AUGMENTED REALITY – ENHANCED CILTURAL HETIRTAGE PRESERVATION USING POINT – CLOUD DATA

A THESIS SUBMITTED TO
THE FACULTY OF ARCHITECTURE AND ENGINEERING
OF
EPOKA UNIVERSITY

BY

ENDI QOSE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

JUNE, 2024

Approval sheet of the Thesis

This is to certify that we have read this thesis entitled **“Augmented-Reality Enhanced Cultural Heritage Preservation Using Point-Cloud Data”** and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc.Prof. Dr. Arban Uka
Head of Department
Date: June, 25, 2024

Examining Committee Members:

Assoc.Prof.Dr. Dimitrios Karras (Computer Engineering) ________________

Dr. Florenc Skuka (Computer Engineering) ________________

Dr. Shkëlqim Hajrulla (Computer Engineering) ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Surname: Endi Qose

Signature: _______________

ABSTRACT

AUGMENTED-REALITY ENHANCED CULTURAL HERITAGE PRESERVATION USING POINT-CLOUD DATA

Qose, Endi
M.Sc., Department of Computer Engineering
Supervisor: Dr. Florenc Skuka

In today's world there is a growing urgency to protect artifacts and important landmarks due, to environmental factors and human activities. As a result, experts and conservationists are reevaluating their methods realizing the need for approaches to address the challenges facing cultural sites. In the realm of technology advancements, tools like 3D point cloud data, laser scanning and augmented reality (AR) have become essential for preservation efforts. These tools go beyond reacting to issues by taking steps to prevent risks. The use of 3D point cloud data allows for replicas of archaeological sites and structures helping conservationists preserve intricate details that are at risk of damage. AR technologies play a role in enhancing preservation work by offering experiences that support virtual exploration and education promoting greater awareness and respect for our shared past. Central to these pursuits is a commitment to safeguarding cultural heritage that goes beyond simple preservation efforts.

Researchers aim to deepen our understanding of humanity's tapestry through interdisciplinary collaboration and innovative methods. Their goal is to develop solutions that can be passed down through generations ensuring access, to and appreciation of our collective heritage. At the intersection of preserving heritage and embracing technology this academic study offers a broad perspective, on the use of 3D point cloud data and augmented reality. It opens up ways to understand and protect our shared heritage.

Keywords: *Point Cloud, Augmented Reality, Venetian Tower, Structure from Motion, Multi-View Stereo, Image, Scene, Object*

ABSTRAKT

RUAJTJA E TRASHËGIMISË KULTURORE NËPËRMJET REALITETIT TË AUGMENTUAR TË PËRMIRESUAR DUKE PËRDORUR TË DHËNAT POINT-CLOUD

Qose, Endi

Master Shkencor, Departamenti i Inxhinierisë Kompjuterike

Udhëheqësi: Dr. Florenc Skuka

Në botën e sotme ka një urgjencë në rritje për të mbrojtur artefaktet dhe monumentet e rëndësishme për shkak të faktorëve mjedisorë dhe aktiviteteve njerëzore. Si rezultat, ekspertët dhe konservatorët po rivlerësojnë metodat e tyre duke kuptuar nevojën për qasje për të adresuar sfidat me të cilat përballen vendet kulturore. Në fushën e avancimeve të teknologjisë, mjetet si të dhënat e resë me pika 3D, skanimi me lazer dhe realiteti i augmentuar (RA) janë bërë thelbësore për përpjekjet e ruajtjes. Këto mjete shkojnë përtej reagimit ndaj çështjeve duke ndërmarrë hapa për të parandaluar rreziqet. Përdorimi i të dhënave të resë me pika 3D lejon kopjet e vendeve dhe strukturave arkeologjike duke ndihmuar konservatorët të ruajnë detaje të ndërlikuara që janë në rrezik dëmtimi. Teknologjitë AR luajnë një rol në përmirësimin e punës së ruajtjes duke ofruar përvoja që mbështesin eksplorimin dhe edukimin virtual duke promovuar ndërgjegjësim dhe respekt më të madh për të kaluarën tonë të përbashkët. Në qendër të këtyre përpjekjeve është një angazhim për ruajtjen e trashëgimisë kulturore që shkon përtej përpjekjeve të thjeshta të ruajtjes.

Studiuesit synojnë të thellojnë të kuptuarit tonë për tapiceri të humanizmit përmes bashkëpunimit ndërdisiplinor dhe metodave novatore. Qëllimi i tyre është të zhvillojnë zgjidhje që mund të kalojnë brez pas brezi duke siguruar akses, dhe vlerësim të trashëgimisë sonë kolektive. Në kryqëzimin e ruajtjes së trashëgimisë dhe përqafimit të teknologjisë, ky studim akademik ofron një perspektivë të gjerë, mbi përdorimin e të dhënave të resë me pika 3D dhe realitetit të shtuar. Ai hap mënyra për të kuptuar dhe mbrojtur trashëgiminë tonë të përbashkët.

Fjalët kyçe: *Point Cloud, Realiteti i Augmentaur, Kulla Veneciane, Strukturë nga Lëvizja, Stereo me shumë Pamje, Imazh, Skenë, Objekt*

ACKNOWLEDGEMENTS

I want to express my thanks to my supervisor, professors, friends, family and colleagues, for their support during the writing of this thesis. Their guidance and encouragement were crucial in helping me throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

In today's world there is a growing need to protect artifacts and important landmarks like never before. This need arises from the increasing impact of human activities and natural factors which present risks to these priceless relics of our past. As a result, experts and conservationists are finding themselves reevaluating their strategies acknowledging the importance of coming up with ways to address the complex issues that cultural heritage sites are up against.

In the world of advancements 3D point cloud data, laser scanning and AR stand out as crucial tools, for those dedicated to preservation. These technologies go beyond reacting to these issues; they proactively work towards addressing the challenges presented by a changing environment. Utilizing 3D point cloud data allows for the creation of digital replicas of archaeological sites and structures enabling conservationists to capture and safeguard intricate details that may be threatened by environmental factors or human actions.

Figure 1 Example of AR applications

AR technologies complement these preservation efforts by offering experiences that facilitate exploration and learning. By superimposing reconstructions onto real world settings, AR enables people to interact with cultural heritage sites in ways promoting greater awareness and respect for our collective past. Additionally, AR applications can assist in interpreting and presenting discoveries helping researchers communicate stories and insights to a wider audience.

This study aims to not only document and analyze historical structures with great precision, but also enhance their ability to withstand external threats ensuring their longevity, for future generations. Situated at the intersection of heritage preservation and technological advancements, we explore the world of 3D point cloud data and augmented reality. Here the conservation of heritage seamlessly merges with cutting edge technology opening up avenues for understanding and protecting our shared heritage.

At the heart of these pursuits is a commitment to strengthening the preservation of our cultural legacy. Yet this commitment goes beyond preserving; it includes a goal of deepening our knowledge of the complex historical fabric that connects humanity. By leveraging collaboration and innovative approaches researchers aim to develop solutions that can withstand the test of time ensuring ongoing access, to and appreciation of our common heritage.

1.2 Main Problems

The collection of papers being reviewed encounters significant challenges across the wide spectrum of heritage preservation, 3D data processing and augmented reality [1]. These scholarly works embark on a journey to tackle issues related to the preservation of archaeological heritage and historical architecture.

One of the concerns is the necessity to protect and uphold these priceless structures from the harmful impacts of pollution, shifting weather patterns and the rising threat of natural disasters as emphasized in [1] [2] discussing multi-layer residual architecture, for compressing point cloud geometry and merging point cloud data with augmented reality technologies, posing difficulties in crafting accuracy-oriented AR experiences. Additionally, various methods for compressing point clouds are explored to meet the growing demand for approaches to data storage and transmission.

The academic discussion surrounding Istanbul's Land Walls [3] serves as a reminder of the importance of monitoring historical structures especially in preventing potential structural changes caused by environmental events like earthquakes.

Many applications of AR have highlighted the needs in compressing point clouds and integrating these data for an AR based environment, such as a campus navigation app [4] which emphasizes the significance of creating user navigation in AR enabled environments. This effort effectively addresses the need for solutions that improve accessibility and usability within AR frameworks.

These scholarly works converge to address issues related to heritage conservation through data analysis and the use of AR applications. They display a commitment to understanding the obstacles to heritage preservation, while also delving into advances to enrich our understanding of history in a more engaging way.

1.3 Objectives

- To develop strategies in order to protect and preserve historical sites and monuments against environmental degradation and human-caused damage.
- To explore techniques for 3D reconstruction through compressing and analyzing point cloud data by maximizing their storage utility.
- To integrate point cloud data with augmented reality technology to provide immersive and user-friendly experiences in relation to historical sites.
- To use augmented reality tools and APIs to develop engaging experiences that increase the appeal and accessibility of historical sites.
- Support the use of technology while acknowledging privacy, sensitivities and concerns in heritage conservation efforts.

1.4 Significance of the Study

Preserving our heritage in the face of present-day threats is crucial, to upholding our shared history. This research aims to bridge the gap between heritage preservation and latest technology by delving into 3D data, analysis of AR and their applications in preserving these landmarks.

The significance of this study extends beyond conserving structures; it lies in leveraging modern technological advancements to combat contemporary risks. Through exploring strategies for point cloud generation and 3D reconstruction along with their integration with AR technology, this research seeks to revolutionize how we document, analyze and safeguard our legacy. Addressing these challenges not only enriches our knowledge of historical sites but also offers practical solutions for mitigating their susceptibility to natural disasters, pollution and other adverse influences.

Moreover, this study holds importance for its impact on heritage conservation and technological progressions. By clarifying issues and proposing solutions in the realms of data processing, AR incorporation and heritage preservation practices it offers valuable insights for scholars as well as professionals, in cultural heritage fields and technology sectors.

The discoveries, methods and suggested frameworks of this research are geared towards shaping strategies for conserving heritage, sparking lively discussions among scholars and driving practical applications in the field.

Essentially the significance of this study lies in its approach to safeguarding our legacy. The impacts of this work extend beyond investigations laying the groundwork for real world solutions that enhance our comprehension of history while ensuring its preservation.

1.5 Thesis organization

This thesis is divided in 9 chapters: following the introduction, this thesis is organized as follows. Section 2 presents a review of the historical context of the key site that will be investigated and used to implement the solution. Section 3 presents essential technologies and techniques identified throughout the project's literature assessment. Section 4 and Section 5 describes the technique as well as all of the procedures that follow for the 3D reconstruction and AR experience. Section 6 a full explanation of the experimental setup, which are addressed and concluded in sections 7 and 8. Section 9 outlines conclusions and potential future work ideas. Figure 2 illustrates the thesis's organization.

