Scatteni from Università degli Studi di Cagliari (UniCa). His academic guidance and constructive feedback have been essential in shaping the direction and quality of this work.

I am grateful for the collaborative and innovative environment provided by my colleagues at Gexcel, specifically Emmanuele Alimonda, Marco Bianchetti, Davide Davelli, Mario Pezzoni, and Pablo Pittaluga. Their support, expertise, and contributions have been beneficial to my research journey.

Special thanks go to the researchers at CRS4, notably Eva Almansa, Moonisa Ahsan, Enrico Gobbetti, and Giovanni Pintore, for their collaboration, insights, and contributions that have enriched my research experience.

Lastly, my thanks also go to my friends and family and, most importantly, my parents, for their understanding and support during this demanding period of my academic career.

Armando Sánchez
Brescia, Italy
24th November — 2023.

Contents

List of Figures

List of Tables

Preface

This thesis represents a summary of the work done from 2019 to 2023 at GEXCEL, a geomatics, 3D mapping and surveying company located in Brescia, Italy, under the direction and supervision of **Prof. Giorgio Vassena**, whom I really want to thank for the opportunity to be part of his company.

The scientific work in this thesis has been performed within the international framework of **EVOCATION** (Advanced Visual and Geometric Computing for 3D Capture, Display, and Fabrication) project, which was a leading European-wide doctoral Collegium for research in Advanced Visual and Geometric Computing for 3D Capture, Display, and Fabrication supported by European Union's H2020 research and innovation program grant 813170. The consortium participants are the University of Rostock (UNIRO), the Center for Research, Development and Advanced Studies in Sardinia (CRS4), the University of Zurich (UZH), the Italian National Research Council (CNR), the Technical University of Vienna (TUW), Fraunhofer IGD (FHG-IGD), and the two companies Holografika (HOLO) and GEXCEL. The goal of this ITN is to address the current and future major challenges in scalable and high-fidelity shape and appearance acquisition, extraction of structure and semantic information, processing, visualization, 3D display and 3D fabrication in professional and consumer applications.

With this fellowship, I was also enrolled as PhD Student in Computer Science Program at the Department of Mathematics and Computer Science in the University of Cagliari under the tutoring of **Prof. Riccardo Scateni**.

My topic was inserted in a research program devoted to ***”Acquisition of complex and structured environments”***, where the goal is to develop a system capable of large-scale acquisition of complex environments by data fusion from mobile visual and depth sensors.

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sk lodowska-Curie Actions Innovative Training Network (MCSA-ITN) grant agreement *No 813170*.

Armando Sánchez
Brescia, Italy
November — 2023.

Chapter 1

Introduction

In various sectors, especially industry and AEC (Architecture, Engineering, and Construction), challenges often revolve around scanning large, structured environments. While aerial images and LiDAR effectively capture outdoor urban areas, mapping complex man-made structures like buildings or diverse outdoor spaces presents unique challenges. These include the need for multiple viewpoints to comprehensively cover areas and issues with accessibility. This chapter describes the motivation for building a mobile mapping system capable of tackling these challenges, a summary of research achievements, and provides an overview of the overall structure of this thesis.

1.1 Background and motivation

In many application domains, in particular industry and AEC, the problems faced are not related to the number of objects, but to the need to scan large, often structured, environments. For the outdoor urban environment case, for example, well-established capture techniques exploit aerial images and LiDAR. Ground capture and indoor capture of complex man-made structures, in particular buildings or outdoor environments, is still a challenging research problem, if only because of the large number of points of view required to cover complex spaces, and of the limited accessibility. Within this thesis and in project EVOCATION as a whole, we have set the goal to develop a state-of-the-art wearable laser+visual scanner system, targeted to capture and map complex man-made structures. The combination of laser and visual input into a small mobile device makes this solution applicable to a variety of applications. In particular, this solution can be used in outdoor and indoor environments or even in tunnels and underground mines, returning dense points clouds with high coverage and a high level of detail. Geometric models alone, however, do not suffice for a variety of needs, such as location awareness for security or guidance, which requires models that are photorealistic enough to recognize real places by just looking at them, object identification, or content creation, which requires visual data. The combination of laser with photographic input provides the color information required for all these applications. Moreover, as photos can be captured at a higher speed and resolution than most laser scanners, combining coarser 3D input with denser photographic input to obtain a better model is also a very interesting research direction.

1.2 Objectives

The primary objectives of this thesis are to advance the mobile mapping field through:

- The development of an integrated mobile 3D mapping system prototype, blending laser scanning with panoramic imaging, to efficiently capture detailed spatial data in diverse settings.
- Utilizing the prototype to create and disseminate open data 3D indoor datasets for research, combining multimodal data from the prototype with ground-truth information from static laser scanners.

These objectives are pursued with a focus on capturing and processing data to create structured datasets, particularly in indoor environments.

1.3 Achievements

The main achievements of this thesis are:

- **Development of an Advanced MMS Prototype:** Successful creation of a state-of-the-art MMS capable of detailed and efficient environmental scanning.
- **Creation of Diverse, Structured Datasets:** Comprehensive datasets have been generated, showcasing the integration of various data types and serving as valuable resources for the research community.
- **Application in AEC Pilot Project:** Demonstrating the practical utility of the developed MMS in real-world AEC scenarios, specifically in construction monitoring and BIM integration.

1.4 Organization

This thesis is structured as follows to present a coherent and comprehensive narrative:

- **Chapter 1**: This chapter sets the stage, introducing the research topic, motivations, objectives, and overarching results.
- **Chapter 2**: Offers a general background, setting the research within the wider context of existing approaches in mobile mapping.
- **Chapter 3**: Delves into the technical development of the mobile mapping system, exploring the integration of laser scanning and panoramic imaging.
- **Chapter 4**: Focuses on the application of the prototype for capturing and processing research-useful datasets, highlighting the fusion of geometric and image data.
- **Chapter 5**: Discusses the deployment of the MMS prototype in an AEC pilot project, underlining its practical applications and implications.
- **Chapter 6**: Concludes the thesis with a summary of achievements, a critical discussion of the results, and contemplation on future research directions.

Chapter 2

General background

This chapter lays the foundational knowledge and contextual landscape essential for understanding the development and application of Mobile Mapping Systems (MMS) in industry and Architecture, Engineering, and Construction (AEC). It provides a comprehensive review of the state-of-the-art technologies, challenges, and advancements in this domain.

2.1 Introduction

The evolution of Mobile Mapping Systems (MMS) has revolutionized the way we capture and interpret spatial data. From the challenges of scanning complex structures to the integration of advanced data processing techniques, this section introduces the core concepts and driving forces behind this transformative technology, highlighting historical milestones in MMS development. This chapter also provides an overview of the main sensor categories that are essential in mobile mapping applications. Finally, this section includes a review of the Simultaneous Localization And Mapping (SLAM) technology that underpins MMS.

Tipically, a MMS platform employs Light Detection And Ranging (LiDAR) and high-resolution cameras for data capture, alongside positioning technologies such as Global Navigation Satellite System (GNSS) for georeferencing and Inertial Measurement Units (IMU) for tracking movements at a smaller scale. To extract valuable information from this data extensive post-processing is usually required, as it will be explained in Sec. 2.3. In Fig. 2.1 it is possible to observe a MMS designed to be mounted on a car, where most of the sensor families mentioned before are clearly distinguishable.

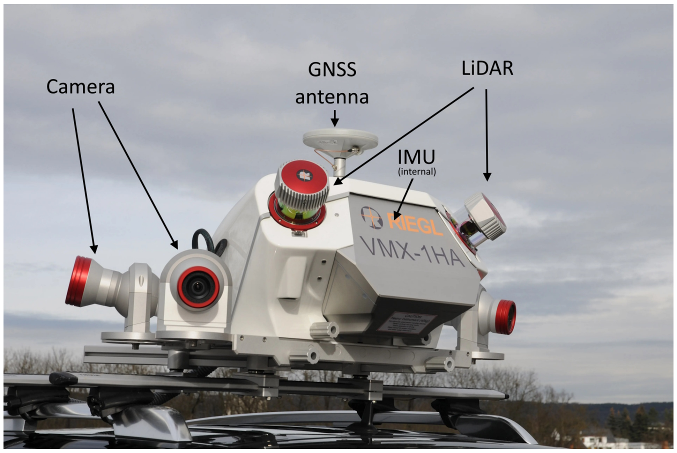


Figure 2.1: *Car-mounted Mobile Mapping System Riegl VMX-1HA [1]*

2.2 Sensors in Mobile Mapping Systems

In Mobile Mapping Systems (MMS), positioning and data collection sensors play pivotal roles. These sensors, such as GNSS, IMU, and DMI (odometers), capture geographical positions and sensor motion to accurately represent 3D data. For enhanced positioning precision, sensor fusion techniques are applied, integrating measurements from various sensors. This fusion is crucial as single sensors like GNSS or IMU/DMI cannot alone ensure reliable navigation data. GNSS signals may vary across environments, with potential signal loss in obstructed areas, whereas IMU and DMI can accumulate errors and typically serve as supplementary data sources. Data collection relies heavily on LiDAR for 3D measurements and digital cameras for colorimetric/spectral data, with advanced imaging techniques enabling additional dense data for 3D fusion. This section delves into these sensor categories and their integration methods for optimized mapping accuracy.

2.2.1 Global Navigation Satellite System (GNSS)

Global Navigation Satellite Systems (GNSS) sensors have revolutionized the way we navigate and map the world. These sensors utilize a constellation of satellites orbiting the Earth to provide geospatial positioning with global coverage. This technology is a cornerstone of Mobile Mapping Systems (MMS), enabling high-precision location data crucial for various applications, from navigation and geodesy to agriculture and autonomous vehicles. GNSS sensors offer several advantages, including wide coverage, high accuracy, and the ability to work in nearly any weather condition. However, they are not without their limitations, such as signal degradation in urban canyons or under dense foliage and dependence on external factors like satellite geometry and atmospheric conditions.

Real-Time Kinematic (RTK) and Real-Time eXtended (RTX) technologies are advanced GNSS correction techniques that significantly enhance the precision of position data. RTK utilizes a fixed base station to provide differential corrections, achieving centimeter-level accuracy. This method is highly effective for applications requiring high precision in real-time, such as surveying, construction, and precision agriculture. However, RTK's effectiveness is limited by its reliance on the proximity to a base station, typically within a few tens of kilometers. In Fig. 2.2 there is an example of a GNSS system compatible with RTK technology.

Figure 2.2: *Trimble DA2 receiver antenna [2], compatible with Trimble Catalyst GNSS positioning service*

RTX technology, on the other hand, extends the possibilities of precision correction through satellite-delivered corrections. It does not require a base station, making it more versatile and suitable for areas lacking RTK infrastructure. RTX can provide sub-meter to centimeter-level accuracy, depending on the service level chosen, and is ideal for remote sensing, asset management, and operations in remote locations.

Choosing between RTK and RTX technologies depends on the specific requirements of the application, including the needed accuracy level, operational environment, and availability of infrastructure. RTK is preferred when the highest precision is needed and a base station is available, while RTX is valuable for broader coverage without the logistical constraints of setting up a base station.

The integration of GNSS sensors into Mobile Mapping Systems has significantly improved the efficiency and accuracy of spatial data collection, driving advancements in many fields. However, users must carefully consider the strengths and weaknesses

2.2.2 Intertial Measurement Units (IMU)

Inertial Measurement Units (IMUs) are key components for Mobile Mapping Systems (MMS), providing critical data for position, orientation, and movement analysis. An IMU typically comprises three sensors: an accelerometer, a gyroscope, and a magnetometer. These sensors work in tandem to measure linear acceleration, angular velocity, and magnetic field, respectively, enabling precise tracking of an object's position and orientation.

Accelerometers measure the rate of change in velocity along an axis, enabling the detection of acceleration and deceleration. **Gyroscopes** assess the rate of rotation around an axis, providing information on orientation changes. **Magnetometers**, detecting magnetic fields, offer compass-like directional insights. Together, these sensors offer a comprehensive understanding of an object's movement in three-dimensional space, crucial for navigation, robotics, and various engineering applications.

The backbone of modern IMUs is Micro-Electro-Mechanical Systems (MEMS) technology, which allows for the miniaturization of mechanical and electro-mechanical elements. MEMS technology has been instrumental in developing compact, cost-effective, and energy-efficient IMUs. These advancements have broadened the application of IMUs, making them integral to consumer electronics, automotive systems, and, notably, Mobile Mapping Systems.

Despite their utility, IMUs have limitations. Their accuracy can degrade over time due to sensor drift, particularly in gyroscopes and accelerometers, leading to cumulative errors. Furthermore, IMUs are sensitive to environmental factors like temperature changes and vibrations, which can affect their performance.

Recent innovations have seen the integration of GNSS data with IMUs to mitigate these weaknesses, enhancing the overall accuracy and reliability of mobile mapping solutions. An example of such advancement is the Xsens MTi-680G, as seen in Fig. 2.3, a state-of-the-art IMU that incorporates GNSS data. This integration allows for real-time kinematic (RTK) positioning, providing high-precision location data. The fusion of GNSS and IMU data enables the MTi-680G to offer continuous and accurate positioning, orientation, and velocity information, even in environments where GNSS signals are weak or unavailable.

The combination of IMU sensors with MEMS technology and GNSS data represents a

Figure 2.3: *MTi-680G XSens [3] Inertial Measurement Unit, with GNSS antenna and serial communication cable*

significant leap forward in mobile mapping and spatial data collection. It exemplifies the ongoing evolution of navigation and positioning technologies, offering enhanced capabilities for a wide range of applications, from autonomous driving to geospatial surveying. However, as with any technology, the selection of an IMU for a specific application must consider the balance between its strengths and the potential limitations posed by sensor drift and environmental susceptibility.

2.2.3 Distance Measuring Instruments (DMI)

Distance Measuring Instruments (DMIs) provide distance and speed data for a variety of mapping and surveying applications. At their core, DMIs measure the distance traveled by a vehicle or device, offering valuable information for calculating areas, mapping routes, and integrating with other sensor data to produce comprehensive geospatial datasets.

The most traditional form of a DMI is the **wheel odometer**, which calculates distance by counting the number of wheel rotations and multiplying it by the wheel's circumference. This simple yet effective mechanism has been utilized in automotive and surveying applications for decades, providing a straightforward means of distance measurement.

Beyond wheel odometers, other sensors and technologies can be classified under DMIs due to their utility in measuring distance. For instance:

- **Laser rangefinders** use the time-of-flight of a laser beam to measure the distance between the sensor and a target, offering high precision over short to medium ranges. This technology is particularly useful in environments where direct contact measurement is impractical or impossible.
- **Ultrasonic sensors** emit ultrasonic waves and measure the time it takes for the echoes to return after bouncing off objects. These sensors are commonly used in robotics and industrial applications for distance measurement and obstacle detection.
- **Infrared (IR) sensors** work on a similar principle to ultrasonic sensors but use infrared light waves. They are often used in consumer electronics and vehicles for proximity sensing and distance measurement.

Despite their utility, DMIs, including wheel odometers, have limitations. The accuracy of a wheel odometer can be affected by factors such as wheel slip, changes in tire pressure, and variations in the terrain, leading to potential errors in distance measurement. Similarly, laser rangefinders, ultrasonic, and IR sensors can be influenced by environmental conditions like atmospheric interference, surface reflectivity, and obstacles that obscure the line of sight to the target.

The integration of DMIs with other technologies, such as GNSS and IMUs, in Mobile Mapping Systems helps to mitigate some of these limitations. By combining data from multiple sources, MMS can provide more accurate and reliable measurements, enhancing the quality of the mapping and surveying outputs.

In the context of modern mobile mapping, DMIs play a role by contributing to the multi-sensor data fusion necessary for creating detailed and accurate geospatial information. Their strengths, such as direct measurement capability and simplicity, make them valuable in certain applications.

2.2.4 Light Detection And Ranging (LiDAR)

LiDAR (Light Detection and Ranging) sensors have emerged as a cornerstone technology in the field of Mobile Mapping Systems (MMS), offering the capability to capture high-resolution, three-dimensional information about the physical world. These sensors emit light pulses and measure the return for each pulse after it bounces off surfaces. This data is then used to generate precise, three-dimensional models of the environment. LiDAR technology is employed across various applications, including autonomous vehicles, forestry management, urban planning, and disaster risk assessment.

Mechanically, LiDAR sensors can be broadly categorized into two types: rotative and solid-state.

Rotative LiDAR sensors, which are more traditional, use mechanically rotating parts to scan the environment. This design, visible in Fig. 2.4, allows for 360-degree coverage and is widely used in applications requiring comprehensive environmental data collection, such as autonomous driving and geographical mapping.

Figure 2.4: *Velodyne [4] VLP-16 rotative LiDAR sensor*

Solid-state LiDAR sensors, on the other hand, have no moving parts. They use optical phased arrays to steer the laser beams. Solid-state LiDARs are more durable, compact, and energy-efficient, making them suitable for integration into smaller devices and vehicles, as seen in Fig. 2.5. However, they may offer a narrower field of view compared to their rotative counterparts.

The wavelength of the laser used in LiDAR systems is a critical factor that influences their operation and application. Common wavelengths include near-infrared (NIR) at around 700 nm to 1550 nm. Shorter wavelengths, closer to 700 nm, are typically used for applications requiring fine detail, as they can achieve higher resolution. The choice of wavelength directly affects the LiDAR's maximum geometric resolution and penetration capabilities, with longer wavelengths generally providing better performance in adverse weather conditions but potentially at the cost of detail.

Figure 2.5: *Velodyne [4] Velarray solid-state LiDAR sensor*

LiDAR sensors measure distance using several principles:

Time of Flight (ToF) involves emitting a laser pulse and measuring the time it takes for the reflection to return. This method is straightforward and widely used but can be influenced by the speed of light's variance in different atmospheric conditions.

Phase Shift measures the phase change in the light wave as it reflects back to the sensor. This technique can offer more precise measurements over short distances and is useful for detailed 3D modeling.

Laser Radar is similar to ToF but uses a continuous beam of light instead of a pulse. The frequency shift in the reflected beam is measured to calculate the distance, providing high accuracy and resolution.

Interferometry involves splitting a light beam into two paths, with one path reflecting off the target and the other remaining within the device. By comparing the phase difference between these two beams upon recombination, highly precise distance measurements can be made, useful in scientific and engineering applications requiring nanometer-level precision.

Each LiDAR technology and measurement principle has its strengths and weaknesses. For instance, ToF LiDAR systems are versatile and widely applicable but can struggle with accuracy over long distances or in challenging environmental conditions. Phase Shift and Interferometry offer higher precision but are more complex and costly to implement.

In summary, LiDAR sensors are essential to Mobile Mapping Systems because they provide detailed and accurate 3D data. The choice between different LiDAR technologies, wavelengths, and measurement principles depends on the specific requirements of the application, including the needed resolution, range, and environmental conditions. Understanding these factors is crucial for leveraging LiDAR technology effectively within the diverse landscape of mobile mapping and geospatial data collection.

2.2.5 Cameras and other Imaging Devices

Cameras and other imaging devices play a crucial role in the ecosystem of Mobile Mapping Systems (MMS), offering visual context and detailed imagery that complement data collected by other sensors like LiDAR and GNSS. These devices capture high-resolution photographs and videos, enabling the creation of detailed maps, 3D models, and virtual environments. Their utility spans across various domains, including urban planning, disaster management, environmental monitoring, autonomous vehicle navigation, and object identification, to name a few.

Visible light spectrum high-resolution digital cameras capture detailed still images, which are essential for creating high-definition maps and photogrammetric models. These cameras can vary from consumer-grade to professional-grade, with the latter offering superior image quality, higher resolutions, and better performance in low-light conditions. They also can provide continuous footage of the environment, useful for dynamic mapping applications such as monitoring traffic flow or pedestrian movements. They are also instrumental in the development and testing of autonomous vehicles, where understanding the environment's dynamic nature is crucial.

Multispectral and Hyperspectral Cameras, by capturing images across multiple wavelengths beyond the visible spectrum, they can detect various materials, vegetation types, and other environmental characteristics. This capability is especially beneficial in agriculture for soil health assessments, in environmental monitoring for assessing water quality, and in urban planning for mapping and analyzing land use. More specifically, cameras sensible to the infrared spectrum can capture images based on infrared radiation, providing valuable data for environmental monitoring, energy auditing, and even agricultural applications by revealing heat distributions and plant health that are not visible in the standard visual spectrum.

Strengths of Cameras and Imaging Devices

- Visual Detail and Context: They provide a visual record of the environment, invaluable for interpretation, analysis, and decision-making in many applications.

- Versatility: The wide range of camera types and imaging technologies available allows tailoring mapping solutions to specific needs and applications.
- Accessibility: Technological advances have made high-quality cameras more accessible and affordable, enabling their widespread use in mobile mapping systems.

Weaknesses of Cameras and Imaging Devices

- Lighting Conditions: Their performance can be significantly affected by lighting conditions. Overexposure or underexposure can lead to loss of detail, and nighttime or low-light conditions pose a particular challenge. They are passive sensors, incapable of generating its own light, as opposed to LiDAR sensors.
- Weather Conditions: Adverse weather conditions, such as rain, fog, or snow, can obscure the captured images, reducing the quality and accuracy of the data collected.
- Data Volume: High-resolution images and videos generate large amounts of data, posing challenges for storage, processing, and analysis.

In conclusion, cameras and other imaging devices are indispensable to the functionality and success of Mobile Mapping Systems. Their ability to capture detailed visual information adds a layer of richness to the data collected, enhancing the accuracy and applicability of mapping outputs. However, to fully leverage their benefits while mitigating their limitations, it is crucial to integrate these devices with other sensors and technologies, ensuring a comprehensive and multifaceted approach to mobile mapping and spatial data collection.

2.3 Simultaneous Localization And Mapping (SLAM)

Simultaneous Localization and Mapping (SLAM) is the engine that powers most approaches in the domain of Mobile Mapping Systems (MMS), enabling devices to construct a map of an unknown environment while simultaneously determining their location within that space. This dual capability is critical for autonomous vehicles, robotics, augmented reality (AR), and various surveying applications, where understanding and navigating an environment in real-time is essential.

Core Aspects of SLAM Technology

SLAM technology integrates data from various sensors, including cameras, LiDAR, IMUs, and GNSS, to perform its dual tasks. It employs complex algorithms to process and fuse this sensor data, creating a detailed spatial map of the environment while tracking the sensor's or vehicle's position and orientation within that map. This process involves several steps, including feature extraction, data association, state estimation, and map updating, which work together to build and refine the map dynamically as new data is collected [5, 6].

Strengths of SLAM Technology

- Autonomy in Unknown Environments: SLAM enables devices to operate in environments where maps are not available or GPS signals are weak or nonexistent, such as indoors or in dense urban areas.
- Real-time Operation: It provides the capability to perceive and navigate the environment in real-time, crucial for applications like autonomous driving and robotic navigation.
- Flexibility: SLAM can be implemented using various sensor configurations, making it adaptable to different applications and budget constraints.

Weaknesses of SLAM Technology

- Computational Requirements: The complex algorithms and the need for real-time processing of large volumes of sensor data make SLAM computationally demanding, requiring powerful hardware.
- Sensor Noise and Uncertainty: The accuracy of SLAM is dependent on the quality and calibration of the sensors used. Noise and errors in sensor data can lead to inaccuracies in localization and mapping.
- Dynamic Environments: SLAM systems can struggle in highly dynamic environments where the scene changes frequently, as this can complicate the process of tracking features over time.

In summary, SLAM technology is a pivotal innovation in Mobile Mapping Systems, offering the ability to navigate and map uncharted environments with a high degree of autonomy and precision. While it presents challenges in terms of computational demands and sensor accuracy, ongoing advancements in algorithms, sensor technology, and computational hardware are continually expanding its capabilities and applications. The integration of SLAM into various domains not only enhances operational efficiency and safety but also paves the way for new possibilities in exploration, automation, and spatial interaction.

2.4 Mobile Mapping Systems platforms

Mobile Mapping Systems can be classified into four main configurations: handheld, backpack, trolley, and vehicle-mounted systems. Each configuration has its unique advantages and disadvantages, making them suitable for different use cases.

Handheld Configuration

Advantages:

- Portability: Handheld MMSs are highly portable, ideal for mapping confined or hard-to-reach areas.
- Ease of Use: They require minimal setup time and can be operated with little to no training, offering convenience and accessibility.
- Cost-effective: Generally, handheld systems are less expensive than more complex setups, offering a cost-effective solution for small-scale projects.

Disadvantages:

- Limited Range and Accuracy: Handheld devices may offer lower accuracy and range compared to more robust systems, affecting the precision of the collected data.
- User Dependence: The quality of data can be highly dependent on the operator, leading to inconsistencies.

Use Cases: Handheld MMSs are suited for indoor mapping, architectural surveys, and archaeological site documentation, where portability and ease of use are paramount.

Backpack Configuration

Advantages:

- Mobility and Accessibility: Backpack systems allow for easy navigation through crowded or obstructed environments, such as urban areas or forests.
- Hands-free Operation: They enable the operator to use their hands for other tasks, enhancing safety and efficiency.

- Enhanced Data Collection: Equipped with advanced sensors, backpack MMPs can collect detailed data over large areas without the need for direct line of sight.

Disadvantages:

- Physical Burden: Carrying a backpack system for extended periods can be physically demanding.
- Cost: These systems can be more expensive than handheld configurations due to the inclusion of more sophisticated technology.

Use Cases: Backpack MMSs are ideal for extensive pedestrian area mapping, disaster assessment in inaccessible areas, and large-scale indoor mapping projects.

Trolley Configuration

Advantages:

- Stability and Precision: Trolley systems offer stable platforms for high-precision sensors, improving data accuracy.
- Ease of Maneuverability: Designed for smooth surfaces, trolleys can be easily maneuvered in urban and indoor environments.
- Extended Battery Life: They often feature larger batteries or power solutions, allowing for longer operation times.

Disadvantages:

- Limited Terrain Accessibility: Trolleys are not well-suited for rough or uneven terrain.
- Transport and Setup: These systems can be cumbersome to transport and set up, requiring more time and effort than more portable configurations.

Use Cases: Trolley MMSs excel in detailed mapping of streets, sidewalks, and indoor spaces like malls and museums, where precision is critical, and surfaces are uniform.

Vehicle-Mounted Systems

Advantages:

- Large-Scale Mapping Capability: Vehicle-mounted MMSs can cover extensive areas quickly, making them perfect for city-wide mapping projects and road surveys.

- High-Speed Data Collection: These systems can collect data at high speeds, reducing the time and cost associated with large-scale mapping efforts.
- Integration of Multiple Sensors: Vehicles can carry a wide array of sensors, including LiDAR, cameras, and GNSS, enabling the collection of comprehensive datasets.

Disadvantages:

- High Cost: The most expensive of all configurations, vehicle-mounted systems require significant investment in equipment and maintenance.
- Limited Access: Vehicles cannot access narrow or indoor spaces, limiting the scope of possible mapping projects.

Use Cases: Vehicle-mounted MMPs are used for highway planning, urban planning, and large-scale environmental monitoring, where speed and the ability to cover large distances are crucial.

Each configuration of Mobile Mapping Platforms offers distinct advantages and poses unique challenges, making them suitable for a wide range of applications. The choice of system depends on the specific requirements of the project, including the desired accuracy, coverage area, and budget constraints. Some existing solutions are shown in Fig. 2.6 for lightweight portable MMSs, and in Fig. 2.1 for a vehicle-mounted solution.

Figure 2.6: *Existing Mobile Mapping Systems. From left to right: Gexcel Heron [7], GeoSLAM Zeb Horizon [8], Navvis VLX [9], LiBackpack DGC50 [10]*

2.5 Advancements and Applications of Mobile Mapping Systems

Mobile Mapping Systems (MMS) have increasingly become essential in various domains, notably in construction progress monitoring, architectural preservation, and the creation of digital twins. Recent studies have shed light on the diverse capabilities and applications of MMS, highlighting their transformative impact in these fields.

2.5.1 MMS in Construction Progress Monitoring

The integration of MMS with Building Information Modeling (BIM) for construction progress monitoring represents a significant leap in construction management. There are significant ongoing efforts exploring how an indoor MMS (iMMS) can be utilized for tracking construction progress [11, 12]. These studies demonstrate that iMMS could produce high-quality point clouds suitable for comparing as-built models with as-designed BIM, thereby enabling efficient construction progress monitoring. Innovative methods combining BIM with iMMS for construction progress evaluation also approach used periodic geometric surveying using iMMS to produce point clouds, which then are compared with 4D BIM models, effectively digitizing monitoring processes and aligning them closely with BIM schedules.

2.5.2 MMS in Architectural Preservation

MMS technology has also shown promising results in the preservation of historical sites [13] by conducting a comparative study using a wearable MMS and a photogrammetric multi-camera prototype to create the survey. Research in this direction highligths the effectiveness of MMS in providing comprehensive coverage and detailed geometric descriptions of trees and landscapes, outperforming traditional photogrammetric approaches in certain aspects.

2.5.3 MMS in Digital Twin Creation

The creation of digital twins for large buildings and sites is another area where MMS demonstrates its versatility [14]. The use of iMMS in creating digital twins for social housing real estate outlines the technical and management methodologies involved in field data acquisition and the subsequent processing workflow. This application of mobile mapping systems in creating digital twins represents a significant advancement in asset management and urban planning.

2.6 Conclusion

The studies presented in these papers collectively underscore the crucial role of MMS in diverse fields. They highlight not only the technological advancements in MMS but also their practical implications in improving the efficiency, accuracy, and comprehensiveness of spatial data collection and analysis. The integration of MMS with BIM, its application in preserving cultural heritage, and its role in creating digital twins all point towards a future where mobile mapping plays a central role in spatial data handling and decision-making processes in architecture, engineering, and construction.

Chapter 3

Multimodal mobile capture of large environments

In an era where the precision and efficiency of spatial data acquisition are paramount, the development of an integrated mobile 3D mapping system stands at the forefront of technological innovation. This chapter delves into the intricacies surrounding the development of such a system, a pivotal component of modern spatial data acquisition and analysis in large environments. The work presented here is particularly pertinent to the Architecture, Engineering, and Construction (AEC) fields, where such technology plays a critical role in accurate and efficient data gathering. This chapter provides an in-depth look into the creation and deployment of such a system, highlighting the challenges and breakthroughs encountered in its development.

3.1 Introduction

Building acquisition and reconstruction pipelines for image and 3D data requires the design of a system that is capable of capturing large and complex environments through visual and depth sensors. Using the data acquired from said sensors, it is possible to recover structural and functional information about the captured models.

Making this whole system wearable and mobile adds a whole new degree of complexity, introducing challenges regarding, size, power consumption, and sensor fusion. The standard integrated mobile mapping system consists of a spherical camera, one or more LiDAR sensors, and an Inertial Measurement Unit (IMU). These systems are designed in such a way that they can be carried by one person, either handheld or as a backpack. While the user moves through the environment, the device performs the data acquisition and some basic processing on the fly. Usually, in a later post-processing stage, all the captured data is processed in order to recreate the trajectory followed during the data acquisition and the environment, with as much precision as possible.

In the case of mobile mapping systems, a Simultaneous Localization And Mapping (SLAM) algorithm plays a central role in this stage, ideally taking advantage of the three available sources of data: image, depth, and inertial. Performing data fusion with the stream of data coming from these three sensors is essential since the same parameter (i.e. Position) can be extracted from multiple sensors. The reliability of these sensors when giving an accurate value for said parameter can vary wildly depending on the external conditions. For example, LiDAR sensors can operate in total darkness because they act as their own light source (infrared laser), while cameras rely on external light sources. IMUs perform excellently in a local environment, while globally they present a large drift that must be corrected through the data collected by other sensors (loop closures).

3.2 Development of a prototype integrated mobile 3D mapping system

To overcome current limitations in the current acquisition approaches, proposed solutions in the design of a Mobile Mapping System (MMS) should be characterized by three main features:

- Flexibility: The solution must be generalized and adaptable to wildly different acquisition scenarios. A common pain of SLAM-based mapping systems is that they work very well in specific places, while they break under different conditions.
- Scalability: To offer guarantees about the optimal acquisition of datasets in a time-efficient way.
- Minimum redundancy: A massive amount of data per unit of time is acquired using an MMS. Most of it is redundant, which can be circumvented with proper filtering techniques.

The first step is building a hardware system that is capable of capturing raw data from visual, LiDAR, and inertial sensors in a robust and reliable way.