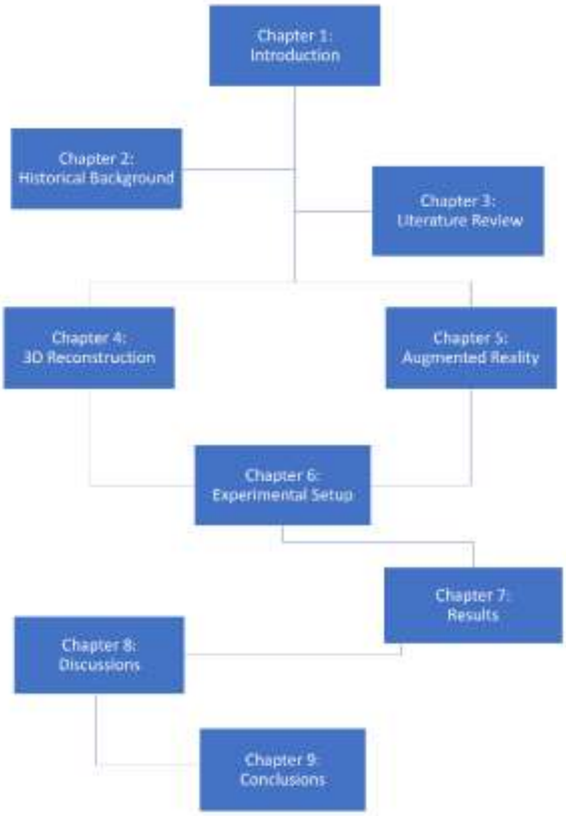


Figure 2 Thesis organization ordered by chapters

CHAPTER 2

HISTORICAL BACKGROUND

2.1 Introduction

This chapter explores the history of Durrës focusing on the importance of the Venetian Tower, as well as studies that aim to digitally preserve the city's historical landmarks. By delving into these topics, we can better grasp how technology today connects the past with the present ensuring that Durrës, or other albanian cities's vibrant history remains honored and safeguarded for generations to come.

2.2 Durrës's Background

Durrës, a city in Albania, has a rich history influenced by landscape modifications, including variations in sea level and environmental changes. It has evolved through various phases as a Greek colony, Roman, Byzantine, Norman, Aragonese, Venetian, Turkish possession, and is now an Albanian city. The city's urban development has been shaped by environmental transformations such as the fluctuating sea levels and the collapse of hillsides. The city's topographical position on a ridge by the coast has been significant in its historical evolution [5].

Durrës, known in antiquity as Epidamnus and Dyrrachium, has a rich history dating back to the seventh century BCE. It served as a significant point of connection between Rome and Constantinople. The city features large protective walls and diverse architecture reflecting its layered history and vibrant civilization. Notably, Dyrrachium was the site of a major battle between Pompey and Julius Caesar and a battle between the Norman and Venetian fleets. The city is home to various ancient sites and monuments like a Roman amphitheater, a Byzantine protective wall, and the Venetian Tower [6].

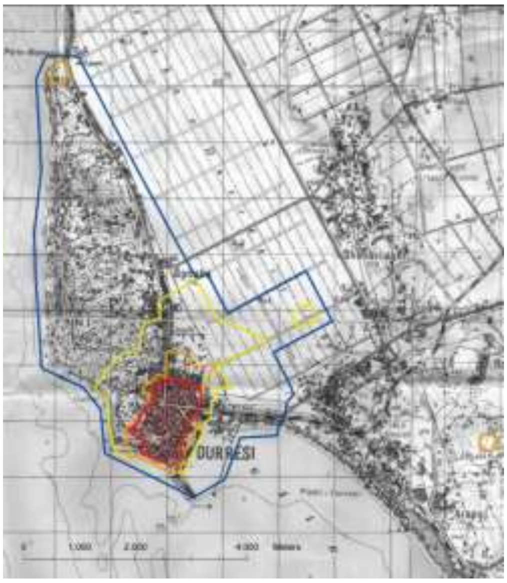


Figure 3 The colored frames showing the historical sites being endangered (blue to red scaling from less to much in risk)

However, the city's trajectory took a downturn following the invasion by the Turks in the 1500s. Durrës faced a period of decline, witnessing a drastic shift in its fortunes. The invasion led to a challenging phase, causing a decline in its economic and social vitality. This tumultuous period almost resulted in the abandonment of the city as it struggled to cope with the aftermath of the invasion.

Nevertheless, Durrës exhibited resilience and gradually began to reclaim its importance in the 17th century. Over time, it embarked on a journey of revival, slowly recovering from the aftermath of the invasion. This resurgence marked a pivotal turning point for the city, as it sought to re-establish itself as a significant center, albeit through a gradual process of rebuilding and resurgence in various facets of life.

2.3 Venetian Tower

Durrës's most unique artifact, preserved over centuries, the Venetian Tower [5] is part of the city's historical infrastructure holding a significant structure with a rich history, reflecting the city's evolution through various periods. The tower's infrastructure likely exhibits characteristics influenced by the city's urban development and historical transformations, with a strategic location linked to important topographical features and historical events.

2.3.1. Preservation Challenges

The city's archaeological heritage faces challenges due to uncontrolled building activities without a proper town plan, and natural disasters, such as the earthquake of November 2019. These factors endanger not the walls of the tower but its iconic Amphitheater. It is essential to make efforts towards safeguarding and documenting the sites to ensure that parts of the Venetian Tower are preserved for future generations.

Figure 4 Images showing the impact of natural disasters on Venetian Tower

2.4 Urban Planning Transformation

Durrës with its origins has a charm that has been transformed over time by Italian architects aiming to modernize and incorporate European elements into its urban design. The citys architectural journey reflects a fusion of elements, with interventions that aim to redefine its visual appeal.

The transformation of Durrës architecture began with the development of plans in 1934, which were later revised in 1942 [7]. These plans focused on revitalizing and constructing features, such, as public buildings, hotels and residential complexes. The overarching goal was to incorporate styles into the cityscape aiming to give Durrës a distinctive European flair.

Specific improvements were targeted at areas in Durrës including the town hall square, main boulevard and mercantile street. These locations were central to implementing changes inspired by rationalist design principles. The rationalist approach prioritized functionality, geometric shapes and simplicity. Reflecting the trends in European architecture during that era.

Durrës urban scenery underwent a purposeful transformation by blending modern architectural elements with historical influences. This fusion of remnants with rationalist aesthetics resulted in a unique architectural identity for Durrës. The influence of architecture is still evident, throughout the citys landscape showcasing a harmonious blend of historical legacy and intentional design interventions.

The deliberate blend of styles showcases the layers of Durrës history serving as a visible symbol of its changing identity, in the rich fabric of Albanias cultural and historical legacy.

2.5 Relevant Albanian Studies

This chapter explores the history of Durrës focusing on the importance of the Venetian Tower, as well as studies that aim to digitally preserve the city's historical landmarks. By delving into these topics, we can better grasp how technology today connects the past with the present ensuring that Durrës, or other albanian cities's vibrant history remains honored and safeguarded for generations to come.

2.5.1. Preservation and Documentation

The Inter-Link Project for Elbasan Citadel [8] focuses on addressing the preservation and restoration challenges associated with the Elbasan fortified citadel. The problem at hand necessitates innovative approaches, prompting the adoption of advanced methods and technologies such as photogrammetry, laser scanning, and geo-radar for comprehensive surveying and rendering of attributes.

Spearheaded by Prof. Roberto Pierini, the project takes a holistic approach aimed at studying, preserving, and leveraging the heritage of the Elbasan citadel. Through the integration of cutting-edge technologies like photogrammetry, laser scanning, and geo-radar, the project endeavors to achieve a thorough understanding of the citadel's architectural intricacies and historical significance, facilitating informed preservation and restoration efforts.

2.5.2. Involvement of Modern Techniques

The Structural Assessment of the Leaden Mosque in Berat [9] addresses the critical issue of deterioration faced by cultural heritage monuments, with a particular focus on the Leaden Mosque in Berat. The problem underscores the urgent need for a comprehensive understanding of the structural integrity and condition of the mosque to guide effective preservation efforts.

The methodology used for this research is based on the assessment of visible “symptoms” that structural defects and distresses had caused throughout the structure. In order to evaluate the structural performance of an existing masonry structure, its geometry, the characteristics of its masonry texture, typical cross section, the actual conditions of the joints, physical, chemical and mechanical properties of stones and mortar should be known.

Figure 5 Point cloud of the Leaden Mosque in ***[9]***

In response, the project adopts a case study approach, employing visual inspection techniques alongside the development of tailored restoration strategies. Beyond mere structural assessment, the project delves into broader considerations, including the preservation of cultural, historical, and architectural values inherent in the mosque. This holistic approach emphasizes the importance of balancing preservation needs with the inherent significance of the cultural heritage site.

Leveraging visual inspection methodologies and customized restoration strategies, the project aims to safeguard the Leaden Mosque's structural stability while upholding its cultural and historical legacy for future generations.

2.5.3. Structural Assessment Strategies

The Conservation and Analysis of Gjirokastra’s Vernacular Heritage project [10] tackles the pressing issue of preserving and safeguarding the unique characteristics inherent in Gjirokastra’s vernacular heritage while addressing emerging threats to its transmission. To address this challenge, the project adopts a multidisciplinary approach encompassing various methodologies such as direct survey, on-site observation, interviews, laser scanning, and photogrammetry. These techniques are employed synergistically to gain a comprehensive understanding of the heritage and devise effective conservation strategies.

Beyond documentation and analysis, the project delves into describing the distinctive features of Gjirokastra, contextualizing its significance within the broader landscape of cultural heritage preservation. Additionally, the project highlights initiatives like the 3D Past project and discusses potential conservation strategies and future measures aimed at ensuring the long-term sustainability of Gjirokastra’s vernacular heritage. Leveraging a suite of technologies including direct survey, on-site observation, interviews, laser scanning, and photogrammetry, the project endeavors to not only conserve but also promote the enduring legacy of Gjirokastra’s cultural heritage for generations to come.

CHAPTER 3

LITERATURE REVIEW

3.1 Introduction

In today’s world, preserving our cultural heritage faces significant challenges due to environmental vulnerabilities and human activities. Amidst these issues the fusion of technology with traditional conservation efforts offers hope. This review explores the connection between heritage preservation, technological advancements and the fast-evolving realm of 3D data analysis. It critically analyzes a selection of works that discuss the intertwined areas of safeguarding marvels, preserving historical edifices and utilizing technologies, like 3D point cloud data, mesh generation and augmented reality to enhance our comprehension, documentation and protection of cultural heritages.

3.2 Theoretical Framework

The theoretical framework of the study extends to how Digital Heritage and 3D Reconstruction technologies were combined to preserve the sites. Digital heritage uses tools to document, preserve and share objects and places, increasing access and ensuring their long-term preservation. On the other hand, 3D Reconstruction generates three-dimensional models of physical environments and objects that enable, in depth analysis, virtual interaction and preservation of historical authenticity. These concepts form a basis for exploring the application of digital methods in safeguarding Durrës rich historical heritage, including notable structures like the Venetian Tower.

This framework shapes the research approach. Supports the evaluation of practices and advancements in digital conservation, within the Albanian setting.

3.2.1. Digital Heritage

The framework discussed in [11] offers a perspective, on how scholars explore the merging of heritage with digital technology. It acts as a concept in understanding the connection between digital tools and efforts towards preserving culture. Within this framework researchers explore definitions and ideas surrounding heritage to form a cohesive understanding of its role and importance in today's world. By highlighting the capabilities of the technologies encouraged by this framework, critical examination is applied into how digitization influences the recognition, safeguarding and communication of cultural heritage.

The research on endangered heritage sites [12], in Kandovan, Iran took an approach by creating an integrated framework using virtual reality technology. This framework aimed to capture and replicate the endangered heritage sites to assess how new urban developments might impact their existence. The study utilized an approach to gather data established in a narrative-based framework and present it interactively through a virtual reality project. The validity of the framework was confirmed through field data collection spanning five years from Kandovan Iran's unique case study. The research outcomes effectively raised awareness encouraging involvement and action by challenging the existing state of the heritage sites.