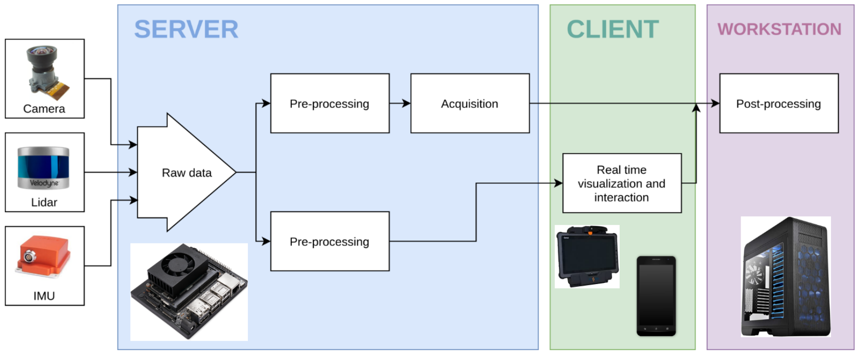


Figure 3.1: *Mobile 3D Mapping System pipeline*

An overview of the mobile 3D mapping system is shown in Fig. 3.1. At first, raw data coming from all sensors is collected and pre-processed in two separate pipelines. The first pipeline is the more robust, dedicated exclusively to data storage for a later post-processing stage. In this case, the pre-processing is minimal, mainly dedicated to guarantee data integrity. On the other hand, the second pipeline is focused in real-time visualization. In this second case, the data must be more heavily processed, involving, among other requirements, the conversion from polar to cartesian coordinates of all LiDAR points, the 3D rendering of the point cloud (it can be sub-sampled if necessary) and a lightweight SLAM algorithm capable of aligning several point clouds in a constrained time window.

The main hardware components in its current version are:

- Sensors
  - Camera: Customized spherical camera Labpano [15] Gexcel MG-1. It offers both 8K stitched and unstitched image data in jpg format, and 4K 24 fps video with h264/h265 codec. RTSP/RTMP video streaming protocols and NTP synchronization with the local system.
  - LiDAR:
    * Velodyne [4] VLP-16 and/or VLP-32. Available signals: Azimuth angle, distance, reflectivity, timestamp, status. Communication through Ethernet UDP packets. An array of 16/32 NIR Laser photodiodes, spinning at 300-600 rpm.
    * Hesai [16] XT16, XT32 and/or XT32M2X. Same signals are available compared to the Velodyne. However, the XT family of sensors from

Hesai offers a sixfold improvement in accuracy, as well as increased maximum range, minimum range of 0, and angle-based triggering of the laser photodiodes, as opposed to time-based.

  - Inertial Measuerement Unit (IMU): Xsens [3] MTi-30 AHRS or MTi-630. Accelerometer, gyroscope, and magnetometer. Xbus communication protocol. Signals: Temperature, orientation, linear acceleration, angular velocity, magnetic field, timestamp, status.

- System on a Chip (SoC):

  - Nvidia Jetson [17] Xavier NX. ARM 6-core CPU + 384-core Nvidia Pascal GPU + 48 Tensor cores. NVDLA Deep Learning accelerator chip and Multimedia Complex with hardware encoding/decoding for image and video. 8GB of shared DRAM memory.

  - Asrock [18] NUC BOX-1115G4. Intel 11th Gen i3 2-core CPU + Intel UHD Graphics. 8GB of Dual Channel DDR4 3200 MHz memory.

Multiple challenges have been faced during the continued development of the mobile data acquisition pipeline, such as an ever-increasing complexity of the system when adding support for new sensors, asynchronous execution of all the software components, and limited computational resources, all of which without compromising neither flexibility nor robustness. Taking all that into consideration, the last iteration of the data acquisition system was designed with the following goals in mind:

- Cross-platform support.

- Performance-oriented. A MMS has limited energy and computational resources.

- Robustness. The system is meant to be used by professionals who need to perform a 3D survey in environments as diverse as AEC, heavy industry, mining operations... A system failure leads to a significant loss of time and resources, especially when the survey requires a difficult expedition.

- Flexibility: It is of utmost importance that it's easy and fast to add new functionalities and support new hardware, while also creating a favorable environment for having access to image and geometric data streams. This makes life easier for whoever is interested in building on top of our platform.

With all these ideas in mind, the data-capturing software system is built with the following characteristics:

- Backend/firmware: Asynchronous Rust [19]. While being a fairly new programming language with limited support in some areas, Rust otherwise aligns perfectly with the established goals of flexibility, robustness, and performance.

- Frontend: Dart [20] + Flutter [21]. Inherently cross-platform. It enables the development of client applications for the Web, Android, Windows, Linux. . .

- OS: Yocto project [22]. A customized Linux operating system, fine-tailored to fit the exact needs of the system, without unnecessary bloat.

The firmware from the device is built around the following main blocks:

- Capture Manager: the central process of the capture system. It keeps the knowledge of the expected system configuration (i.e. which kind and which number of sensor processes to expect), keeps track of each sensor's health, and acts as a service discovery interface for clients.

- Supervisor: it handles the following tasks:

  - Communication with the OS

  - Management of storage devices

  - Update system

  - Launching of all processes except itself

  - I/O of system configuration and calibration files

- HTTP Server: implemented through nginx. Only serves static files for frontend applications.

- 1 or more clients

- 1 or more Sensor Processes, which share a common communication protocol with the capture manager

All of the components above share a common communication protocol based on messages. This communication is implemented using grpc [23] and the protocol buffers [24] library. The main goal with the choice of this architecture is to create layers of abstraction from the physical devices (LiDAR sensor, IMU, cameras. . .) to the user application. For this reason, a sensor process must handle all the details regarding the interaction with a specific device. On its interface, a sensor process exposes its capabilities as a set of services or abstractions over the capabilities of the specific device. For example, a camera sensor process could expose: write to file, data stream and single shot.

Following this abstraction, a client can ask the capture manager to activate one or more services, without having knowledge of the specific sensor that is connected to the capture box. To simplify some interactions of the client with the capture box firmware, the capture manager also implements an additional layer of abstraction over the services: it defines a virtual sensor, which is intended to collect a set of services that are provided by the sensor processes and expose them as a single service. In Fig. 3.2 it is shown how 3 of the main processes families (Client, Capture manager and Sensor) interact with each other through grpc messages.

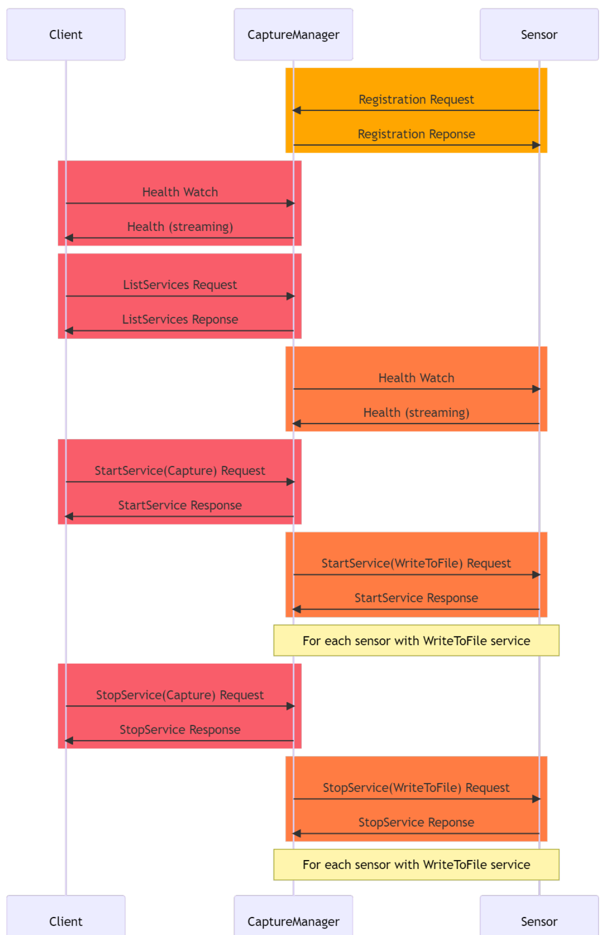


Figure 3.2: *Firmware grpc communications sequence diagram*

3.3 Conclusions

In summary, the main goals set within EVOCATION regarding the implementation of a prototype integrated mobile 3D mapping system have been achieved. The system combines inputs from different sensors such as LiDAR, RGB panoramic cameras, and IMUs, and is able to generate useful data that can be used both in industry or in research, as demonstrated in Chapter 5. The future improvements roadmap includes:

- Implementation of a guided procedure for configuring the system by adding sensors and extracting metadata from them (IP address, serial number) through a guided procedure. This is already in an advanced stage, being successfully used in the last batch of Labpano Gexcel MG-1 cameras received.
- Implement service to create scans from raw data: this service should gather data from sensors through streaming services and implement a system to generate scans (full rotations). This would allow the development of a lightweight 3D viewer client as well as allowing smart subsampling of data to be streamed to clients. A repurposed version of the Depth Maps renderer described in Chapter 4 is being tested for this purpose.
- Implementation of ICP / SLAM on board by exploiting hardware acceleration (GPU) as well as researching deep learning approaches to process or enhance data by exploiting hardware acceleration (Tensor cores, DL engines). This has been put to a halt due to the progressive switch from systems based on Nvidia Jetson SoCs to Intel NUCs, where a purely CPU-based approach is preferred.

3.4 Bibliographic notes

Most of the content of this chapter was presented in the EVOCATION Deliverable D2.5 titled "New technologies for high-level structured reconstruction with visual and depth sensing" [25].

Chapter 4

Exploiting data fusion for deep panoramic depth prediction and completion for indoor scenes

The evolution of 3D indoor mapping technologies has reached a pivotal juncture with the advent of deep learning and advanced data fusion techniques. This chapter outlines the journey of creating an innovative approach to generate datasets that facilitate panoramic depth prediction and completion in indoor environments. We discuss the challenges of integrating multimodal data and the implications of these advancements for indoor 3D mapping. Our work, rooted in the collaborative efforts between CRS4 and Gexcel, demonstrates the potential of these technologies in transforming our understanding and interaction with indoor spaces.

4.1 Introduction

Collaborative work between CRS4 and Gexcel has opened the possibility to develop acquisition methodologies for complex indoor environments and the construction, starting from geometric and visual data, of a dataset suitable both for testing multimodal reconstruction methods and for the use in modern deep learning networks. While there is now ample availability of indoor datasets based on single views, there are no such datasets that combine, at the same time, laser data with panoramic video or image sequences to support more complex and dense reconstructions. Panoramic and spherical capture, both visual and geometric, is a consolidated technique in the acquisition of indoor environments, thanks to the wide coverage of the environment and its context [26], most of the real and synthetic datasets available cover single room environments or more complex environments with scattered images. On the other hand, modern techniques of data fusion [27] and reconstruction of larger and more complex models are based on conventional RGB-D perspective video sequences [28, 29, 30, 31].

Starting from this background, the goal was to design an efficient and scalable acquisition technique for this type of panoramic dataset, and had, as a demonstrative result, the release of an innovative dataset of this type, Indoor3Dmapping [32]. To the best of our knowledge, ours is the first prototype of such a kind of dataset, where large, commonly used datasets provide scene coverage by panoramic images only from a limited number of predefined views [33, 34, 35].

4.2 3D Indoor datasets for research

The dataset introduced in Sec. 4.1 has been generated from the survey of the environment with the prototype wearable mobile laser scanner coupled with a spherical camera already presented in Chapter 3. This survey is supported by a capture of the same area with a high-resolution Terrestrial Laser Scanner (TLS). This task is directly relevant to flexible and massive 3D acquisition and registration of multimodal data [25], as well as to the Pilot project described in Chapter 5, which focuses on indoor mapping from AEC. The dataset will serve as a basis for developing and testing advanced techniques for high-level structured reconstruction, in particular, to support novel deep learning pipelines based on visual, pure geometric, or hybrid data. Of particular importance, in this case, is the availability of the full TLS capture to serve as ground truth. Moreover, the dataset will be useful for testing algorithms and systems that focus, respectively, on scalable visualization techniques for captured datasets and scalable navigation techniques for massive data. We expect in the future that the extension of this type of dataset may support the use of modern dense fusion techniques [36, 37], currently based on RGB-D perspective video, to modern panoramic capture of large indoor environments [38].

Dataset overview

Two distinct datasets are generated. The first one consists of both the sparse mobile point cloud and the dense TLS point cloud in an open format, the RGB panoramic images, and the set of poses from where the images were captured. The second dataset contains the same file with the poses, the RGB images, and two sets of depth images: mobile and TLS; generated from the point of view of every individual RGB image.

It is required that both datasets are clean and complete and that no specific software is required for their use. All the information required for generating the second dataset is present in the first dataset in such a way that different versions of the depth images can be generated from scratch. It is also useful for prototyping different versions of any algorithm of interest.

4.3 Indoor3Ddataset

It involves a set of RGB images, two sets of depth images (mobile and TLS), generated from the point of view of each individual RGB image, and a set of poses from which the panoramic images were captured. A purpose-built rendering application, which inner workings are drafted in Fig. 4.1, is used to load a multi-layer point cloud file, a trajectory file, a compounded images file, and an images index file. These inputs are processed to produce a set of paired depth and RGB images. The specific inputs are as follows:

- **Point cloud:** Multi-layered point cloud file.
  - **TLS:** Range, Color, Reflectance, Inclination, Confidence.
  - **Mobile:** Range, Timestamp, Color, Confidence, Reflectance, Inclination, Radius.
- **Trajectory:** Discrete set of timestamped poses.
- **Images:** Set of JPEG files.
- **Images index:** Start offset, end offset, timestamp.

The settings that can be selected when rendering with this tool are:

- **Maximum rendering distance:** Points that are further away than a certain distance will have a value equal to $0xFFFF$.
- **Iteration distance:** Images are captured at a frequency of 15 Hz, regardless of the walked distance. This setting allows for the output of an RGB + depth image pair every time a certain distance threshold is traversed (1 meter by default).
- **Rendering resolution:** It is possible to render the depth images at any resolution, although very high resolutions are discouraged due to file size and augmented data sparsity. The default resolution is 1920x960.

- **Time window:** For the point clouds that have a timestamp layer, it is possible to render the depth images within a constrained time window centered around the timestamp of the associated RGB image. Unbounded by default.

The specific settings used for Chapter 5 are represented in Table 4.1

Max. render distance (m)	Iter. distance (m)	Render resolution (px)	Time window (ms)
16	0.3	1024 x 512	± 150

Table 4.1: *Specific settings used for the rendering application.*

Since the trajectory contained in the raw trajectory file is discretized and does not match the poses of the images, an iteration process is necessary. An interpolation process based on Lie algebra is used, as it reliably produces smooth and realistic-looking trajectories when interpolating between rigid motions (combined translation and rotation) in 3D. After this iteration process is completed, the pose for every image in the dataset is available. Now, the depth map rendering sequence can finally be initiated for both the sparse and dense point clouds:

- For every pose, each individual point is projected onto the surface of a unit sphere, resulting in a value in normalized spherical coordinates. A radius value of $0xFFFF$ is assigned to points beyond the maximum rendering distance specified in the settings, and those points are discarded.

- Conversion from normalized spherical coordinates to image coordinates. If a pixel with a higher value (closer to POV) was previously recorded, the pixel is discarded due to an occlusion.

- The resulting depth values are encoded as 16-bit grayscale PNG images, producing the final depth images.

The flowchart in Fig. 4.1 illustrates how each pixel of the final depth image is drawn. The final result of all this process is a set of 3 aligned images that can be seen in Fig. 4.2: RGB (Fig. 4.2a), dense depth map from the TLS point cloud (Fig. 4.2b) and sparse depth map from the mobile mapping system point cloud (Fig. 4.2c).

4.3.1 Dataset A: Organization

Under the root directory for the whole acquisition there is a positions.csv file and 3 subdirectories: img, dense and sparse.