3.2.2. 3D Reconstruction

The theoretical framework discussed in [12] relies on utilizing homology, a tool in topological data analysis (TDA). Persistent homology serves as an approach to assessing the characteristics of a dataset by examining various scale values to derive qualitative insights from the data. This technique tracks how the homology of structures evolves with increasing scale parameters, as shown in Figure 6, thereby identifying which features persist through changes. Moreover, the paper also explores the application of a scale kernel-based learning method [12], which is linked to scale space theory.

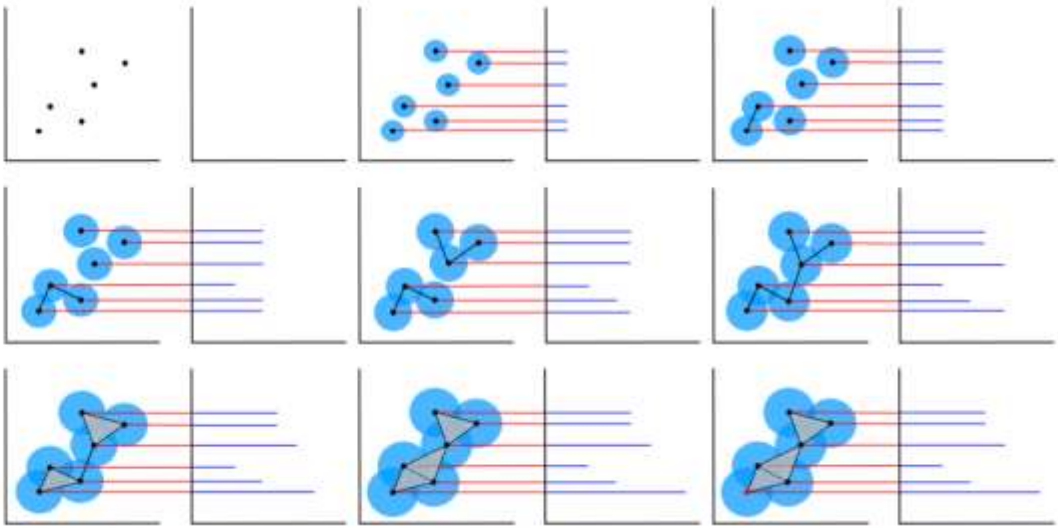


Figure 6 Effects of persistent homology of a scale parameter defining the neighborhood radius

The study, in [13] is centered around the idea of Approximate Intrinsic Voxel Structure. AIVS, Figure 7, introduces a voxel box-based system for organizing point clouds enabling simplification while controlling point distances intrinsically. This approach merges voxel structure with FPS to effectively streamline point clouds while preserving uniform characteristics. Additionally, it provides a simplification strategy to cater to needs, such, as sampling based on curvature and preserving sharp features.

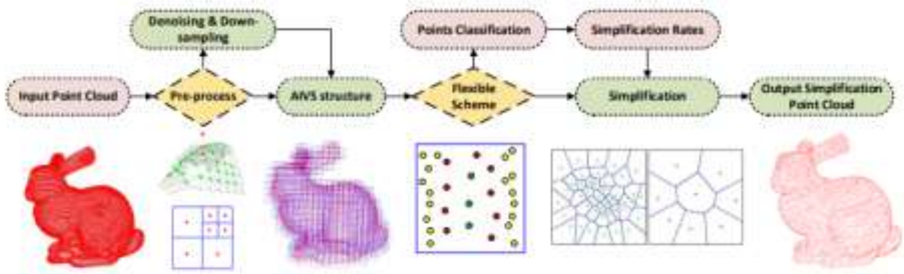


Figure 7 The pipeline proposed by AIVS for point cloud simplification

The study described in [14] introduces superpoint graphs as a way to represent 3D point clouds. These graphs capture the shapes within the point cloud. Include edge features that show how different parts of objects relate to each other. The approach uses learning techniques such as PointNets for embedding superpoints and graph convolutions for context-based segmentation. It also includes ideas from DenseNet about combining states and GRUs to disseminate information. The goal is to address the challenges of object classification in 3D point cloud data.

The authors discuss a segmentation method for point clouds that uses region growing together with neighborhood search, filter sampling, Euclidean clustering, and region growing steps to improve segmentation accuracy and efficiency. This technique is specially adapted for indoor point cloud data. Demonstrates enhanced segmentation results. The potential use of this algorithm in home robots is also highlighted.

3.3 Methodologies and Approaches

3.3.1. Reconstruction Methods

Structure from Motion is a photogrammetric range imaging technique used to estimate three-dimensional structures from two-dimensional image sequences that may be coupled with local motion signals. SfM assumes that the camera capturing the images is moving through a static scene. By identifying and tracking feature points across multiple overlapping images, SfM can determine the camera's motion and the 3D positions of the feature points [13] [15].

The process involves:

- Feature Detection and Matching: Detecting distinct points in the images (features) and matching these points across different images.
- Camera Pose Estimation: Estimating the position and orientation of the camera for each image.
- 3D Point Cloud Generation: Using the estimated camera poses to triangulate the 3D positions of the matched feature points, thus creating a sparse 3D point cloud.

- Optimization: Refining the 3D structure and camera poses through bundle adjustment to minimize reprojection error.

Multi-View Stereo (MVS) builds upon the initial reconstruction provided by SfM to produce a detailed and dense 3D model [16] [17]. While SfM generates a sparse point cloud, MVS densifies this cloud by leveraging the multiple views of the scene.

The MVS process involves:

- Dense Matching: Identifying corresponding points across multiple images to increase the density of the 3D reconstruction.
- Depth Map Estimation: For each image, estimating a depth map that represents the distance from the camera to the scene points.
- Fusion: Combining the depth maps from all images to form a consistent and dense 3D surface model.
- Surface Reconstruction: Creating a mesh or other surface representation from the dense point cloud, often using techniques like Poisson surface reconstruction or volumetric methods.

3.3.2. Data Acquisition Methods

A novel labeled dataset of driving sequences was generated specifically for the purpose of conducting experiments. This dataset consists of video sequences recorded from a moving vehicle, and each sequence has been meticulously annotated with ground-truth semantic segmentation [15]. The annotation process involved the manual application of per-pixel semantic labels to every 30th frame of the high-definition videos. Subsequently, sequences from this meticulously annotated dataset were utilized as either training or testing data, playing a pivotal role in the empirical evaluation and refinement of the proposed methodologies.

Red dots represent points in the 3D point cloud, along with their projections from world space to the camera's image plane. Any feature data linked to a 3D point also appears on the image plane and is summed. The algorithm evaluates each pixel by sliding over them with a randomized decision forest, as indicated by the yellow and green crosses. Feature responses are computed at a set 2D offset (indicated by the white dashed line) and rectangle *r*. Two sample rectangles, r_1 (yellow) and r_2 (green), with their respective truncated pyramids p_1 and p_2, are shown here. Rectangle r_1 is positioned up and to the left of pixel (x_i, y_i), utilizing the context of f_c to determine the category at (x_i, y_i). Rectangle r_2 is centered on pixel (x_j, y_j) (no offset), thus pooling the local information of f_x at that point, shown in Figure 8.

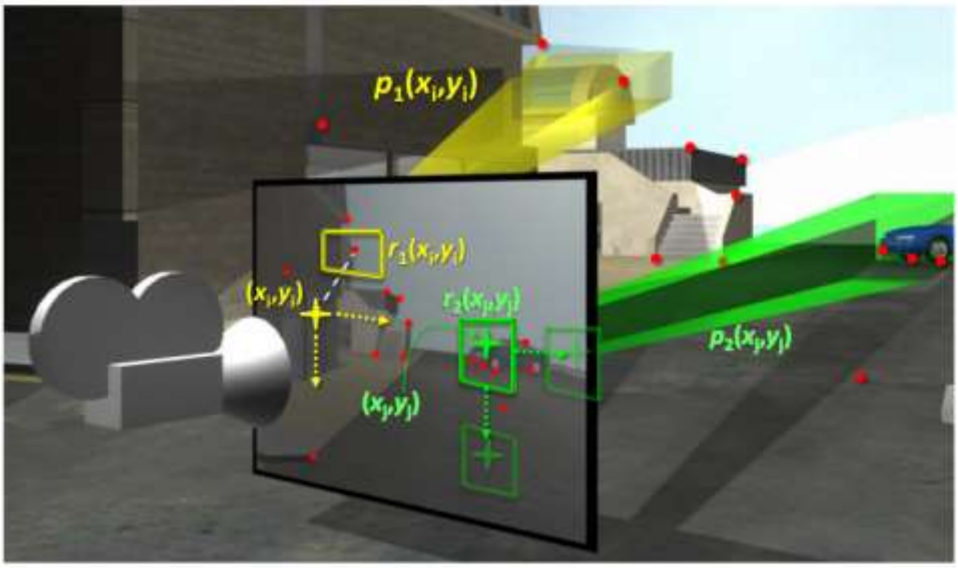


Figure 8 Principle of 3D projections in 2D

3.3.3. Data Processing

Conventional structure-from-motion (SFM) techniques were used to create a 3D point cloud from video sequences recorded from moving vehicles in [15]. The primary movement in these sequences determined the positioning of the camera, thereby creating a 3D point cloud relative to all identified 2D features accounting for any irregularities. 2D image feature tracking was facilitated through Harris-Stephens corners using localized normalized cross correlation with a 20 × 20-pixel patch tracked over time within a search window at 15% of the image size. Differences between frames were attributed only to camera movement.

The process of extracting structure from ego-motion involves using methods to automatically generate a 3D point cloud from video sequences captured by a moving vehicle. The dominant motion in these sequences defines the position of the cameras in the world by creating a relative 3D point cloud for all tracked 2D features, including outliers. It begins by tracking image features such as Harris-Stephens corners using localized normalized cross-correlation to monitor pixel patches over time within a defined search window. Captured by a camera mounted on a vehicle, the differences between the images are attributed to movement alone, for three-dimensional reconstruction.

The assumption made in this process helps to calculate a world point for each tracked 2D feature using the matrices determined in a pre-processing step by adjusting the solutions uniformly until the correlation falls below 0.97 [15].

A set of five basic motion and texture cues were suggested as cues to the categories of objects present in a specific scene. These cues were then transferred from the 3D point cloud to the 2D image to create a random decision forest classifier that enables a segmentation. In particular, the algorithm demonstrated how semantic segmentation can be achieved based on motion-derived 3D world structure, eliminating the need for image-based descriptors.

3.3.4. AR Development

In the real-time occlusion handling method, the process for selecting the occluding object involves an interactive interface where the user labels some pixels as foreground and others as background. All the pixels in the image are divided into two classes, object and background, based on the hidden information provided by the labeled pixels. This interactive segmentation method allows the user to specify the occluding object by selecting it in the first frame. The proposed method [16] then finds the object boundary, even in complex scenes, to determine the occluding object.

3.4 Case Studies

The project in [16] addresses the challenge of enhancing cultural experiences and tourism by focusing on the promotion and dissemination of architectural heritage. To tackle this challenge, the approach involves thorough testing and validation of ICT tools. These tools aim to create virtual networks that capture shared historic, environmental, and technical characteristics among architectural landmarks. Key to this effort is integrating advanced technologies such as 3D GIS for detailed spatial representations and photo-realistic models.

Figure 9 Images showing the building in AR (right) and in real life (left)

Since all the virtual tours, as well as the digital contents from the hot spots and the highlights, are stored as records of the WebGIS environment, the expert users who are more interested in specialized data.

Additionally, the project incorporates immersive VR and AR digital environments to provide engaging experiences for visitors. By leveraging these innovative tools, the project seeks to make architectural heritage more accessible, offering immersive journeys that deepen understanding and appreciation of cultural significance and historical context. Through this integration of technology and heritage, the project aims to enrich tourism experiences while preserving and celebrating architectural treasures for future generations.

3.5 Open-Source Projects

3.5.1. Digital Heritage

To overcome the challenge of incorporating affordable 3D point cloud sensors into AR systems, we explored the application discussed in [18] within the AR framework. Their main focus was on simplifying the calibration and registration processes. Additionally, we tackled the face alignment issue using 3D point cloud data together with open-source libraries. Our goal is to improve the integration of these elements and optimize the collaboration between 3D point cloud sensors and AR systems leading to efficient and budget-friendly AR applications.

3.5.2. Digital Heritage

To meet the challenge of displaying large point cloud data and simplifying the task of organizing it into a multi-resolution octree, for tools like Potree, an extension of Potree was suggested in one study [19]. This improved version includes features such as a download tool. A search function for geographic names names a package of measurement tools and a 2D orientation map. The project uses a divide-and-conquer approach to reduce the time required to build octets, especially when dealing with datasets. The main goal is to improve the accessibility and user friendliness of point cloud data.

In addition, the integration of functions such as geographic name search and measurement tools enhances the capabilities of the Potree tool. Implementation of an octree data structure acts as a core component in enhancing visualization and interaction with extensive point cloud datasets. Through these advancements, the project aims to simplify working with and analyzing large cloud data across platforms.

3.5.3. Digital Heritage

In response to the demand for preserving cultural heritage items using 3D scanning tools, which often produce detailed point clouds that require extensive post-processing, the methodology described in [20] provides a way to create simplified 3D models. The goal is to remove the points while ensuring that the integrity of the model remains intact. Focusing on incorporating these simplified models into on-device AR applications, the research performs an evaluation of various methods for optimizing 3D meshes.

This evaluation demonstrates the effectiveness and reliability of these techniques through the examination of objects. The proposed method presents an approach for the digitization and integration of cultural heritage items into applications. Applying this methodology to 3D-scanned Dacian artefacts proves its versatility and utility in heritage preservation and presentation, in AR environments.

CHAPTER 4

3D RECONSTRUCTION

4.1 Structure from Motion

Expanding on the core concepts of computer vision, especially highlighting the role of Structure from Motion, our research framework incorporates methods to tackle the various challenges in 3D reconstruction technologies. SfM is well known for its ability to create three environments from a series of two images or video frames, which plays a key role in equipping computational systems with essential perceptual skills for understanding spatial complexities [21]. This methodological approach not only drives progress in help of AR applications, but also broadens its scope to diverse fields like heritage preservation.

At the heart of computer vision lies Structure from Motion, a technique used to reconstruct the three-dimensional layout of an environment, using a sequence of two-dimensional images or video frames. This methodology plays a role in enabling systems to perceive and interpret the intricate details of three-dimensional spaces making it valuable across various fields such as robotics, augmented reality and autonomous navigation.

The process of SfM involves an orchestrated series of steps each contributing uniquely to the ultimate goal of reconstructing three dimensional scenes. It begins with identifying and extracting features or points, within the two images. These characteristics, which can include shapes, like corners, edges or changes, in texture form the fundamental elements on which the rest of the reconstruction process is based.

4.1.1. Steps for Reconstruction

For each image I_i, Structure from Motion identifies sets

$$\boldsymbol{F_i = \{(x_j, f_j) \mid j = 1 \dots N_{F_i}\}}$$

of local features, where each feature is located at $x_j \in \mathbb{R}^2$ and characterized by an appearance descriptor f_j. These characteristics should remain consistent when subjected to changes in lighting and perspective, allowing SfM to reliably identify them in images. The Scale Invariant Feature Transform and its variations are widely regarded as the gold standard for extracting features that are robust. On the other hand, binary features provide enhanced efficiency but come at the cost of reduced robustness.

Figure 10 Keypoints extracted from an image taken of Durrës's Venetian Tower

Next, SfM discovers images that capture the same scene part by leveraging the features F_i as an appearance description of the images. The naive approach tests every image pair for scene overlap; it searches for feature correspondences by finding the most similar feature in image I_a for every feature in image I_b, using a similarity metric comparing the appearance f_j of the features. This approach has a computational complexity of $O\left(N_I^2 N_{F_i}^2\right)$ and is prohibitive for large image collections. A variety of approaches tackle the problem of scalable and efficient matching. The output is a set of potentially overlapping image pairs $C = \{\{I_a, I_b\} | I_a, I_b \in I, a < b\}$ and their associated feature correspondences $M_{ab} \in F_a \times F_b$.

Figure 11 Matching features between two images

In the phase it checks the image pairs *C* that may overlap. Matching depends on how things look. It's not certain that matching features really represent the same spot, in a scene. Hence SfM double checks the matches by attempting to figure out a transformation that links feature points across images, through geometry. Depending on the spatial configuration of an image pair, different mappings describe their geometric relation. A homography *H* describes the transformation of a purely rotating or a moving camera capturing a planar scene.

Epipolar geometry describes the relation for a moving camera through the essential matrix *E* (calibrated) or the fundamental matrix *F* (uncalibrated), and can be extended to three views using the trifocal tensor. If a valid transformation maps a sufficient number of features between the images, they are considered geometrically verified. Since the correspondences from matching are often outlier-contaminated, robust estimation techniques, such as RANSAC, are required. The output of this stage is a set of geometrically verified image pairs $\bar{C}$, their associated inlier correspondences $\bar{M}_{ab}$, and optionally a description of their geometric relation G_{ab}.

Figure 12 Projective geometry between two images

Starting from a metric reconstruction, new images can be registered to the current model by solving the Perspective-n-Point (PnP) problem using feature correspondences to triangulated points in already registered images (2D-3D correspondences). The PnP problem involves estimating the pose P_c and, for uncalibrated cameras, its intrinsic parameters. The set P is thus extended by the pose P_c of the newly registered image. Since the 2D-3D correspondences are often outlier-contaminated, the pose for calibrated cameras is usually estimated using RANSAC and a minimal pose solver. For uncalibrated cameras, various minimal solvers or sampling-based approaches exist.

4.1.2. Bundle Adjustment

Image registration and triangulation are separate procedures, even though their products are highly correlated – uncertainties in the camera pose propagate to triangulated points and vice versa, and additional triangulations may improve the initial camera pose through increased redundancy. Without further refinement, SfM usually drifts quickly to a non-recoverable state. Bundle Adjustment (BA) [22] is the joint non-linear refinement of camera parameters P_c and point parameters X_k that minimizes the reprojection error:

$$E = \sum_j \rho_j \left(\left\| \pi(P_c, X_k) - x_j \right\|_2^2 \right)$$

using a function π that project scene points to image space and a loss function ρ_j to potentially down-weight outliers.

Levenberg-Marquardt is the method of choice for solving BA problems. The special structure of parameters in BA problems motivates the Schur complement trick, in which one first solves the reduced camera system and then updates the points via back-substitution. This scheme is commonly more efficient, since the number of cameras is usually smaller than the number of points.

There are two choices for solving the system: exact and inexact step algorithms. Exact methods solve the system by storing and factoring it as a dense or sparse matrix with a space complexity of $O(N^2P)$ and a time complexity of $O(N^3P)$. Inexact methods approximately solve the system, usually by using an iterative solver, like preconditioned conjugate gradients (PCG), which has $O(NP)$ time and space complexity.

Indirect methods are preferred for handling a number of cameras due, to the cost associated with direct algorithms in such scenarios. While sparse direct approaches can simplify things for problems, they become impractical when dealing with extensive and unstructured photo collections that typically have more interconnected elements. In these instances, indirect algorithms prove to be more effective. When it comes to photos a significant amount of effort is put into optimizing similar images.

Once distinct features are identified, the focus shifts towards matching these features across images. This matching process plays a role in the SfM process as it helps reveal the relationships, between features observed in various frames of the image sequence. By using matching algorithms, the system aims to understand how these key features evolve over time and interact spatially throughout the image dataset.

After establishing feature correspondences, the next step involves estimating camera poses to determine the positions and orientations of the cameras used to capture the image sequence.

This project showcases capabilities, where the system aims to track the movement of cameras over time in a series of images understanding how they move and rotate in a three-dimensional space.

Building on the information gathered from estimating camera positions the process smoothly moves into creating a 3D representation of the scene by combining matching features and estimated camera positions into 3D point clouds. This step is, at the core of SfM techniques, for ensuring accuracy and reliability in reconstructing the scene.

Alongside reconstruction there is an adjustment process called bundle adjustment that fine tunes and brings together all elements of the reconstructed scene. By optimizing camera positions and 3D points coordinates carefully the computational system works to reduce errors in projecting images onto this 3D representation aiming for a realistic portrayal of the scene captured in the image sequence.

4.2 Multi-View Stereo

The aim of multi-view stereo is to create a full 3D model of an object using a set of images captured from known camera angles. In years several advanced algorithms have been developed, leading to advancements, in technology. However, the absence of datasets poses challenges in evaluating these algorithms and directing research efforts effectively. In contrast binocular stereo, which aims to generate depth maps from image pairs has seen progress due to the availability of databases containing ground truth data, for algorithm comparison and improvement.

4.2.1. Scene representation

Various methods [23] can be used to depict the shape of an object or scene with view algorithms opting for voxels, level sets polygon meshes or depth maps. While some algorithms stick to one representation others utilize ones at stages of the reconstruction process.

Many techniques choose to represent geometry on a sampled 3D grid (volume) either as an occupancy function or, as a function that encodes the distance to the nearest surface. These 3D grids are favored for their simplicity, consistency and capability to approximate any surface shape.

Polygon meshes portray a surface as a collection of interconnected planar facets. They are efficient for storage and rendering purposes. Are commonly used as an output format in view algorithms. Meshes excel in visibility calculations. Serve as the representation, in certain algorithms.

The process of expanding the surface in view stereo reconstruction entails gradually enlarging flat circular disks in a direction tangent to cover the entire surface. This expansion halts when a disk exceeds a threshold angle [24] during the phase. The spatial sampling density, which governs how disk centers are distributed on the surface of the scene is modified based on how a new disks moved from its parent disk. This modification ensures that the sampling density adjusts automatically to match the image resolution enabling reconstruction of geometric features.

4.2.2. Photo-consistency measure

Many methods have been suggested [23] to assess how well a reconstruction matches a set of input images visually. Most of these methods work by comparing pixels, in one image with pixels in images to determine their correlation referred to as photo consistency measures. The choice of method is not tied to an algorithm. Can usually be interchanged between different techniques. Photo consistency measures are classified based on whether they're defined in scene space or image space.