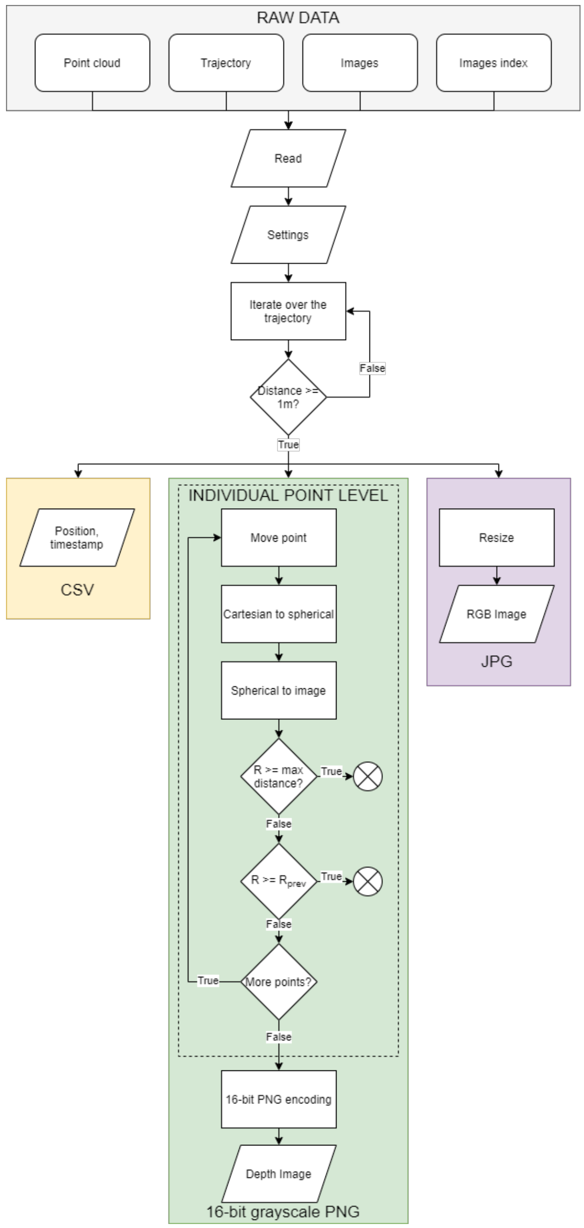
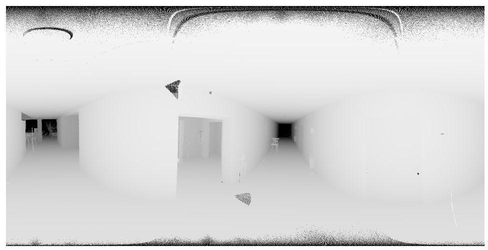


Figure 4.1: *Rendering flowchart*

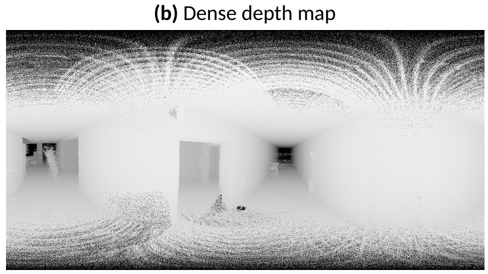
positions.csv

- File format: One ASCII file.

(a) RGB Image



(b) Dense depth map



(c) Sparse depth map

Figure 4.2: *3-way comparison of the data present in Indoor3Dmapping [32].*

- File structure Rows: Each image is one record.
- File structure Columns: Comma-separated headers, with exact order described below.
  - Filename, column 0: Panorama file name as on disk, without the file extension.
  - Timestamps, column 1: Absolute time at which the panorama was captured, Decimal notation, without thousands separator (microseconds).
  - X,Y,Z, columns 2 through 4: Position of the panoramic camera in decimal notation, without thousands separator (meters).
  - w,x,y,z, columns 5 through 8: Rotation of the camera, quaternion.

```
acquisition
├── img
│   ├── <positionID>.jpg
│   └── ⋮
├── dense
│   └── TLS.las
├── sparse
│   └── MOBILE.las
└── positions.csv
```

Figure 4.3: *Dataset A: Folder Tree.*

sparse

MOBILE.las unstructured pointcloud file, following LAS Specification v.1.4 [39], Point Data Record Format 3. Contrary to what is described in the LAS Specification, the 8 bytes GPS Time field is a floating point value representing number of seconds elapsed since 2014-12-31 23:00 UTC.

dense

TLS.las unstructured pointcloud file, following LAS Specification [39] v.1.4, Point Data Record Format 2.

img

A set of equirectangular panoramic images taken with a 360° color camera in 1920x960 resolution. They follow the same trajectory.

4.3.2 Dataset B Organization

Under the root directory for the whole acquisition there is a positions.csv file and 3 subdirectories: img, dense and sparse.

positions.csv

- File format: One ASCII file.
- File structure Rows: Each image is one record.

Item	Format	Size	Required
X	long	4 bytes	yes
Y	long	4 bytes	yes
Z	long	4 bytes	yes
Intensity	unsigned short	2 bytes	no
Return Number	3 bits (bits 0-2)	3 bits	yes
Number of Returns (Given Pulse)	3 bits (bits 3-5)	3 bits	yes
Scan Direction Flag	1 bit (bit 6)	1 bit	yes
Edge of Flight Line	1 bit (bit 7)	1 bit	yes
Classification	unsigned char	1 byte	yes
Scan Angle Rank (-90 to +90) – Left Side	signed char	1 byte	yes
User Data	unsigned char	1 byte	no
Point Source ID	unsigned short	2 bytes	yes
GPS Time	double	8 bytes	yes
Red	unsigned short	2 bytes	yes
Green	unsigned short	2 bytes	yes
Blue	unsigned short	2 bytes	yes
Minimum PDRF Size		34 bytes	

Table 4.2: *LAS Specification v.1.4 [39], Point Data Record Format 3.*

```
acquisition
├── img
│   ├── <positionID>.jpg
│   └── ⋮
├── dense
│   ├── <positionID>.png
│   └── ⋮
├── sparse
│   ├── <positionID>.png
│   └── ⋮
└── positions.csv
```

Figure 4.4: *Dataset 2 folder tree.*

Item	Format	Size	Required
X	long	4 bytes	yes
Y	long	4 bytes	yes
Z	long	4 bytes	yes
Intensity	unsigned short	2 bytes	no
Return Number	3 bits (bits 0-2)	3 bits	yes
Number of Returns (Given Pulse)	3 bits (bits 3-5)	3 bits	yes
Scan Direction Flag	1 bit (bit 6)	1 bit	yes
Edge of Flight Line	1 bit (bit 7)	1 bit	yes
Classification	unsigned char	1 byte	yes
Scan Angle Rank (-90 to +90) – Left Side	signed char	1 byte	yes
User Data	unsigned char	1 byte	no
Point Source ID	unsigned short	2 bytes	yes
Red	unsigned short	2 bytes	yes
Green	unsigned short	2 bytes	yes
Blue	unsigned short	2 bytes	yes
Minimum PDRF Size		26 bytes	

Table 4.3: *LAS Specification v.1.4 [39], Point Data Record Format 2.*

- File structure Columns: Comma-separated headers, with exact order described below.
  - Filename, column 0: Panorama file name as on disk, without the file extension.
  - Timestamps, column 1: Absolute time at which the panorama was captured, Decimal notation, without thousands separator (microseconds).
  - X,Y,Z, columns 2 through 4: Position of the panoramic camera in decimal notation, without thousands separator (meters).
  - w,x,y,z, columns 5 through 8: Rotation of the camera, quaternion.

sparse

- Set of equirectangular rendered depth images.
  - 1920x960 resolution
  - 16-bit grayscale PNG

- White $\rightarrow$ 0 m
- Black $\rightarrow$ $\geq$16 m or absent geometry
- Occlusions: If a pixel was hit by several rays, only the value of the closest one is represented.

dense

- Set of equirectangular rendered depth images.
  - 1920x960 resolution
  - 16-bit grayscale PNG
  - White $\rightarrow$ 0 m
  - Black $\rightarrow$ $\geq$16 m or absent geometry
  - Occlusions: If a pixel was hit by several rays, only the value of the closest one is represented.

img

A set of equirectangular panoramic images taken with a 360° color camera in 1920x960 resolution. They follow the same trajectory.

4.3.3 Acquisition and setup techniques

Two very distinct setups and techniques were required in the preparation of the complete dataset. On the one hand, for obtaining the dense point clouds, a Faro Focus3D X 330 Terrestrial Laser Scanner was used (Fig. 4.5).

The procedure for creating a multiroom indoor scan based on this kind of scanner requires choosing a series of fixed positions within the area that needs to be scanned with the goal of maximizing coverage and minimizing dead angles. Scanning from positions where there is a transition from one room to the next one is highly desirable (such as below doorways) since it helps in the process of stitching together a global point cloud. The set of fixed positions chosen for our dataset can be seen in Fig. 4.6.

On the other hand, for the generation of the sparse point cloud data, a sensor head from an Heron MS Twin Color Mobile Mapping System was used (Fig. 4.7), connected to the prototype capture system described in Chapter 3. The point cloud scanning capabilities of this instrument are provided by two Velodyne Puck

Figure 4.5: *Faro Focus3D X 330 TLS*

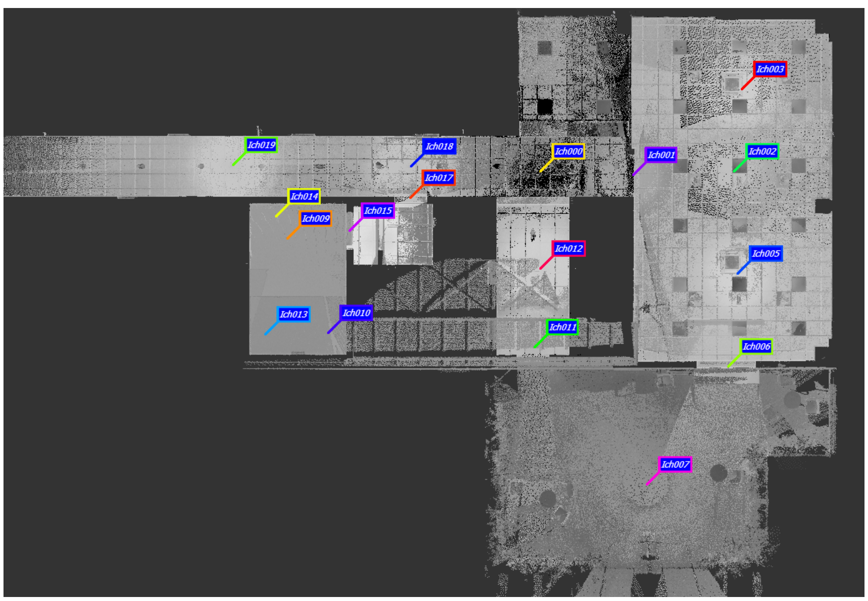


Figure 4.6: *17 distinct positions for the static TLS scans*

sensors, one aligned with the horizontal plane and the other one tilted 45°. On top of that, it has a camera that can capture 5K images in single shot mode and FullHD (1920x1080) in continuous acquisition mode at 15 frames per second.

A SLAM algorithm is responsible for aligning one point cloud to the next one while

Figure 4.7: *Heron [7] MS Twin Color Mobile Mapping System*

the user is walking and wearing the Heron backpack MMS. This SLAM algorithm makes use of an inertial sensor also present in the instrument. In Fig. 4.8 there is a representation of the trajectory that was followed during the acquisition of the mobile sparse point cloud data.

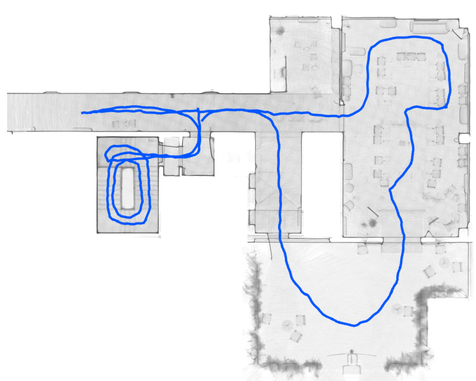


Figure 4.8: *Trajectory followed by the mobile capture device*

4.4 Conclusions

Indoor3Dmapping dataset [32] is already available and ready to be used as it is. However, there is room for improvement. For example, by taking advantage of the already developed tools and methodologies, it would be interesting to expand it into a comprehensible family of datasets where many different scenarios are available to the public, with their corresponding descriptions. There is also the possibility of, not only widening the selection of different environments, but also the characteristics of the sparse LiDAR depth maps, simulating situations in which different slices of information are available in each image file, such as variations in the maximum and minimum rendering distance, and constraining the time window of the rays that are represented in the image. It is worth mentioning that all the data that is being acquired and processed for Chapter 5 opens a new opportunity in this regard since it can potentially be repurposed for the creation of new publicly available datasets.

4.5 Bibliographic notes

Most of the content of this chapter was presented in the EVOCATION Deliverable D2.5 titled "New technologies for high-level structured reconstruction with visual and depth sensing" [25].

Chapter 5

Indoor mapping for AEC pilot

This thesis was conducted within the framework of the European Union's H2020 research and innovation program, under grant 813170 (EVOCATION). As part of this project, we executed a pilot study to assess the applicability of our research to real-world scenarios in the Architecture, Engineering, and Construction (AEC) sectors. Specifically, I was involved in data acquisition for the pilot using Gexcel's prototype mobile mapping system. Additionally, I participated in the processing of the captured information, producing comprehensive datasets by merging geometric, image, and inertial data. This chapter provides an overview of the work undertaken and the results obtained.

5.1 Introduction

The EVOCATION project, within which this thesis was carried out, targeted the advancement in the state-of-the-art in 3D data capture through the development of a mobile system for the rapid acquisition of indoor environments using synchronized visual (panoramic) and depth sensing to improve the exploitation of such data through methods that extract structure information from the captured data, in particular through the analysis of panoramic input and point cloud data, and to improve the exploration capabilities through the development of novel methods for presenting point cloud data at the scale generated by mobile captures. In this chapter, we report on the pilot ”Indoor mapping for AEC”. The primary aim of this pilot is to show the application of the methods and technology developed within the project to the Architecture, Engineering, and Construction (AEC) fields. In particular, we describe the mobile mapping system for 3D data capture that was designed within the project and describe its practical application to the mapping of large and complex multi-room environments, construction sites, and large buildings. The device is capable of quickly capturing a collection of panoramic images aligned with 3D data in point cloud form. We summarize the design and implementation of the mapping device and system developed within the project and show its application to relevant use cases in the AEC domain. In particular, we present the initial experimental activities conducted at the indoor facilities of the Engineering University of Brescia (Italy), where the prototype mobile mapping system has been tested in small to medium-sized rooms located across three floors and connected by large stairs. We then show how the mobile mapping system has been tested in the context of a project focused on capturing social housing buildings in the municipality of Milan (Italy). This project was initiated by Metropolitana Milanese (MM), an Italian engineering company, to conduct the census of the assets of almost 500 buildings. These complexes are characterized by tall multi-floor buildings with narrow stairs connecting small rooms. We also describe the resulting data.

5.2 Data capture

The prototype mobile mapping system used for the mapping activity is the culmination of extensive research and development activities conducted throughout the project. As detailed in Chapter 3, the development of the mobile system encountered numerous challenges arising from the system's complexity and the asynchronous execution of its software components, all within the constraints of limited computational resources. The system employed for data capture represents the final development iteration as seen in Fig. 5.1.

The survey and mapping activity was conducted in 2 indoor environments with increasing degrees of complexity. The initial testing of the prototype mobile mapping system took place in the indoor facilities of the Engineering University of Brescia, Italy. The data capture process was primarily focused on a number of small to medium-sized rooms located on the first 3 floors, connected by large stairs, as depicted in Fig. 5.2.

The data, consisting of 3D point clouds and images (Fig. 5.3), was captured along 9 trajectories, covering a total length of 2680 m. The capture process took 1 hour and 11 minutes.

For each trajectory, the synchronized 3D point cloud and panoramic images are available, as shown in Fig. 5.4.

In a second phase, the prototype mobile mapping system was employed to capture

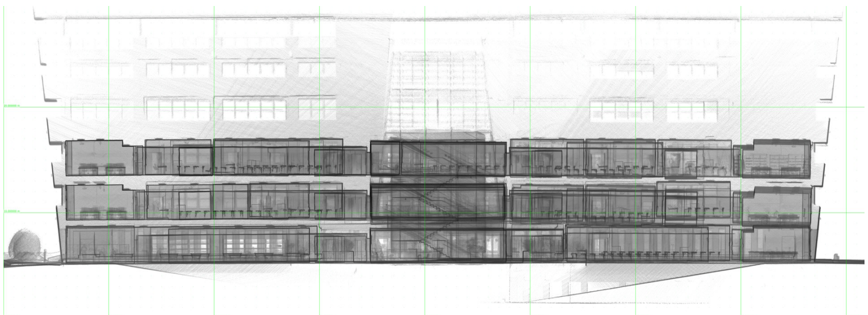
Figure 5.1: *Mobile mapping system.*

several social housing buildings in the municipality of Milan, Italy, as illustrated in Fig. 5.5.