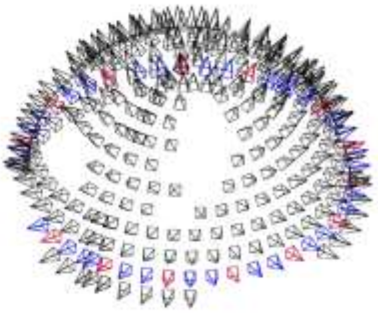


Figure 13 Camera positions and orientations for the temple dataset in ***[7]***

Scene space measures involve taking a point, patch or volume of geometry projecting it onto the input images and assessing the level of agreement between these projections. One way to measure this agreement is by looking at the variance of the projected pixels in the input images. Alternatively, some methods compare pairs of images. Use window matching metrics, like sum of squared differences or normalized cross correlation. A unique aspect of scene space window-based techniques is that the current estimation of geometry can influence the size and shape of the window used for comparison.

4.2.3. Reconstruction Steps

There are generally four types of view stereo algorithms. The first type involves calculating a cost function on a volume and then creating a surface based on this volume. One example is the voxel coloring algorithm, which processes the volume, in one go computes costs and reconstructs voxels with costs below a threshold simultaneously. Other approaches vary in how they define the cost function and extract the surface. Some methods use a Markov Random Field (MRF). Employ max flow or multi way graph cut to determine an optimal surface.

The second type of techniques evolves a surface iteratively to reduce or minimize a cost function. This category includes methods that utilize voxels, level sets and surface mesh. Space carving and similar techniques gradually eliminate voxels, from a volume. Variants of this method allow for both adding and deleting voxels to minimize an energy function. Level set techniques focus on minimizing a set of equations within a volume.

Similar, to space carving techniques level set methods usually begin with a volume that gradually decreases inward. What sets them apart from space carving methods is their ability to expand locally when necessary to reduce an energy function. Alternatively, some methods depict the scene as a changing mesh influenced by external forces.

CHAPTER 5

AUGMENTED REALITY

5.1 Techniques

5.1.1. Object Alignment and Placement

In augmented reality, object alignment entails positioning virtual objects within the user's physical environment in a manner that seamlessly integrates with their surroundings. This process involves accurately determining the spatial orientation and location of virtual objects relative to real-world points of reference.

One fundamental principle underlying object alignment in AR is the utilization of computer vision techniques for tracking and anchoring virtual objects to real-world surfaces or features. However, challenges arise when these tracking algorithms encounter issues such as occlusion, poor lighting conditions, or insufficient visual cues. In such cases, manual intervention becomes necessary to align virtual objects accurately.

To address these challenges, [25] proposes interaction algorithms aimed at simplifying the manual alignment process. Techniques like HoverCam and SHOCam leverage intuitive gestures or input mechanisms to enable users to manipulate virtual objects effectively. Additionally, the application of a Signed Distance Field provides a mathematical representation of the space surrounding virtual objects, facilitating precise alignment even in the absence of robust tracking data.

By incorporating these interaction algorithms to streamline the user interface for manual 3D pose alignment within AR systems, WE reduce the time and effort required for users to align virtual objects manually, thereby enhancing the overall user experience in AR applications.

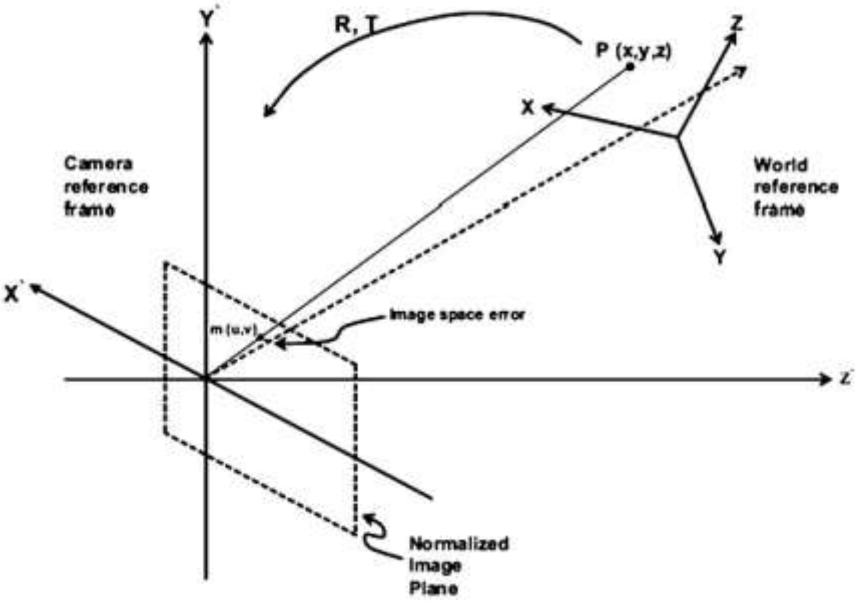


Figure 14 Point constraints for the camera pose in ***[8]***

5.1.1.1. Challenges in AR Object Alignment and Placement

The interactive alignment of large virtual objects within augmented reality systems poses several challenges. Firstly, tracking failures represent a significant hurdle. AR systems heavily rely on 3D tracking technologies to overlay virtual objects onto the real world. However, no contact-less 3D tracking method is entirely infallible, resulting in potential misalignment in 3D poses despite possessing full knowledge of the geometry.

One of the fundamental aspects of AR is accurately registering virtual objects within the real world. This involves determining the precise position and orientation of the user's viewpoint or device. The mathematical foundation for this principle is established through the concept of camera pose estimation, which is often achieved using homography and the pinhole camera model.

The pinhole camera model can be represented by the collinearity equation, which defines the relationship between 3D points in the real world and their 2D projections on the camera's image plane:

$$h_i = \frac{1}{r_3^T p_i + t_z}(Rp_i + T)$$

where $h_i = (u_i, v_i, 1)^T$ is the homogeneous coordinate of the image point, p_i is the 3D point, R is the rotation matrix, and T is the translation vector.

The rotation matrix R and the translation vector T can be expressed as:

$$R = \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix}, T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

Another crucial concept is the image space error, which measures the difference between the observed image points and the projected points. The error is given by:

$$E_{p_i} = \sqrt{\left(\hat{u}_i - \frac{r_1^T p_i + t_x}{r_3^T p_i + t_z}\right)^2 + \left(\hat{v}_i - \frac{r_2^T p_i + t_y}{r_3^T p_i + t_z}\right)^2}$$

where u_i and v_i are the coordinates of the observed image points.

Environmental conditions further compound these challenges. Variations in lighting, system noise, or issues related to visibility and occlusion can disrupt the performance of pose tracking algorithms, leading to temporary tracking failures and subsequent misalignment.

Particularly problematic are scenarios involving large objects. In instances where the object of interest surpasses the camera's field of view, tracking algorithms are more prone to failure. This issue is especially pronounced in tasks such as inspecting extensive objects supported by AR systems, like a 300-meter-long container ship, where limited features and poses hinder the computer vision method's ability to detect anchoring points effectively.

Moreover, user interface limitations contribute to the complexity of interacting with and aligning large virtual objects within AR systems. Utilizing a 2D input surface, such as a tablet, for controlling all six degrees of freedom poses inherent challenges.

Rendering in AR involves the process of displaying the virtual content in a way that it blends seamlessly with the real world. This requires sophisticated algorithms to handle lighting, shadows, occlusions, and perspective correction. The rendering equation is a fundamental concept that describes the transfer of light in a scene:

$$L_o(x, \omega_o) = L_e(x, w_o) + \int_{\Omega} f_r(x, \omega_i, \omega_0) L_i(x, \omega_i)(\omega_i \cdot n)\, d\omega_i$$

where L_o is the outgoing light, L_e is the emitted light, f_r is the bidirectional reflectance distribution function (BRDF), L_i is the incoming light, ω_i and ω_o are the incoming and outgoing directions, and n is the surface normal [26].

To achieve realistic rendering, AR systems often employ techniques such as environment mapping and shadow mapping. Environment mapping involves capturing the surrounding environment and using it to render reflections on virtual objects. Shadow mapping ensures that virtual objects cast shadows on real-world surfaces, enhancing the depth and realism of the scene.

Manual alignment becomes necessary when the anchoring algorithm of the tracking system falters. Users may need to manually adjust the pose of virtual models to better match reality, a task that becomes particularly daunting when adjustments in more than two degrees of freedom are required.

The impact of misalignment in 3D poses extends to the user experience realm. Such discrepancies can detract from the potential benefits of AR systems. Consequently, there is a pressing need for the development of improved interaction algorithms to facilitate manual alignment of virtual objects, thereby enhancing the overall user experience in AR applications.

5.1.2. Plane Recognition

Plane recognition constitutes a fundamental concept in augmented reality, essential for accurately overlaying virtual content onto real-world surfaces. In AR, plane recognition algorithms analyze the visual input from a device's camera to identify and characterize flat surfaces, such as walls or tabletops, within the environment. Once recognized, these planes serve as the foundation upon which virtual objects or annotations can be integrated, ensuring spatial coherence and enhancing the user's immersive experience. Despite advancements in plane recognition technology, variations in lighting conditions, surface texture, and occlusions can present challenges, potentially affecting the accuracy and stability of AR applications reliant on plane detection. Nonetheless, ongoing research and advancements aim to refine and optimize plane recognition algorithms, thereby bolstering the efficacy and utility of AR systems across diverse use cases.

AR technology enables the measurement of vertical surfaces, as demonstrated in the context provided by the authors [27] who developed an AR application utilizing Apple's ARKit 1.5, incorporating vertical surface recognition capabilities. This application permits users to measure surface areas by strategically placing virtual anchors to delineate the area of interest. The overarching objective of the thesis was to assess the feasibility and accuracy of measuring vertical surfaces through AR.

Figure 15 Vertical and horizontal planes being extracted in AR experience

The advantages of plane detection-based object recognition lie in its pivotal role within the domain of computer vision, particularly in facilitating markerless augmented reality AR. Plane detection serves as a fundamental feature, enabling developers to comprehend the geometric layout of the surrounding environment and integrate three-dimensional content into real-world settings. However, one of the primary challenges in AR implementation is maintaining real-time synchronization between physical environments and digital objects, necessitating efficient tracking of surfaces such as walls, floors, and objects [28].

Future applications of object detection span diverse domains, including damaged building of cultural heritage. Nonetheless, while plane recognition has witnessed significant advancements, challenges persist, such as balancing speed and accuracy in real-time processing. Additionally, the lack of annotated datasets and the need to address multiple scales and class imbalances pose ongoing hurdles. Despite these challenges, ongoing research and advancements aim to refine plane recognition algorithms and overcome existing limitations, ensuring their efficacy across a wide range of applications.

5.1.3. Object Occlusion

AR technology, now accessible through mobile devices empowered by sophisticated tracking algorithms like Google's ARCore and Apple's ARKit, has diverse applications spanning gaming, shopping, learning and most importantly cultural preservation, which is the main issue of this thesis. However, existing AR implementations often struggle with occlusion, where virtual objects fail to be hidden behind real-world elements, leading to a disjointed experience.

The process of handling occlusion in augmented reality can be broken down into several key steps, each involving specific mathematical computations. One of the primary methods to address occlusion involves using key-views and reconstructing the 3D occluding boundary from these views. The mathematical formulation begins with the calculation of camera parameters and motion estimation, followed by stereo triangulation to determine the 3D occluding boundary, and finally, the projection of this boundary into the intermediate frames to predict the occluded areas [29].

Camera motion is computed by minimizing a function that accounts for 3D knowledge of the scene and 2D/2D correspondences over time. Given the viewpoint $[R_k, t_k]$ in frame k, the viewpoint in the next frame $k+1$ is determined using both 3D model points and interest points. The quality of the viewpoint is assessed by minimizing the distance between interest points and their corresponding epipolar lines, defined by the function:

$$\Phi(p) = \frac{1}{n}\sum_{i=1}^{n} dist^2(m_i, proj(M_i)) + \frac{\lambda}{2m}\sum_{i=1}^{m}\left[dist^2\left(q_{i,k+1}, ep_{k+1}(q_{i,k})\right) + dist^2\left(q_{i,k+1}, ep_{k+1}(q_{i,k})\right)\right]$$

where m_i are the interest points, M_i are the 3D model points, and *ep* represents the epipolar line.

To build the 3D occluding boundary, stereo triangulation is used. Given two key-views with the occluding object outlined as C_1 and C_2, a point m_1 on C_1 is matched with a point on C_2 by intersecting the epipolar line from m_1 with C_2. The correspondence problem is resolved using the order constraint, which ensures consistent matching of points.

Motion estimation errors significantly affect the accuracy of the 3D reconstruction. To address this, the concept of the ∈ indifference region is used, which defines a region in parameter space where the error function value is within an acceptable range. This region is described by:

$$\epsilon_r = \{p \text{ such that } |\Phi(p) - \Phi(p^*)| \leq \epsilon\}$$

where $\Phi(p^*)$ is the minimum value of the error function. In a small neighborhood around the minimum, Φ can be approximated using its Taylor expansion:

$$\Phi(p) \approx \Phi(p^*) + \nabla\Phi(p^*)^T \Delta p + \frac{1}{2}\Delta p^T H(p^*) \Delta p$$

where $H(p^*)$ is the Hessian matrix of Φ at $p = p^*$. Given that $\nabla\Phi(p^*) = 0$ at the minimum, the equation simplifies to:

$$\Phi(p) \approx \Phi(p^*) + \frac{1}{2}\Delta p^T H(p^*)\Delta p$$

The ∈ indifference region is then an ellipsoid defined by:

$$|\Delta p^T H(p^*)\Delta p| \leq 2\epsilon$$

This formulation allows the calculation of the reconstruction error on the occluding boundary by considering the extremal viewpoints within the indifference region.

The 3D occluding boundary is projected into intermediate frames to estimate the 2D occluding boundary. The uncertainty in the 2D boundary is calculated by projecting the extremal 3D reconstructions using extremal viewpoints, forming a convex hull that represents the spatial uncertainty.

Refinement of the occluding boundary involves region-based tracking and active contour models. Region-based tracking adjusts the predicted boundary by fitting a deformable model to the actual occluding object, minimizing a correlation measure that considers both the boundary shape and texture.

The affine transformation that best matches the occluding template to the current image is found by minimizing:

$$E(a) = \sum_i \square \sum_{dx,dy=-W}^{W} (I_{template}(m_i + d) - I(transf_a(m_i + d)))^2$$

where$transf_a(m)$ is the affine transformation, and W is the window size for local image transformation.

To address this challenge, [30] introduced a method to densify sparse 3D points obtained from SLAM systems, enabling depth information for every pixel. Unlike traditional multi-view stereo methods, this approach prioritizes sharp depth edges aligned with image features, smooth transitions across textures, temporal consistency to prevent flickering, completeness in depth assignment, and real-time processing speed.

By computing sparse reconstructions using a SLAM system like DSO-SLAM, providing camera parameters and intermittent 3D points, a novel approach was employed to identify soft depth edges, ensuring continuity and reliability across frames. By leveraging optical flow fields from adjacent frames and merging forward and backward results, the accuracy of depth edge detection was enhanced while mitigating occlusion-related issues.

To maintain temporal coherence, nearby frames were selected with significant translational motion, ensuring robust optical flow computation. The fusion of forward and backward flow information is guided by a reliability measure, prioritizing regions with diverging flow projections indicative of dis-occlusions. This nuanced approach improves the definition of occlusion boundaries, crucial for realistic AR experiences.

5.2 3D Data Collection

Following the successful completion of the Structure-from-Motion reconstruction process, the three-dimensional model generated served as a pivotal dataset for our subsequent endeavors within the Blender project. This model was seamlessly integrated into the Blender environment, where further enhancing and refining was managed. Within this iterative process, attention was dedicated to rectifying any imperfections or discrepancies present within the initial SFM-generated model.

Utilizing the powerful tools and functionalities offered by Blender, we scrutinized the 3D model, identifying and addressing any inaccuracies or errors that may have arisen during the reconstruction phase. Through a combination of manual adjustments, sculpting techniques, and texture mapping, the model underwent a comprehensive refinement process aimed at enhancing its visual fidelity and structural integrity.

Moreover, the Blender project facilitated the implementation of additional enhancements and features, further augmenting the overall realism and immersive quality of the model. This included the incorporation of intricate details, such as surface textures, lighting effects, and environmental elements, all crafted to evoke an authentic representation of the Durrës Tower and its surroundings.

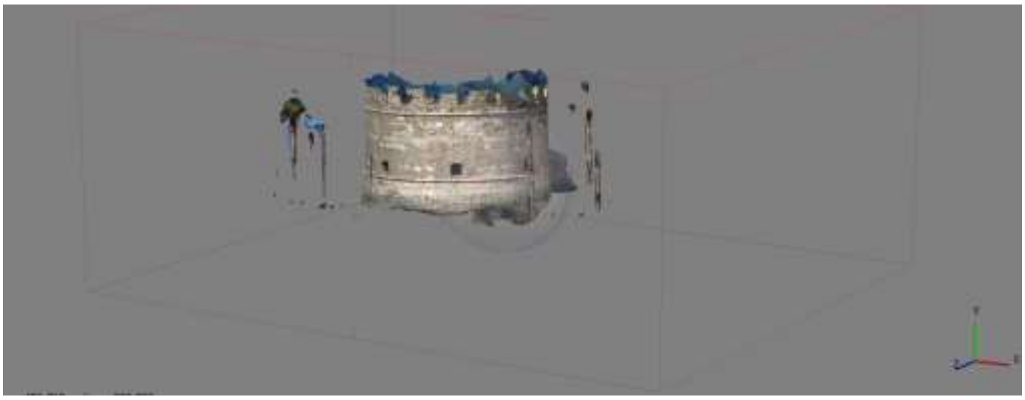


Figure 16 Image showing the finished 3D model from the point cloud

5.2.1. Object Alignment and Placement

After refining the 3D model in Blender, it became a crucial asset in our Unity project for implementing augmented reality features. Leveraging Unity's robust development environment, we seamlessly integrated the polished 3D model into our AR project, laying the groundwork for immersive and interactive experiences.

Within Unity, we transformed the 3D model into a reusable prefab, optimizing it for real-time rendering and interaction. This involved fine-tuning the model's materials, textures, and mesh properties to ensure optimal performance and visual quality within the AR application.

Moreover, Unity provided a comprehensive toolkit for AR development, enabling us to implement features like object tracking, user interaction, and spatial mapping. By integrating ARCore or ARKit frameworks, depending on the platform, we facilitated the seamless overlay of virtual content onto the real-world environment, enriching users' interaction with the Durrës Tower and its surroundings.

Additionally, Unity's scripting capabilities allowed us to add custom logic and behaviors, enabling dynamic interactions and immersive storytelling within the AR experience. Whether triggering animations, simulating environmental effects, or providing contextual information, Unity empowered us to create engaging and memorable AR experiences centered around the historical significance of the Durrës Tower.

5.3 Tools and Instruments

5.3.1. Blender Modelling

Blender, an open-source 3D modeling software, served as a tool in the creation and refinement of the digital assets crucial for the AR experience. Leveraging Blender's versatile suite of modeling, sculpting, and texturing tools, intricate details of the Durrës Tower and its surroundings were crafted. From architectural elements to surface textures, each component underwent scrutiny and refinement to ensure accuracy and visual fidelity. Additionally, Blender's animation capabilities facilitated the creation of dynamic elements, further enriching the immersive experience for users exploring the historical site within the AR environment.

5.3.2. Unity Environment Setup

Unity, a powerful cross-platform game engine, provided the framework for integrating the 3D assets created in Blender into a cohesive AR experience. The setup process involved configuring Unity's environment to accommodate the specific requirements of the project, including asset importation, scene organization, and interaction scripting. Through Unity's intuitive interface and extensive documentation, the integration of digital assets and implementation of AR functionalities were streamlined, allowing for efficient development iterations and rapid prototyping.

CHAPTER 6

EXPERIMENTAL SETUP

6.1 Introduction

In this section, we delve into the practical experiments conducted as an extension of the comprehensive literature review, with a specific emphasis on two primary objectives: generating sparse point clouds from sequences of images and the innovative creation of initial AR objects seamlessly integrated into real-world environments. These experiments are pivotal in bridging the theoretical underpinnings of our research with tangible, applied outcomes.

The first set of experiments focuses on the generation of sparse point clouds, a crucial step in understanding and reconstructing physical spaces in digital form. This process involves the intricate conversion of image sequences into a 3D model, using image data collected from the historical Durrës fortress as a case study. The choice of the Durrës fortress as a subject provides a rich tapestry of architectural complexity and historical significance, making it an ideal candidate for this high-precision digital reconstruction endeavor.

Parallel to the point cloud generation, our research extends into the realm of AR, where we venture into the creation and placement of AR objects within real environments. This segment of the experiments is facilitated by the advanced capabilities of the Unity Engine, a leading platform for AR development. By leveraging Unity, we are able to not only create highly detailed and interactive AR objects but also strategically place them within real-world settings, thus achieving a harmonious blend of virtual and physical realities.\par

These experiments, encompassing both point cloud generation and AR object creation, are instrumental in advancing our understanding of spatial digitization and interactive virtual experiences. Through the meticulous application of cutting-edge technologies and innovative methodologies, we aim to push the boundaries of what is possible in the digital representation of physical spaces and the immersive interaction with augmented environments.

6.2 Model Generation

6.2.1. Incremental SfM

To begin generating the point cloud, we initially gathered data by capturing images of the Venetian Tower in Durrës's fortress from various angles. We utilized an iPhone Xs Max for this task, known for its high-quality camera capable of recording at 4K resolution with a frame rate of 60 frames per second. Once the data was collected, we followed a methodology outlined in the existing research [31], which discusses the challenges and advancements in creating 3D models from 2D images. It highlights the growing popularity of using standard cameras like the one in our iPhone due to their cost-effectiveness and improving technology. This transition enables easier creation of 3D models, which holds significance for applications such as autonomous driving and AR.

As an initial experiment, we began by extracting two frames from the video sequences for every 60 frames, equating to one frame per half-second interval. These selected frames were then saved into dedicated folders for further processing. These meticulously curated folders served as the primary data inputs for the Python program, as referenced [31], employed in our study.

The paper reviews the process of SFM, a technique for deriving 3D structures from multiple photographs by solving pixel-wise correspondence issues, thereby approximating 3D models from different viewpoints. It highlights the incremental approach to SfM for addressing concerns of resilience, precision, completeness, and scalability, contrasting it with the cost and complexity of LIDAR-based systems.