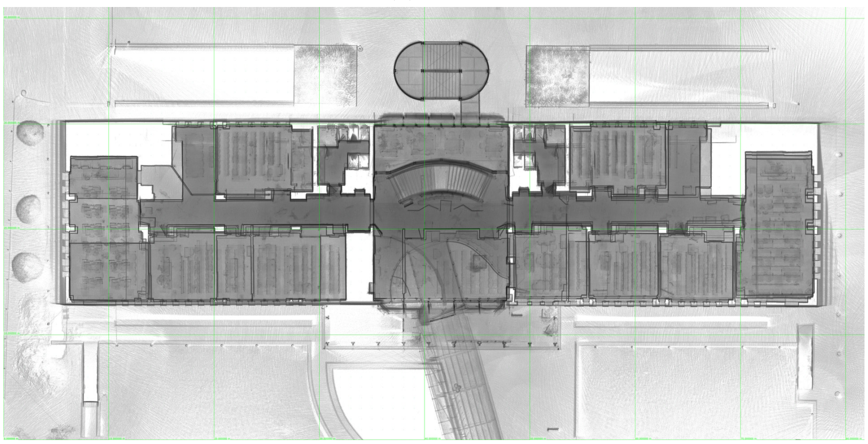
This site was selected as part of an extensive mapping and documentation project undertaken by Metropolitana Milanese (MM), an engineering company responsible for managing the social housing buildings of the Municipality of Milan in Northern Italy. The real estate consists of 498 buildings arranged into 124 complexes of structures, as illustrated in (Fig. 5.6).

The goals of the MM mapping project can be summarized in the following list of objectives:

- Recognition and categorization of all required assets, including rooms, stairs, public areas, external paths, and technical rooms, as depicted in Fig. 5.7.
- Enable accurate local measurements at the centimeter level.



(a) *facade*



(b) *map*

Figure 5.2: *Blueprint and facade view extracted from the mobile mapping survey of the University of Brescia.*

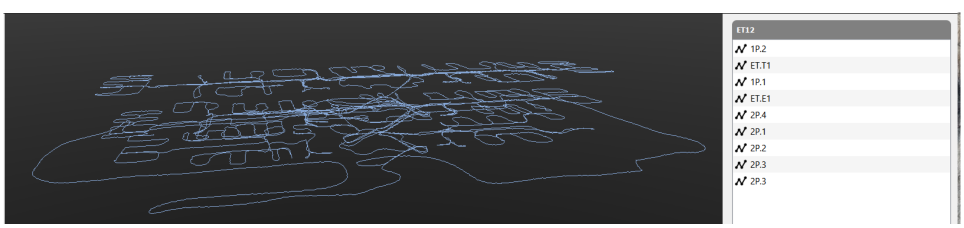


Figure 5.3: *Mobile mapping system trajectories to survey 3 floors of the University of Brescia, Italy. Total length = 2680 m, capture time = 1 hour and 11 min.*

- Publication of survey results on the cloud in the form of a virtual tour, allowing MM technical staff to access and navigate the data.
- Generation of 1:200 orthophoto images depicting the visible front views of

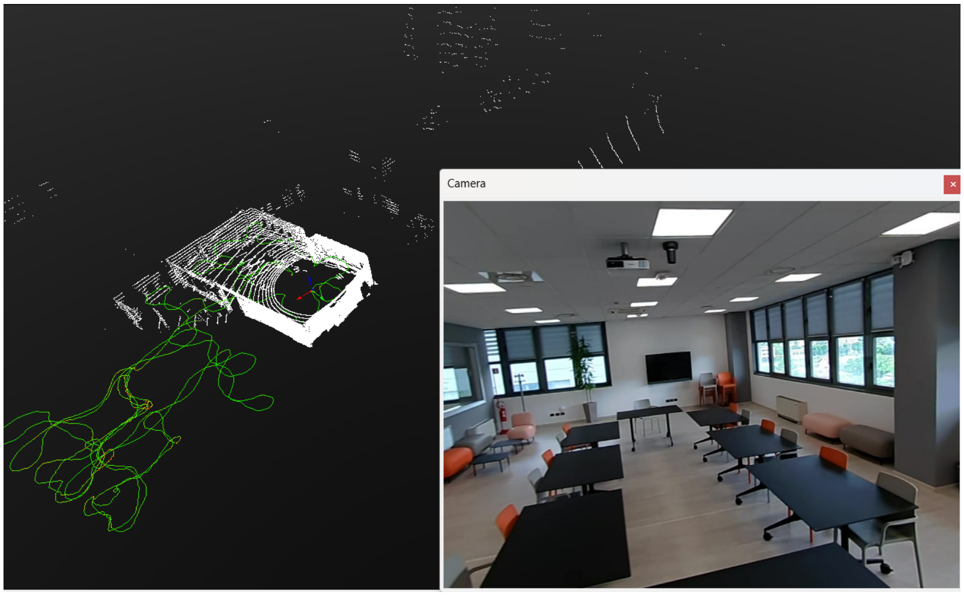


Figure 5.4: *Mobile mapping system trajectories with synchronized LiDAR point cloud and panoramic image of an indoor room of the University of Brescia, Italy.*

Figure 5.5: *Starting the acquisition with the mobile mapping prototype in Milan. The capturing system is controlled by a PDA connected via WIFI with the SoC.*

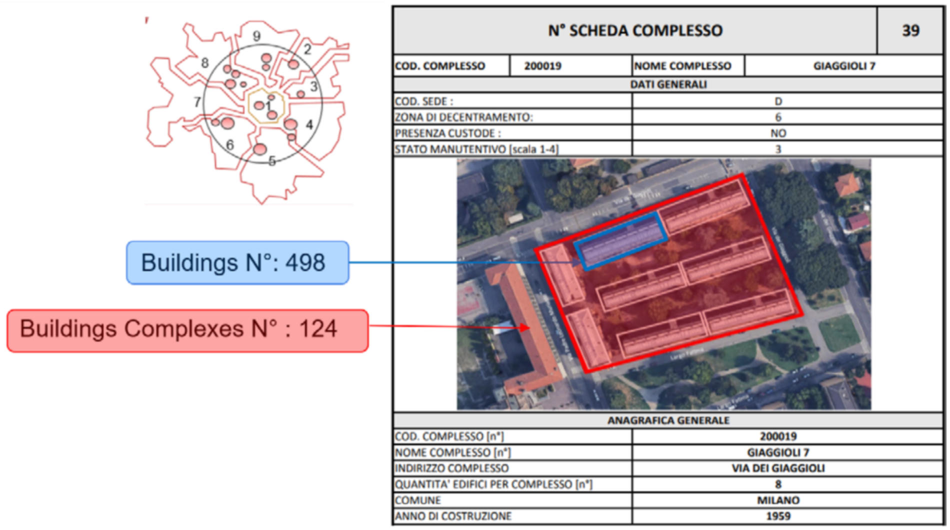


Figure 5.6: *Example of a building complex sheet and its distribution in the city of Milan.*

all buildings, enabling asset recognition and database population.

- Storage of Point Clouds and RGB data on a server using non-proprietary formats such as E57, JPEG, TIFF, and LAS.
- Implementation of a quality check procedure for real-time assessment of survey accuracy.

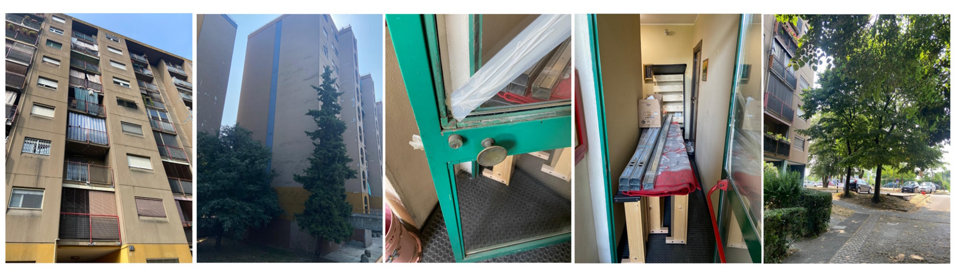


Figure 5.7: *Examples of assets that MM needs to catalog in its inventory database.*

As part of the EVOCATION project Pilot deliverable, the field survey was focused on a specific section of an MM complex, aiming to comprehensively document an indoor stairway, its associated access compartments, and a partial outdoor area, which can be seen in Fig. 5.8.

The designated area was surveyed along 7 trajectories, covering a total length of

(a) *facade* **(b)** *map*

Figure 5.8: *Blueprint and facade view extracted from the mobile mapping of the MM building in Milan.*

1171 meters and requiring a capture time of 36 minutes, as illustrated in Fig. 5.9.

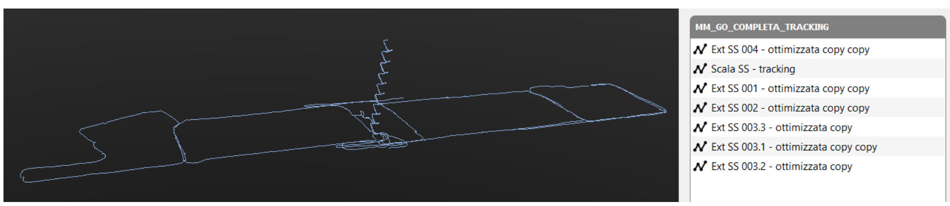


Figure 5.9: *Mobile mapping system trajectories to survey MM building in Milan. Total length = 1171 m, capture time = 36 min.*

Furthermore, for each of these trajectories, synchronized 3D point cloud data and panoramic images were collected, as shown in Fig. 5.10.

The acquisition can be performed in 2 modes:

- **Photo mode:** The user can capture equirectangular 8K images by standing still in user-defined positions for a few seconds.
- **Video mode:** The camera captures equirectangular panoramic footage at a resolution of 4K and a frame rate of 24 fps.

For the scope of this Pilot deliverable, the acquisition had to be performed in both modes. As described in Sec. 5.4, the photo mode acquisition is required to support the automatic 3D reconstruction of structured indoor environments [40]. On the other hand, the video mode enables the automatic colorization of the LiDAR data. This 3D point cloud layer was used in the processing of large point cloud datasets [40].

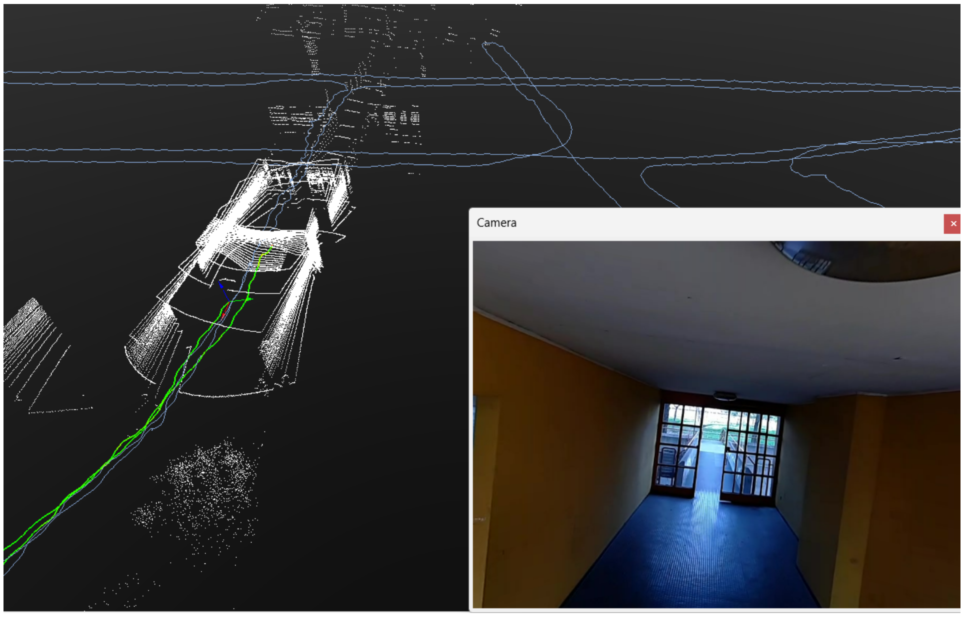


Figure 5.10: *Mobile mapping trajectory synchronized with a 4K video stream, indoors of the MM building in Milan.*

5.2.1 Static scans for ground truth and dense point clouds

In order to fulfill the requirements of data processing pipelines [41] a suitable section of the UNIBS dataset has been selected, comprising of a short corridor connected to a lounge area and a medium-sized conference room, as seen in Fig. 5.11.

The special requirement for this section is the generation of ground truth data for denser point clouds through the usage of a Faro Focus3D X 330 Terrestrial Laser Scanner (TLS), the same static scanner used in Chapter 4.

The procedure for generating a multi-room indoor dataset using this type of scanner requires selecting a sequence of fixed positions within the target area. The aim is to maximize coverage while minimizing blind spots. It is preferable to scan from positions where there is a transition between rooms, particularly beneath doorways, as this aids in the seamless integration of a global point cloud. The chosen set of fixed positions for our dataset can be observed in Chapter 4.

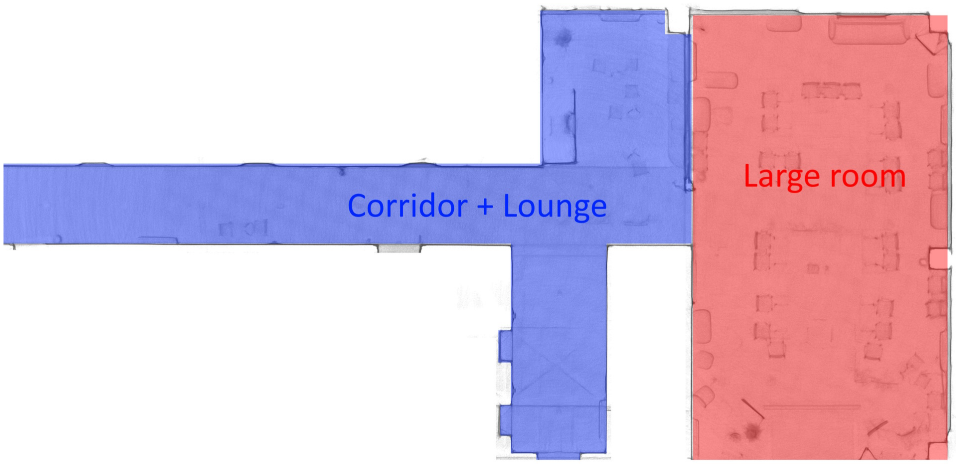


Figure 5.11: *Ortographic map view of the indoor dataset section from UNIBS selected to perform depth prediction, completion, and extraction of structural details from panoramic images.*

5.3 Data processing

During the acquisitions, the captured data is stored in a USB drive connected to the SoC. The raw data consists of the following items:

- LiDAR data (1 file per sensor, with a total of 2 sensors)
- IMU data
- 8K images (in camera mode) or 4k/24fps video (in video mode)

To generate an RGB 3D point cloud with synchronized images, the acquired data are fused together using a geometrical SLAM approach. The data is processed using two desktop software applications:

- Heron Desktop [42]
- Reconstructor [43]

Heron Desktop is dedicated software used to generate trajectories and optimize the data captured by the mobile mapping prototype. With knowledge of the final trajectory, it is possible to generate a 3D point cloud with RGB color and linked 8K images, which can be sent to Reconstructor. In Reconstructor, denoising filters are applied to the data. Reconstructor has been used to export the final point cloud models and images in non-proprietary formats that can be used by other EVOCATION partners.

The workflow in Heron Desktop follows the schematic representation shown in (Fig. 5.12).

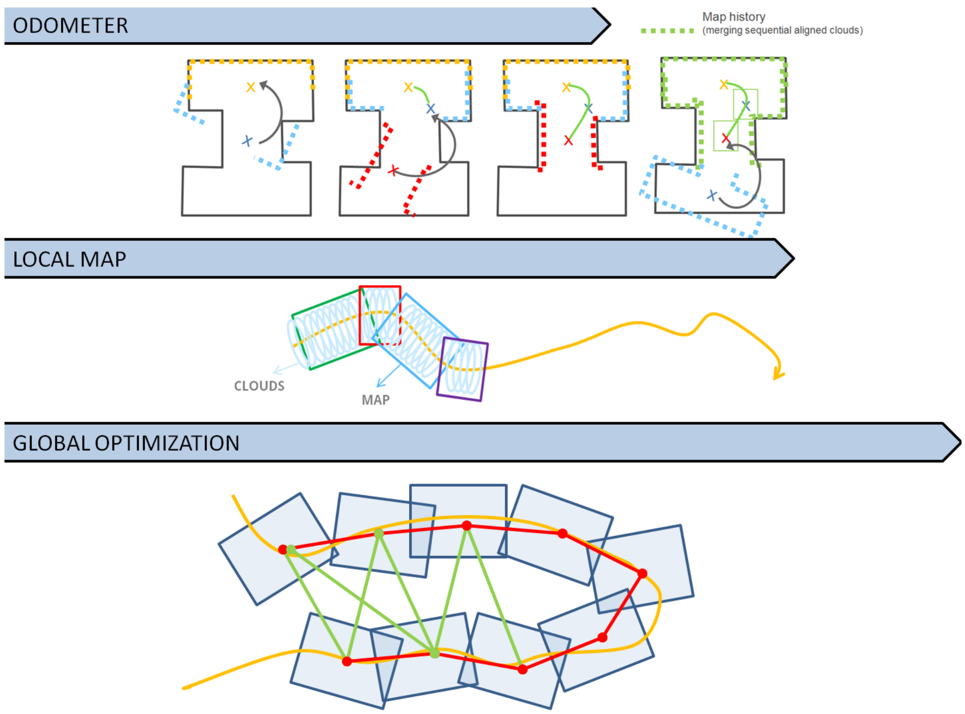


Figure 5.12: *Three-step workflow: 1) odometer, the LiDAR point clouds are sequentially aligned to create the geometric SLAM trajectory; 2) local map, the trajectory is subdivided into chunks, and the corresponding point clouds are accumulated in local maps.; 3) global optimization, the local maps are connected in a bundle adjustment process to reduce residual drifts.*

The odometer phase is responsible for fusing the IMU and LiDAR data and unwarping the original LiDAR point clouds, taking into account the system's motion during the acquisition. The unwarped clouds are sequentially aligned using a robust ICP approach [44, 45]. To reduce drift accumulation during the sequential registration, the aligned clouds from the so-called odometer step, depicted in the top row of Fig. 5.12, are merged to create a reference map. This reference map serves as a persistent reference for aligning the subsequent acquired clouds.