During the experimenting phase, shown in Figure 17, Figure 18 and Figure 19, it was observed that processing the images and generating the point cloud significantly consumed time, even with the parameter `bundle adjustment` set to `false. This observation underscores the computationally intensive nature of the 3D reconstruction process, especially in the context of SFM.

Bundle adjustment is a key optimization process in SFM and other 3D reconstruction techniques. It involves refining the 3D coordinates of scene points, as well as the camera positions and orientations, to minimize reprojection errors. Reprojection error is the difference between the observed image points and the projected points of the 3D model as seen from the camera's perspective. In simpler terms, bundle adjustment optimizes the 3D points and camera parameters so that the re-projected points align as closely as possible with the actual image points captured by the camera.

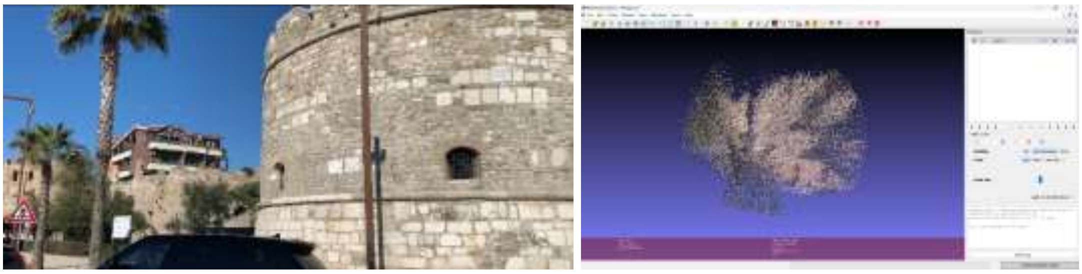


Figure 17 Figures showing the real construction of the main tower and the generated point cloud of it in MeshLab

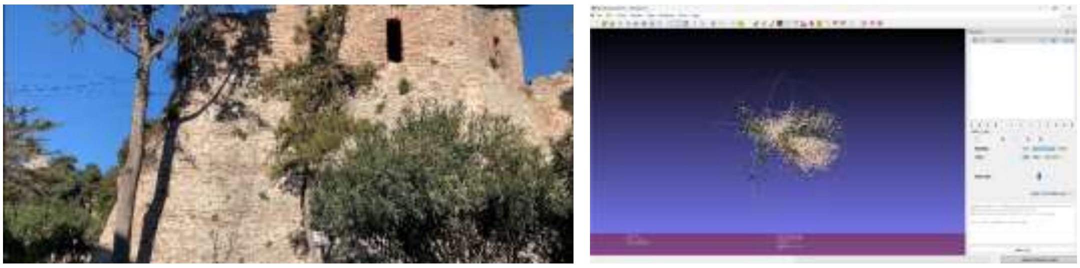


Figure 18 Figures showing the real construction of the surrounding walls and the generated point cloud of it in MeshLab

Figure 19 Figures showing the real construction of an arched wall and the generated point cloud of it in MeshLab

The sparse nature of the generated point cloud, evident after rendering it in MeshLab, indicates the need for a new dataset collection and refinement of camera calibration parameters for the SFM program.

The sparsity of the point cloud suggests that the current dataset may not adequately cover the scene from various angles. Certain areas or perspectives might be underrepresented or missed entirely in the initial dataset collection process.

The quality of camera calibration directly impacts the accuracy of the reconstructed 3D scene. If the camera parameters such as focal length, distortion coefficients, and principal point are not accurately calibrated, it can lead to inaccuracies in the reconstruction process, resulting in a sparse point cloud.

By gathering a new dataset with improved coverage from a wider range of angles and refining the camera calibration parameters to ensure greater accuracy, we aim to address the shortcomings observed in the initial reconstruction attempt.

6.2.2. SfM Integration With MVS

During our project we also used COLMAP, a project, on GitHub to create a 3D point cloud from a set of photos. We studied the COLMAP [22] in depth to understand its methods better. This knowledge helped us effectively use the COLMAP software and assess the quality of the resulting point cloud.

COLMAP relies on two techniques: Structure, from Motion and Multi View Stereo. SfM determines camera positions and 3D structures from overlapping images by identifying points and matching them across images. It creates a sparse point cloud that outlines the scenes geometry like putting together puzzle pieces from the same photos, we used in section 6.2.1.

After completing the SfM phase we moved on to Multi View Stereo. This method improves the sparse point cloud by adding 3D points enhancing the reconstruction into a detailed 3D model. MVS uses viewpoints to estimate depth information and enriches the model with intricate details that were previously missing.

After going through all the steps to process our collection of photos we ended up with a detailed point cloud, shown in Figure 20 and Figure 21, that we carefully checked for accuracy and level of detail. Comparing this point cloud, to the scene showed a level of accuracy accurately capturing all the intricate features and structures in the environment. This validation confirmed that COLMAPs methods are reliable in creating top notch reconstructions.

Figure 20 Part of the tower being reconstructed. In red are the camera positions, in relation to the extracted features

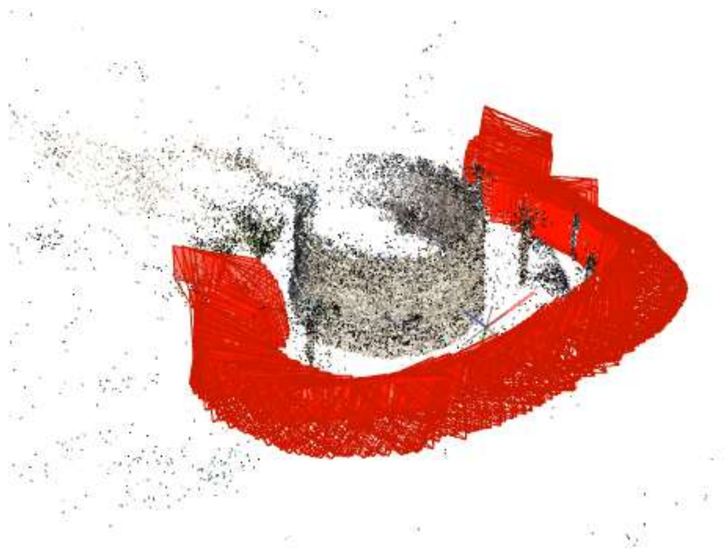


Figure 21 Sparse point cloud of the tower

Our hands on experience with COLMAP really showcases how effective and precise it is for 3D modeling tasks. The softwares combination of SfM and MVS tools creates a toolkit for generating 3D models from regular photos. This capability is especially valuable for fields like documenting heritage, architecture and autonomous navigation.

6.3 AR Setup

For our AR experiments we started by studying the documentation provided by Google. This guide is essential, for understanding the complexities of AR development. Offers insights into core concepts, key functionalities and best practices for using Googles AR technologies to create immersive experiences. The documentation covers a range of topics.

It also covers concepts to AR, such as motion tracking, understanding environments estimating lighting conditions and user interactions. Mastering these core principles is crucial for unlocking the potential of AR technology.

Furthermore, it offers explanations about the range of features and APIs provided by ARCore. These include functionalities, like plane detection rendering point clouds conducting hit tests and enabling occlusion capabilities. These features empower developers to create engaging AR experiences that seamlessly blend elements with the world.

After thoroughly studying the documentation [32], we began setting up and arranging the necessary components to smoothly incorporate AR objects into real-life scenarios. For our initial trial, we decided to create a blue sphere as the AR object, equipped with dynamic scaling abilities based on its distance from the camera.

Testing our AR Unity program on a mobile device like the Samsung Galaxy S10 Plus was a crucial step in our development process. It provided us with important information about how well our augmented reality app performs in the real world and how users experience it. One thing we noticed right away was how well the AR content tracked and stayed stable. Thanks to the device's advanced motion tracking sensors and powerful processing capabilities, our AR objects stayed firmly anchored to the physical environment. They didn't shake or drift much as the device moved around. This reliable tracking made the AR experience feel more realistic, which helped users feel more immersed and engaged.\par

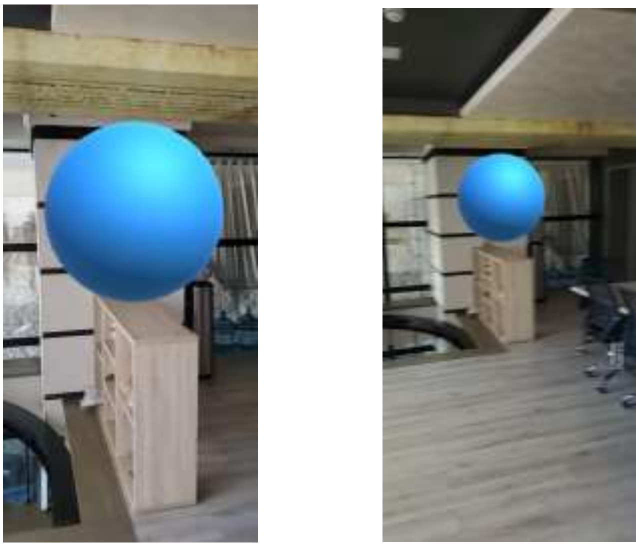


Figure 22 Figures showing how the sphere's size changes by the position of the camera

This clever setup gives the virtual sphere a sense of realism by mimicking perspective effects. In this scenario, objects appear larger when closer and smaller as the camera moves away, as illustrated in Figure 22.

After that, we started looking closely at real-world surfaces through practical investigations. Our main goal was to identify and interact with flat surfaces in the physical world. We mainly concentrated on detecting and extracting flat surfaces, using the features built into Unity's ARCore Kit.

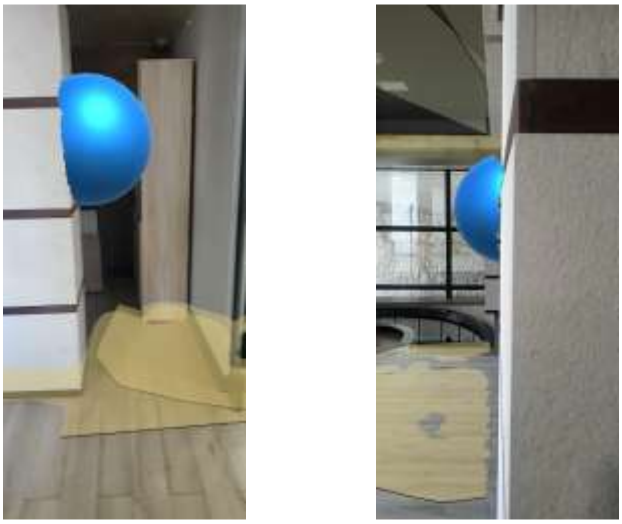


Figure 23 Figures showing how the sphere occludes behind other objects

This capability allowed the program to recognize and outline both horizontal and vertical flat surfaces in the surroundings, using the camera of the device. This skill lets AR objects blend not quite smoothly with real-world surfaces, creating a harmonious connection between the digital and physical worlds, as shown in Figure 23.

Figure 24 Plane extraction and recognition from the environment

During our experiments, we noticed something interesting when we placed the AR sphere behind a glass surface. It looked like the sphere was floating in water, like shown in Figure 25, with only parts of it visible through the glass. It created a cool effect where the virtual and real worlds mixed together. Even a small change in the camera angles it made the sphere's appearance change a lot.