The size and persistence (map history) can be iteratively adjusted based on the size of the captured environment. If the captured point clouds along the path have good line-of-sight visibility with each other, a longer map history can be maintained, resulting in reduced drift. During the odometer process (see Fig. 5.13) a robust geometrical registration is performed to generate a trajectory. When the ICP algorithm provides a position estimate that aligns well with the IMU data, the trajectory is colored green to indicate its good quality.

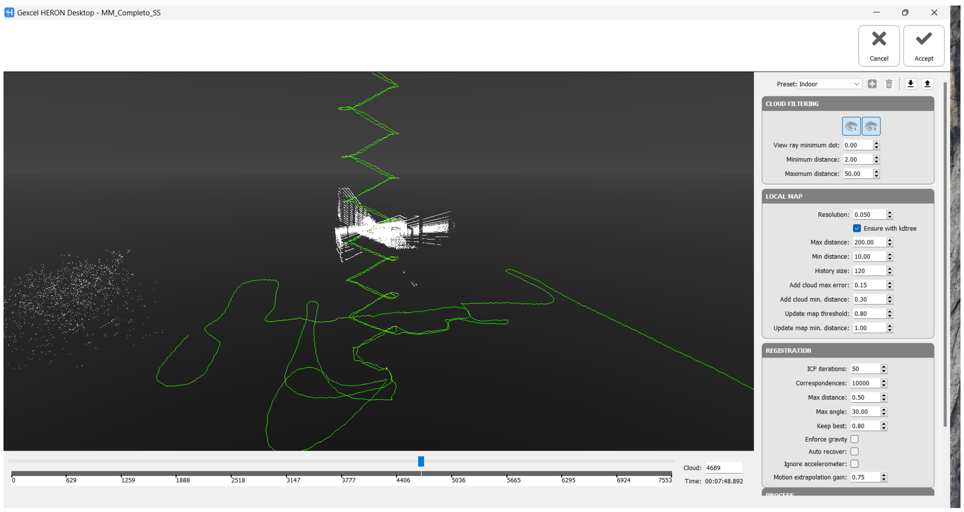


Figure 5.13: *During the odometer phase, the trajectory is generated using a geometrical SLAM approach, the green color indicates the good quality of the trajectory.*

The local map phase, represented in the middle row of Fig. 5.12, involves subdividing the trajectory into chunks and accumulating the corresponding clouds into local maps. Within each local map, the drift is considered to be null. The size of the local maps can be adjusted based on the characteristics of the environment. In open spaces where the captured clouds are more likely to have line-of-sight visibility, larger local maps can be used (Fig. 5.14).

The global optimization is the final step in refining the trajectory and addressing potential drift. In the bottom row of Fig. 5.12, the schematic representation illustrates the goal of global optimization: the local maps, which are initially aligned only along the trajectory, can be connected in the overlapping areas closing loops along the path. Once the overlapping local maps are aligned using ICP, a global optimization is run to minimize registration errors (Fig. 5.15).

In this last phase, the point clouds captured by the Terrestrial Laser Scanner (TLS), as described in Sec. 5.2.1, were used as local maps in the global optimization of certain trajectories. Due to the overlap with the local maps derived from the mobile mapping system, multiple connections could be established between the TLS cloud and the mobile cloud. By considering the TLS data as ground truth for the mobile survey, its position was fixed during the global optimization process. This

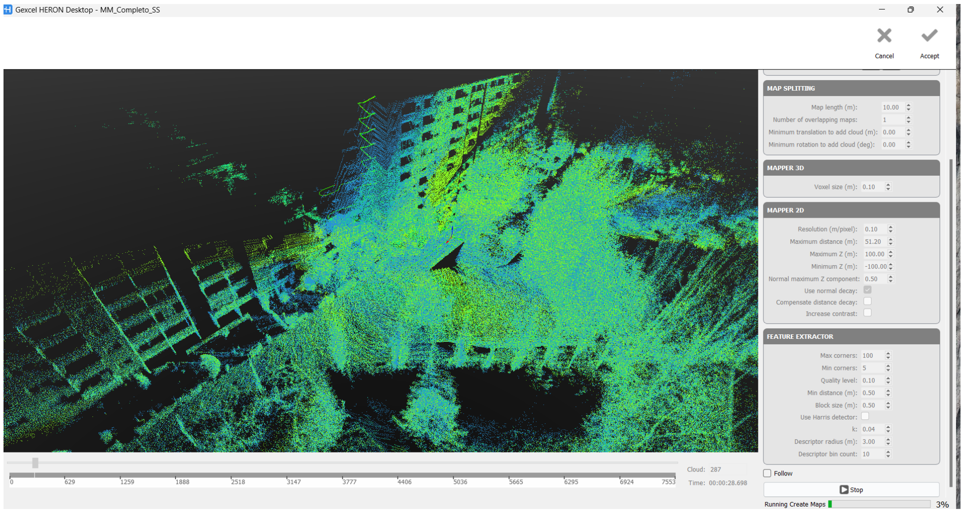


Figure 5.14: *During the local map phase, the generated local maps are displayed using different colors.*

incorporation of TLS constraints benefited the final mobile trajectory, ensuring its alignment with the TLS data. As a result, the final point cloud exhibited accurate alignment with the TLS data.

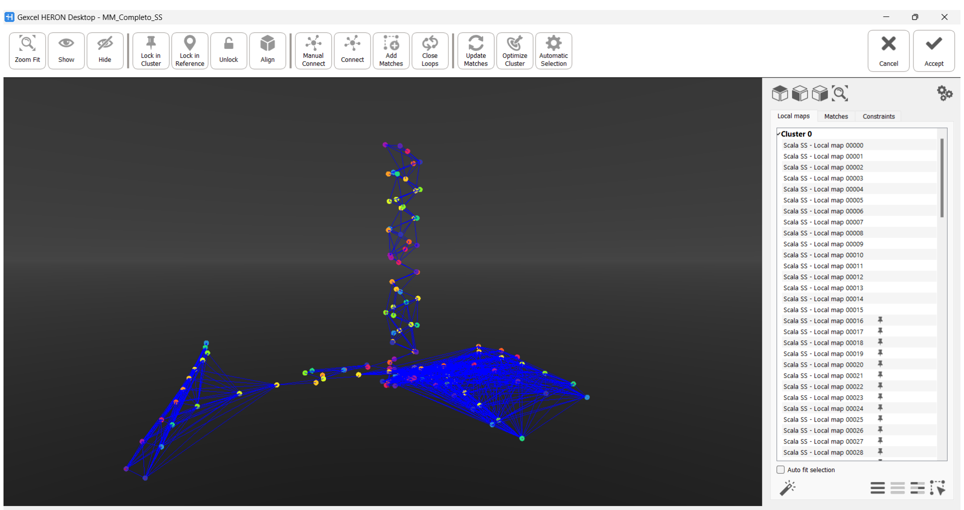
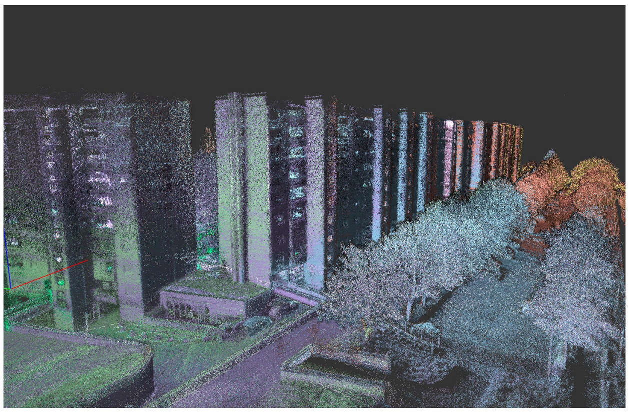
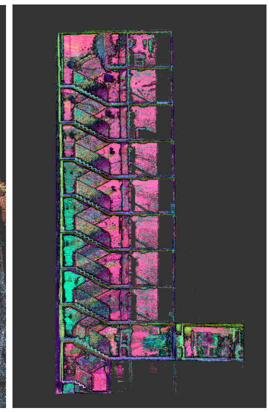


Figure 5.15: *Local maps, represented by the colored dots, are connected together for the final bundle adjustment.*

The trajectory resulting from the global optimization allows for the generation of the final point cloud, which can be rendered and optionally filtered using the Reconstructor software [43]. The final MM survey results in a point cloud with a resolution of 2 cm and approximately 500 million points (Fig. 5.16).



(a) *Exterior view*



(b) *Indoor stairs*

Figure 5.16: *MM building survey: 2 views of the final point cloud, one showcasing the outdoor area and the other highlighting the indoor stairs.*

A portion of the indoor survey of the University of Brescia is visible in Fig. 5.17, showcasing RGB colorization obtained from the camera capturing at 24fps/4k.

In the Reconstructor [43] software, it is possible to navigate along the trajectories while overlapping the 8K single-shot images, as shown in Fig. 5.18. Additionally, the processed data can be shared in open formats, as discussed in Sec. 5.4 on Data Delivery.

5.4 Data delivery

After the data capture (Sec. 5.2) and the data processing (Sec. 5.3), the final results of the indoor mobile mapping survey are as follows:

- A 3D point cloud with multiple color layers (such as intensity, RGB, inclination, range, etc.), which can be visualized at different resolutions.
- Trajectories with timestamps and 3D positions (X, Y, Z, roll, pitch, yaw).
- Panoramic (equirectangular) video at a resolution of 4K and a frequency of 24fps. Each video frame is accompanied by timestamp, X, Y, Z, roll, pitch, and yaw.

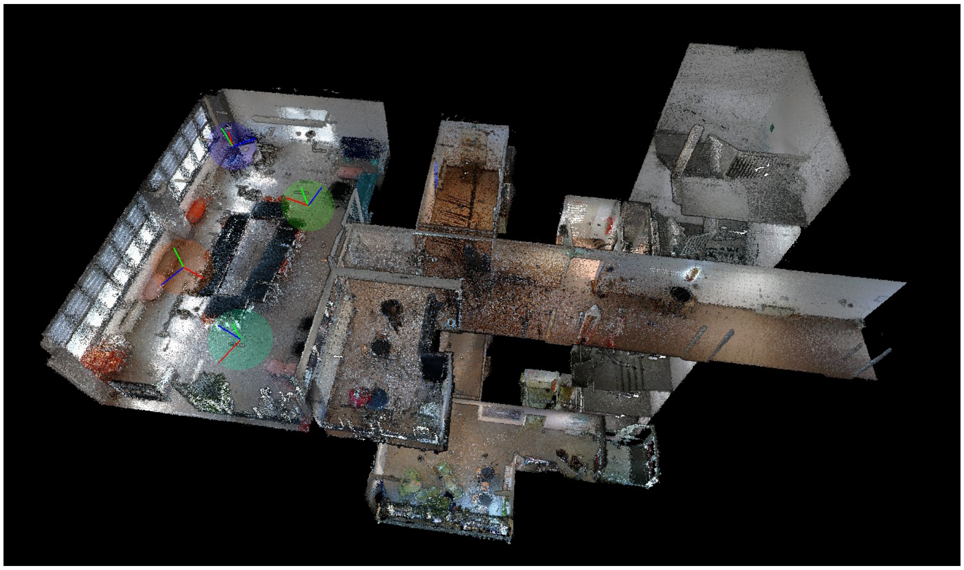


Figure 5.17: *3D points of a multi-floor capture with RGB information layer.*

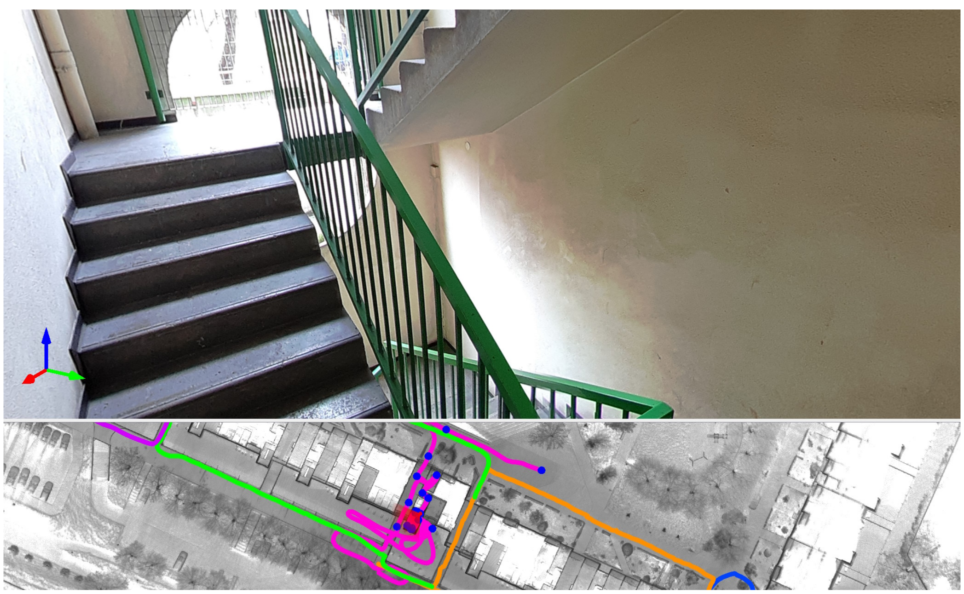


Figure 5.18: *A map view with trajectories allows for navigation while overlapping the 8K images onto the captured 3D geometry.*

- Panoramic images (equirectangular) at a resolution of 8K. Timestamp, X, Y, Z, roll, pitch and yaw are available for each single-shot image.

All the acquired data is synchronized, and the relative positions between the different sensors are known. To leverage the captured data through methods that extract structural information from synchronized panoramic input and point cloud data, as well as to explore novel methods for presenting point cloud data at the scale generated by mobile capture, the aforementioned results have been exported and shared with the EVOCATION partners in three main ways:

- Depth maps corresponding to single indoor panorama.
- RGB structured data of single rooms
- Large point clouds with multiple color layers

Detailed descriptions of the methods used to share the data and the specific data formats are reported in the following three sections.

5.4.1 Depth maps in correspondence of single indoor panorama

The steps performed to prepare the data for delivery are very similar to the ones followed for the elaboration of the Indoor3Dmapping dataset [32], already described in Chapter 4.

5.4.2 RGB structured data of single connected rooms

The processing steps carried out to prepare the data for the automatic reconstruction process [40] are as follows:

- Mobile mapping survey segmentation in separated rooms. The indoor survey, which covers multiple contiguous rooms connected by a corridor, was subdivided using the Reconstructor [43] software. The entire point cloud was displayed in an orthographic top view, and the point clouds corresponding to each room were manually selected. A sub-copy of the point cloud covering each individual room was created (Fig. 5.19). Additionally, inside each room a 8K equirectagular image is included, and the relative position of each image with respect to the 3D point cloud was known and represented in (Fig. 5.19) with circles and reference system drawings.
- Virtual scan from the camera position. For each camera position, the Reconstructor [43] software allows the definition of a spherical camera view (360° lat/long). The spherical camera was positioned at the same X, Y, Z coordinates

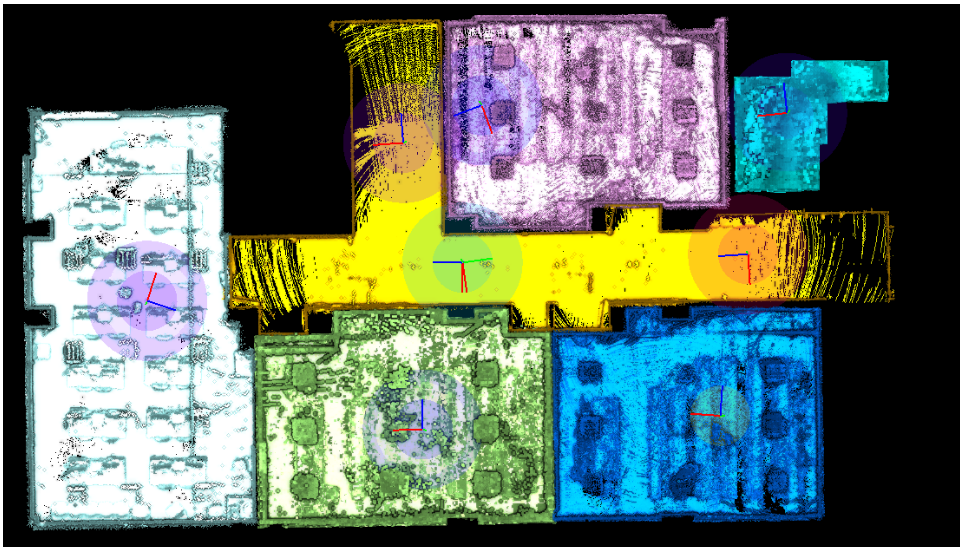


Figure 5.19: *Orthographic map view of the indoor survey with the rooms segmentation. The circles inside each room represents the position of the 8K images.*

as the 8K images, but with the Z-axis in a vertical position relative to the room floor (Fig. 5.20).

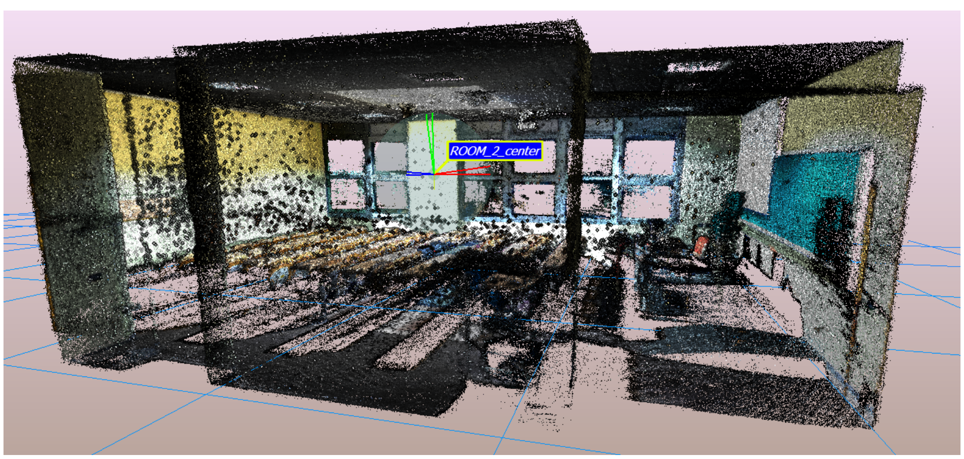


Figure 5.20: *The 8K spherical images are re-projected onto the mobile system cloud. A spherical camera is positioned at the same X, Y, Z coordinates as the mobile system, with the Z-axis oriented vertically.*

The 8K image captured along the mobile trajectory was re-projected onto the point cloud by creating six perspective views (up, down, left, right, forward, and backward) in a cubemap format. This allowed for mapping each pixel of the image onto the mobile point cloud coordinate system. The points on the faces of the cubemap were then mapped into a new common reference system to create an equirectangular grid point cloud. This process, known as virtual scan, was performed using GPU off-screen rendering technique (Fig. 5.21).

Figure 5.21: *The virtual scan process maps each pixel of the 8K image onto the coordinate system of the mobile point cloud.*

- Export point cloud in E57 format [46]. The virtual scan step generates a point cloud with the origin at the camera position (Fig. 5.22).

  The output point cloud is organized in rows and columns, following the structure of the spherical camera (Fig. 5.23).The Z-axis is orthogonal to the room's floor, and the point cloud includes multiple color layers such as RGB, inclination, and range. The E57 open format [46] was chosen to share the data while preserving these characteristics.

5.4.3 Large point cloud with color layers in LAS format

As reported in Sec. 5.2 (Data capture), the mobile system survey conducted at the University of Brescia followed 9 trajectories, covering a total distance of 2680 m and taking 1 hour and 11 minutes to complete. The full resolution point cloud obtained from the survey contains approximately 34 billion points. To facilitate

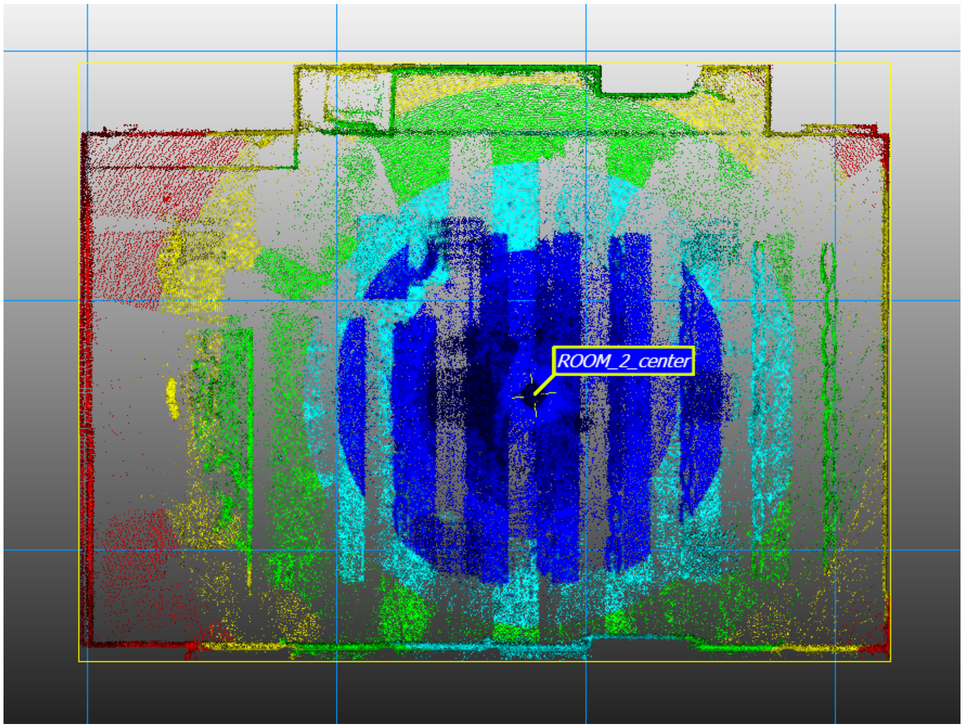


Figure 5.22: *The virtual scan generates a point cloud with the origin at the camera position, as seen in the range color layer of the resulting cloud.*

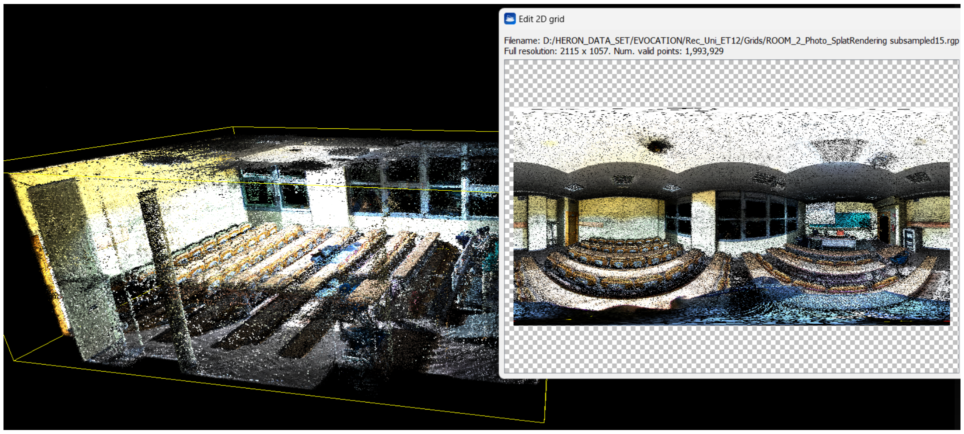


Figure 5.23: *Structured point cloud resulting from the virtual scan step, ready to be exported in E57 format.*

data processing and visualization, the full resolution point cloud was voxelized at a resolution of 3 cm, resulting in a point cloud of 102 million points that includes an intensity color layer (Fig. 5.24). Additionally, a subset of the entire point cloud was voxelized at a resolution of 1 cm, generating a point cloud with 70 million points that includes an RGB color layer (Fig. 5.17).

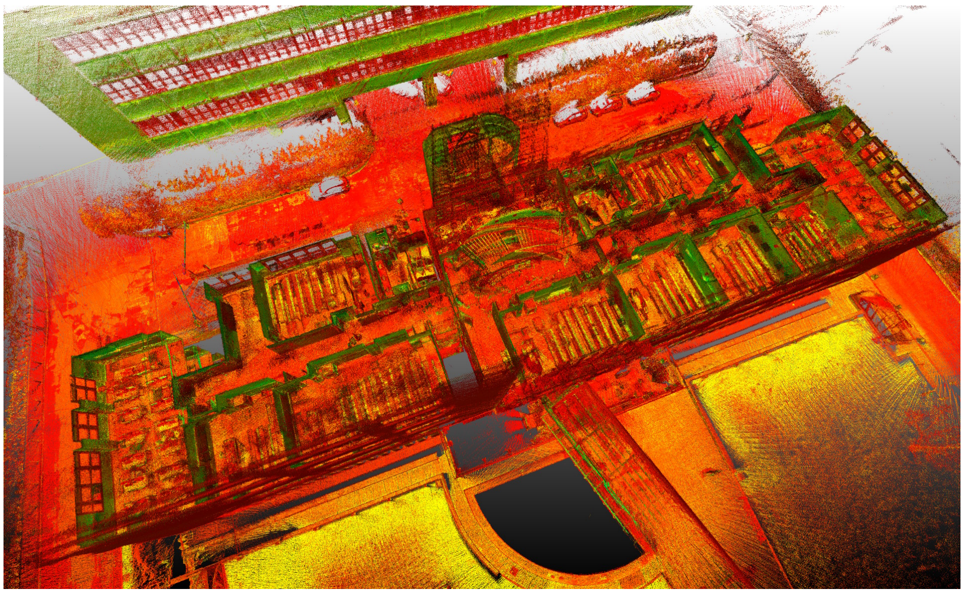


Figure 5.24: *Final point cloud of the University of Brescia (102 million points, ready to be exported in LAS format).*

The MM buildings in Milan were surveyed along 7 trajectories for a total length of 1171 m and a capture time of 36 min. The full resolution point cloud obtained from this survey contains approximately 17 billion points. The full resolution point cloud was then voxelized at a resolution of 3 cm, creating a point cloud of around 231 million points. Before sharing the clouds from both environments, a denoise filter was applied. This filter reduces sensor noise when the underlying geometry is clearly characterized. It describes the geometry with a linear transformation of the eigenvalues associated with the Principal Component Analysis computed in a variable neighborhood around each point. If the geometry behaviour cannot be determined, the original data is left untouched. The denoise filter achieves the best performance when the data show planar or smooth behaviour (Fig. 5.25).

Both datasets were exported in LAS open format [39]. This format was chosen to facilitate the visualization of large point cloud data [40].

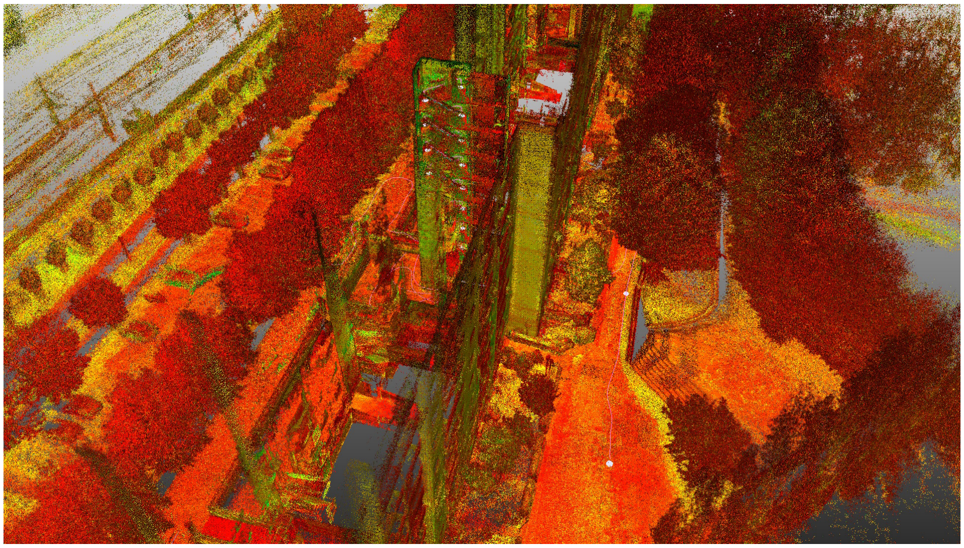


Figure 5.25: *Final point cloud of the MM Building (231 million points, ready to be exported in LAS format).*

5.5 Conclusions

The pilot study "Indoor Mapping for AEC," as part of the EVOCATION project, represents a significant milestone in the integration of advanced mobile mapping technologies within the Architecture, Engineering, and Construction (AEC) sectors. The successful deployment and testing of the prototype mobile mapping system in complex, real-world environments not only demonstrate the system's robustness and versatility but also highlight its potential to revolutionize the field of indoor mapping.

The utilization of this technology in diverse settings, from the University of Brescia's multi-floor environment to the intricate social housing complexes in Milan, showcases its adaptability and efficiency in capturing detailed spatial data. The combination of panoramic imagery with 3D LiDAR data, processed through sophisticated algorithms, has yielded rich datasets that open new avenues for accurate and comprehensive indoor analysis and documentation.

Key achievements of this pilot include:

- The successful capture and processing of extensive indoor data, covering thousands of meters in various settings, emphasizing the system's capability for large-scale deployment.

- The development of innovative data processing techniques, including the creation of structured depth maps and RGB data, facilitating deeper analysis and visualization of indoor spaces.
- The demonstration of the system's practical applicability in AEC scenarios, providing valuable insights into asset management, structural analysis, and renovation planning.
- Contributing to the broader goal of EVOCATION in enhancing data exploitation methods and exploration capabilities in 3D spatial data acquisition.

Furthermore, the insights gained from this pilot study lay the groundwork for future developments in indoor mapping technologies. The challenges encountered and overcome during the project have provided valuable lessons that can inform subsequent iterations of mobile mapping systems, particularly in enhancing their efficiency, accuracy, and user-friendliness.

As we look ahead, the integration of such technologies in AEC and other sectors holds great promise. The potential for creating more dynamic, interactive, and detailed models of indoor environments can significantly impact design, maintenance, and management processes within these industries. Moreover, the advancements in data processing and visualization techniques developed through this project have the potential to transform how professionals and stakeholders interact with spatial data, leading to more informed decision-making and innovative solutions.

In conclusion, the ”Indoor Mapping for AEC” pilot within the EVOCATION project not only achieved its immediate objectives but also made substantial contributions to the field of indoor mapping. It paves the way for further innovation and application of mobile mapping technologies, marking a significant step forward in our ability to understand and interact with the built environment.

5.6 Bibliographic notes

A major part of this chapter has been taken from my contribution to the EVOCATION Deliverable D6.2 titled ”Indoor Mapping for AEC” [40].

Chapter 6

Conclusion

This chapter encapsulates the journey and achievements of this PhD thesis, discussing the solutions developed, their implications, future research directions, and the scholarly contributions made through publications and demonstrations.

6.1 Overview of achievements

This research addressed the challenges in mobile mapping systems (MMS) with a focus on large environment mapping and data fusion techniques. Key achievements include:

- Development of a prototype integrated mobile 3D mapping system, enhancing data acquisition efficiency in large and complex environments.
- Advanced data processing methodologies, enabling the fusion of geometric, image, and inertial data into comprehensive datasets.
- Successful implementation and testing of the developed system in real-world scenarios, demonstrating its practical applicability in the Architecture, Engineering, and Construction (AEC) sectors.

6.2 Discussion

The solutions developed in this thesis present several advantages, such as enhanced efficiency in data acquisition and improved accuracy through data fusion techniques. However, limitations such as the computational demands of processing large datasets and the need for further refinement in sensor fusion algorithms were identified. Future work should focus on optimizing these aspects to extend the system's applicability in more diverse and challenging environments.

6.3 Future directions

Future research directions include:

1. **Enhanced Data Integration and Fusion**: Further research could focus on algorithms for integrating data from diverse sources like aerial drones, ground-based MMS, and static sensors, to create comprehensive models.
2. **Autonomous Navigation and Data Acquisition**: Exploring autonomous navigation in MMS, using AI and machine learning for obstacle detection and optimal route planning, presents a significant area for future development.
3. **Real-Time Processing and Visualization**: Developing methods for real-time data processing and visualization could greatly enhance the applicability of MMS, especially in critical applications.
4. **Advancements in SLAM Algorithms**: The refinement of SLAM algorithms, especially for indoor environments, is an essential area for ongoing research.
5. **Integration with Building Information Modeling (BIM)**: Further development in integrating MMS data with BIM models can significantly benefit construction and maintenance sectors.
6. **Machine Learning for Object Recognition**: Implementing machine learning for automatic object recognition within point clouds can open new possibilities in various domains.
7. **Improved User Interfaces and Interaction**: Research into developing intuitive user interfaces, possibly incorporating augmented reality, can make MMS data more accessible.
8. **Ethical and Privacy Considerations**: As MMS usage expands, addressing privacy concerns, especially in public spaces, will become increasingly important.

Each direction not only addresses current challenges but also expands the potential applications of MMS technology.

6.4 Publications

The scientific results obtained during this PhD work also appeared in related publications:

- **Deep Panoramic Depth Prediction and Completion for Indoor Scenes.**
  Giovanni Pintore, Eva Almansa, Armando Sanchez, Giorgio Vassena, and Enrico Gobbetti, Computational Visual Media, 2023.
  DOI: 10.1007/s41095-023-0358-0.

- **Indoor3Dmapping dataset.**
  Armando Arturo Sánchez Alcázar, Giovanni Pintore, and Matteo Sgrenzaroli, Zenodo, 2022.
  DOI: 10.5281/zenodo.6367381.

- **Indoor mobile mapping systems and BIM digital models for construction progress monitoring.**
  M. Sgrenzaroli, J. Ortiz Barrientos, G. Vassena, A. Sanchez, A. Ciribini, S. Mastrolembo Ventura, and S. Comai,
  The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2022, 2022, Pages 121–127.
  DOI: 10.5194/isprs-archives-XLIII-B1-2022-121-2022.

- **New technologies for high-level reconstruction with visual and depth sensing.**
  Moonisa Ahsan, Fabio Marton, Ruggero Pintus, Armando Sanchez, Matteo Sgrenzaroli, Eva Almansa, Enrico Gobbetti, Giovanni Pintore, Lizeth Fuentes, and Renato Pajarola.
  Tech. rep. EVOCATION Project, H2020 MSCA 813170, Feb. 2023.

- **Indoor Mapping for AEC.**
  Armando Sanchez, Matteo Sgrenzaroli, Giorgio Vassena, Eva Almansa, Enrico Gobbetti, Giovanni Pintore, Lizeth Fuentes, Luciano Romero, and Renato Pajarola.
  Tech. rep. EVOCATION Project, H2020 MSCA 813170, May 2023.

6.5 Demonstration videos

In the context of the EVOCATION project, I have also illustrated the outcomes of my research in the following demonstration videos that are available on the project website at the URL evocation.eu/videos/:

- **Pilot 2: Indoor Mapping for AEC - 1. Introduction** — Demo video.
  This pilot project showcases the efforts focused on utilizing a prototype mobile mapping system to capture indoor 3D environments and demonstrates its applicability in relevant use cases within the AEC domain. All collected data is prepared for further analysis and processed, aiming to enhance and extract valuable information with techniques such as depth prediction and completion, clutter removal, semantic segmentation for CAD models generation and point cloud visualization.

- **Pilot 2: Indoor Mapping for AEC - 1. Introduction** — Demo video.
  A comprehensive overview of the data acquisition phase, which includes data capture using a prototype mobile mapping system in indoor environments, data preparation and its delivery in compatible formats suitable for the requirements of subsequent experimental and innovative data processing pipelines and visualization techniques.

Bibliography

[1] *RIEGL Laser Measurement Systems*. http://riegl.com/. (Visited on 02/06/2024).

[2] *Trimble DA2 — GNSS Systems*. https://geospatial.trimble.com/products/hardware/trimble-da2. (Visited on 02/06/2024).

[3] Movella. *Home - Xsens 3D motion tracking*. https://www.xsens.com. [Visited on 09-August-2022].

[4] https://velodynelidar.com/. [Visited on 09-August-2022].

[5] H. Durrant-Whyte and T. Bailey. "Simultaneous Localization and Mapping: Part I". In: *IEEE Robotics & Automation Magazine* 13.2 (June 2006), pp. 99–110. ISSN: 1558-223X. DOI: 10.1109/MRA.2006.1638022.

[6] T. Bailey and H. Durrant-Whyte. "Simultaneous Localization and Mapping (SLAM): Part II". In: *IEEE Robotics & Automation Magazine* 13.3 (Sept. 2006), pp. 108–117. ISSN: 1070-9932. DOI: 10.1109/MRA.2006.1678144.

[7] Gexcel. *HERON - Gexcel*. https://gexcel.it/en/solutions/heron-portable-3d-mapping-system. [Visited on 09-August-2022].

[8] *ZEB Horizon: SLAM LiDAR for 3D Mapping with Drones*. https://geoslam.com/solutions/zeb-horizon/. [Visited on 09-August-2022].

[9] NavVis GmbH. *NavVis VLX 2nd generation*. https://www.navvis.com/vlx. [Visited on 09-August-2022].

[10] *GreenValley International - 3D Laser Scanning and Lidar360 Point Cloud Service Provider*. https://www.greenvalleyintl.com/. (Visited on 02/06/2024).

[11] M. Sgrenzaroli, J. Ortiz Barrientos, G. Vassena, A. Sanchez, A. Ciribini, S. Mastrolembo Ventura, and S. Comai. "Indoor mobile mapping systems and BIM digital models for construction progress monitoring". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B1-2022 (2022), pp. 121–127. ISSN: 1682-1750. DOI: 10.5194/isprs-archives-XLIII-B1-2022-121-2022.

[12] Giorgio P. M. Vassena, Luca Perfetti, Sara Comai, Silvia Mastrolembo Ventura, and Angelo L. C. Ciribini. "Construction Progress Monitoring through the Integration of 4D BIM and SLAM-Based Mapping Devices". In: *Buildings* 13 (2023), p. 2488. ISSN: 2075-5309. DOI: 10.3390/buildings13102488.

[13] L. Perfetti, F. Marotta, F. Fassi, and G. P. M. Vassena. "Survey of historical gardens: multi-camera photogrammetry vs mobile laser scanning". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-1/W1-2023 (2023), pp. 387–394. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLVIII-1-W1-2023-387-2023.

[14] G. P. M. Vassena and M. Sgrenzaroli. "Innovative and efficient data census on a cloud based digital twin provided by mobile mapping". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-1-W1-2023 (2023), pp. 531–537. ISSN: 1682-1750. DOI: 10.5194/isprs-archives-XLVIII-1-W1-2023-531-2023.

[15] *Labpano*. https://www.labpano.com/. [Visited on 09-August-2022]. (Visited on 08/11/2022).

[16] https://www.hesaitech.com/. [Visited on 01-February-2023].

[17] NVIDIA Developer. *Jetson Modules*. https://developer.nvidia.com/embedded/jetson-modules. [Visited on 09-August-2022].

[18] https://www.asrockind.com/. [Visited on 01-February-2023].

[19] *Rust Programming Language*. https://www.rust-lang.org/. [Visited on 09-August-2022].

[20] *Dart programming language*. https://dart.dev/. [Visited on 09-August-2022].

[21] *Flutter - Build apps for any screen*. //flutter.dev/. [Visited on 09-August-2022].

[22] *Yocto Project – It's not an embedded Linux distribution – it creates a custom one for you*. https://www.yoctoproject.org/. [Visited on 09-August-2022].

[23] *gRPC*. https://grpc.io/. [Visited on 09-August-2022].

[24] Google Developers. *Protocol Buffers*. https://developers.google.com/protocol-buffers. [Visited on 09-August-2022].

[25] Moonisa Ahsan, Fabio Marton, Ruggero Pintus, Armando Sánchez, Matteo Sgrenzaroli, Eva Almansa, Enrico Gobbetti, Giovanni Pintore, Lizeth Fuentes, and Renato Pajarola. *New technologies for high-level reconstruction with visual and depth sensing*. Tech. rep. EVOCATION Project, H2020 MSCA 813170, Feb. 2023.

[26] Giovanni Pintore, Claudio Mura, Fabio Ganovelli, Lizeth Fuentes-Perez, Renato Pajarola, and Enrico Gobbetti. "State-of-the-art in Automatic 3D Reconstruction of Structured Indoor Environments". In: *Comput. Graph. Forum* 39.2 (2020), pp. 667–699.

[27] Stanford University. *Bundle Fusion Dataset*. https://graphics.stanford.edu/projects/bundlefusion. [Accessed: 2019-09-25]. 2016.

[28] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. *ScanNet Data*. http://www.scan-net.org/. [Accessed: 2019-09-25]. 2017.

[29] Jrgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. "A benchmark for the evaluation of RGB-D SLAM systems". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012). Vilamoura-Algarve, Portugal: IEEE, Oct. 2012, pp. 573–580. DOI: 10.1109/IROS.2012.6385773. URL: http://ieeexplore.ieee.org/document/6385773/ (visited on 11/30/2021).

[30] Pushmeet Kohli Nathan Silberman Derek Hoiem and Rob Fergus. "Indoor Segmentation and Support Inference from RGBD Images". In: *Proc. ECCV*. 2012, pp. 746–760.

[31] J. Xiao, A. Owens, and A. Torralba. "SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels". In: *2013 IEEE International Conference on Computer Vision*. Dec. 2013, pp. 1625–1632.

[32] Armando Arturo Sánchez Alcázar, Giovanni Pintore, and Matteo Sgrenzaroli. *Indoor3Dmapping dataset*. Type: dataset. Mar. 18, 2022. DOI: 10.5281/zenodo.6367381. URL: https://zenodo.org/record/6367381 (visited on 08/10/2022).

[33] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. "Matterport3D: Learning from RGB-D Data in Indoor Environments". In: *Proc. 3DV*. 2017, pp. 667–676.

[34] Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. "Structured3D: A Large Photo-realistic Dataset for Structured 3D Modeling". In: *ArXiv e-print arXiv:1908.00222* (2019).

[35] Steve Cruz, Will Hutchcroft, Yuguang Li, Naji Khosravan, Ivaylo Boyadzhiev, and Sing Bing Kang. "Zillow Indoor Dataset: Annotated Floor Plans With 360o Panoramas and 3D Room Layouts". In: (2021), p. 11.

[36] Xuan Luo, Jia-Bin Huang, Richard Szeliski, Kevin Matzen, and Johannes Kopf. "Consistent Video Depth Estimation". In: *arXiv:2004.15021 [cs]* (Aug. 26, 2020). arXiv: 2004.15021. URL: http://arxiv.org/abs/2004.15021 (visited on 11/29/2021).

[37] Zak Murez, Tarrence van As, James Bartolozzi, Ayan Sinha, Vijay Badrinarayanan, and Andrew Rabinovich. "Atlas: End-to-End 3D Scene Reconstruction from Posed Images". In: *arXiv:2003.10432 [cs]* (Oct. 14, 2020). arXiv: 2003.10432. URL: http://arxiv.org/abs/2003.10432 (visited on 11/24/2020).

[38] Aljaž Božič, Pablo Palafox, Justus Thies, Angela Dai, and Matthias Nießner. "TransformerFusion: Monocular RGB Scene Reconstruction using Transformers". In: *arXiv:2107.02191 [cs]* (July 5, 2021). arXiv: 2107.02191. URL: http://arxiv.org/abs/2107.02191 (visited on 11/29/2021).

[39] *LASer (LAS) File Format Exchange Activities – ASPRS*. https://www.asprs.org/divisions-committees/lidar-division/laser-las-file-format-exchange-activities. [Visited on 20-January-2023].

[40] Armando Sánchez, Matteo Sgrenzaroli, Giorgio Vassena, Eva Almansa, Enrico Gobbetti, Giovanni Pintore, Lizeth Fuentes, Luciano Romero, and Renato Pajarola. *Indoor Mapping for AEC*. Tech. rep. EVOCATION Project, H2020 MSCA 813170, May 2023.

[41] Giovanni Pintore, Eva Almansa, Armando Sanchez, Giorgio Vassena, and Enrico Gobbetti. "Deep Panoramic Depth Prediction and Completion for Indoor Scenes". In: *Computational Visual Media* (2023). To appear. DOI: 10.1007/s41095-023-0358-0. URL: http://vic.crs4.it/vic/cgi-bin/bib-page.cgi?id='Pintore:2023:DPD'.

[42] *Gexcel - Heron Destkop - Processing software for SLAM data from Heron*. https://gexcel.it/en/software/heron-desktop. [Visited on 12-May-2023].

[43] *Gexcel - Reconstructor - The powerful processing software for lidar data*. https://gexcel.it/en/software/reconstructor. [Visited on 12-May-2023].

[44] Carlos Sánchez Belenguer, Pierluigi Taddei, Simone Ceriani, Erik Wolfart, and Vítor Sequeira. "Localization and tracking in known large environments using portable real-time 3D sensors". In: *Computer Vision and Image Understanding*. Special issue on Assistive Computer Vision and Robotics - "Assistive Solutions for Mobility, Communication and HMI" 149 (Aug. 1, 2016), pp. 197–208. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2015.11.012.

[45] Carlos Sánchez Belenguer, Simone Ceriani, Pierluigi Taddei, Erik Wolfart, and Vítor Sequeira. "Global matching of point clouds for scan registration and loop detection". In: *Robotics and Autonomous Systems* 123 (Jan. 1, 2020), p. 103324. ISSN: 0921-8890. DOI: 10.1016/j.robot.2019.103324.

[46] *The E57 File Format*. http://libe57.org/. [Visited on 18-May-2023].

Appendix A

Curriculum Vitae

Contact Information

Name	**Armando Arturo Sánchez Alcázar**
E-Mail	armando.sanchez.alcazar@gmail.com
ORCID	0000-0002-6011-3299

Education

2019 - 2024

Ph.D in Computer Science, specialty Computer Vision and Graphics. / **Marie Sk lodowska-Curie Fellow**
Doctoral program in Department of Mathematics and Computer Science.
University of Cagliari (UniCa), Italy.

2017 - 2019

MSc in Industrial Engineering
Universitat Politècnica de València, UPV
Main topics: **Automatic Systems, Control and Robotics.**
Master Final Project: Implementation of a mobile manipulator robot for equipment and materials transportation in the health care industry.

2012 - 2017

BSc in Industrial Engineering
Universitat Politècnica de València, UPV.
BSc Final Project: Design of an image capture system for recording laser speckle patterns.

Relevant Work Experience

2019 - 2023

Marie Sk lodowska-Curie Early Stage Researcher and Software Engineer
Gexcel, Italy.
Participated in the development of integrated mobile 3D mapping systems. Worked on software solutions combining laser scanning and panoramic imaging for large-scale environmental mapping.

Scientific Publications

Journal Articles

1. **Deep Panoramic Depth Prediction and Completion for Indoor Scenes.**
   Giovanni Pintore*, Eva Almansa*, Armando Sanchez, Giorgio Vassena, and Enrico Gobbetti, Computational Visual Media, 2023.
   DOI: 10.1007/s41095-023-0358-0 — View PDF.

2. **Indoor mobile mapping systems and BIM digital models for construction progress monitoring.**
   M. Sgrenzaroli, J. Ortiz Barrientos, G. Vassena, A. Sanchez, A. Ciribini, S. Mastrolembo Ventura, and S. Comai,
   The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2022, 2022, Pages 121–127.
   DOI: 10.5194/isprs-archives-XLIII-B1-2022-121-2022.

Datasets

1. **Indoor3Dmapping dataset.**
   Armando Arturo Sánchez Alcázar, Giovanni Pintore, and Matteo Sgrenzaroli, Zenodo, 2022.
   DOI: 10.5281/zenodo.6367381.

Personal Details

Languages Proficiency	Spanish (native), English (fluent), Italian (conversational)
Place of Birth	Spain