Figure 26 Human occlusion with the AR object

Figure 25 Errored occlusion of AR object placed behind transparent obstacle

Finally, we explored occlusion experimentation to understand ARCore's ability to render virtual objects hidden behind real-world entities. While object occlusion showed promising effectiveness, with the blue sphere occasionally positioned behind tangible objects, human occlusion revealed certain limitations, as show in Figure 26. Specifically, attempts to hide the sphere by introducing human subjects or their body parts, like hands, failed to consistently hide its visibility. This practical investigation was conducted with reference to [16] which scrutinizes the complexities of occlusion in AR environments.

CHAPTER 7

RESULTS

7.1 Point Cloud Generation

The research utilized two methods for reconstruction; Structure from Motion with Multi View Stereo (SfM + MVS) and Incremental Structure from Motion. These techniques led to the creation of two point cloud models of the tower. Following the generation of these point clouds, dense clouds and meshes were developed. This process involved bundle adjustment to enhance the accuracy of the reconstructions. The final 3D models were then prepared for export as objects that can be further modified in Blender.

The statistical measurements gathered from both reconstruction techniques include the quantity of image inputs the tally of registered images the overall number of points the count of observations the average track length, the average number of observations per image and the average reprojection error. These metrics offer an evaluation of how well each reconstruction method performs and its level of accuracy.

Quantity of image inputs denotes the number of images utilized as input for creating a 3D reconstruction. It plays a role as the quality and completeness of the reconstruction heavily rely on both the quantity and coverage provided by these input images.

Both methods, for creating 3D models used a set of 770 images that show the structure of the Venetian Tower. Each picture in this collection has a resolution of 3840 x 2160 pixels, uses 24 bits for colors ensuring color representation and intricate details. Moreover, the images have horizontal resolutions of 96 dpi (dots per inch) providing the level of detail for an accurate reconstruction. The quantity and quality of these images play a role in how the reconstruction process works. Having images allows for gathering data points, which increases the density and precision of the point clouds.

Number of registered images indicates how many input images were successfully aligned or registered during the reconstruction process. A higher count signifies algorithm performance since more images are accurately incorporated in constructing the model.

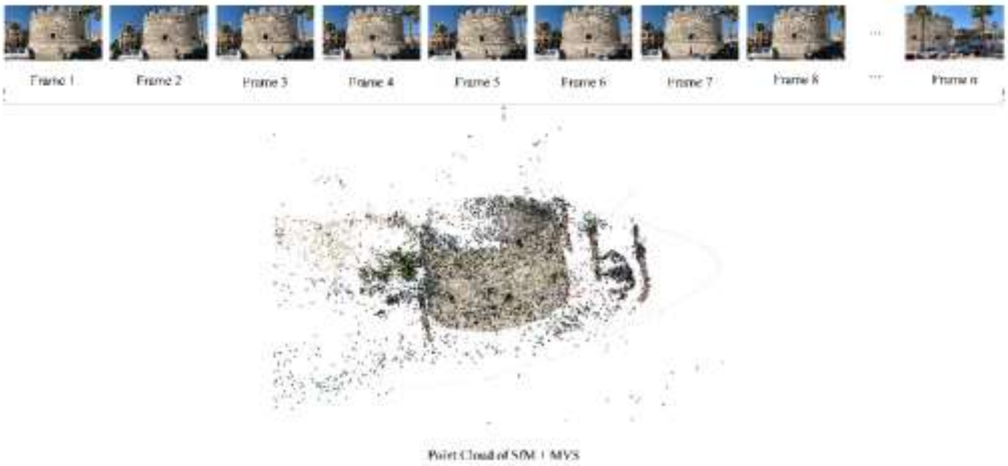


Figure 27 Image showing the input dataset and the result of the SfM + MVS

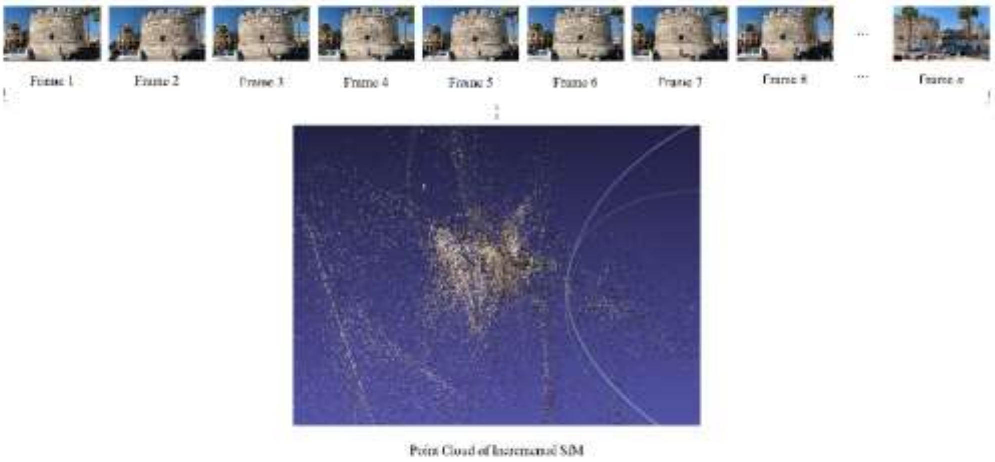
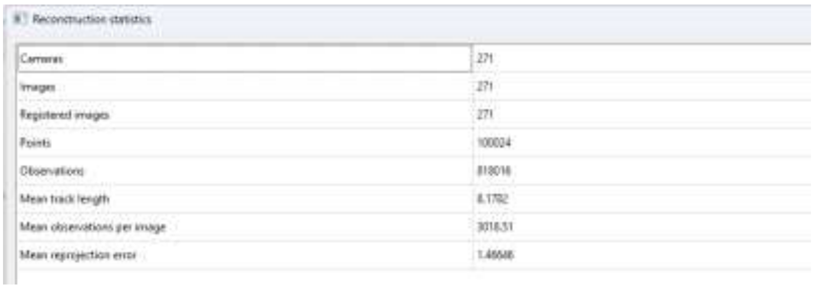


Figure 28 Image showing the input dataset and the result of the Incremental SfM

Another metric considered during the experimental results of the two methods is the overall number of points, which refers to the count of points within the point cloud produced through reconstruction. A greater number typically suggests a higher resolution model that captures finer details of the structure. For test purposes only, the point cloud of the tower without its surrounding walls was reconstructed. For the SfM+MVS model 100,024 points were generated, while for the incremental SfM 89,560 points. This suggests that the first model reconstructed offers a higher resolution when its mesh is later produced.


Reconstruction statistics

Cameras	271
Images	271
Registered images	271
Points	100024
Observations	818016
Mean track length	8.1782
Mean observations per image	3018.51
Mean reprojection error	1.48646

Figure 29 Metrics calculated from COLMAP

Apart from points count, we considered the count of observations, which represents how many times points in the point cloud are observed across all registered images. By this we can undoubtedly say that the SfM+MVS model offers more observations, thus offering more accuracy.

A critical metric we calculated, in order to choose the better reconstructed model, is average reprojection error. This measure reflects the disparity between where points are observed in images and where they're projected from the model. A lower average reprojection error signifies accuracy in the reconstruction indicating that points in the model closely match their positions in 2D images.

Comparing the performance of reconstruction methods (SfM + MVS and Incremental SfM) allows us to assess how well the input images work, the durability of the point cloud and the overall quality of the model. This analysis helps us identify which approach provides a dependable depiction of the Venetian Tower, in Durres.

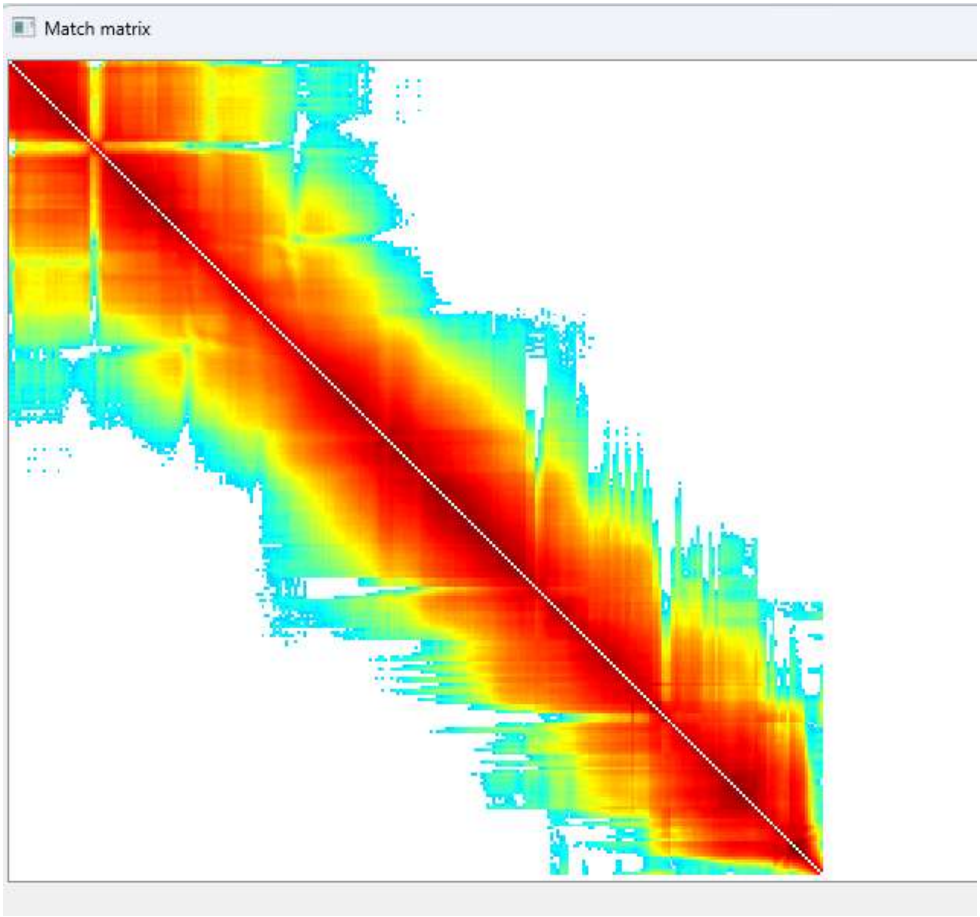


Figure 30 Match matrix of the SfM + MVS reconstructed model

The comparison chart shown in Figure 30 visually represents the degree of connection, between the pairs of images during the reconstruction phase. Each box in the chart indicates how similar two images are. The noticeable red line running diagonally suggests a correlation indicating that the images match and align with each other. The recurring yellow sections, along this line indicate a similarity suggesting that shared features have been successfully employed in the reconstruction process.

7.2 Model Modification

To accurately reconstruct the missing parts of the Durrës fortress, we took into consideration both architectural maps and historical findings that describe the buildings and the surrounding walls. This multi-faceted approach allowed us to create a more comprehensive and accurate 3D model of the fortress.

- Architectural Maps: We utilized detailed architectural maps of the Durrës fortress to identify the precise locations and dimensions of the missing walls. These maps provided a foundational layout that guided the reconstruction process. By overlaying these maps onto current satellite images, we could ensure that the reconstructed walls aligned accurately with existing structures.
- Historical Findings: Historical documents and records were consulted to gather additional information about the design and construction of the fortress walls. These records included descriptions of the materials used, the construction techniques employed, and the historical modifications made to the walls over time. This historical context was crucial in ensuring that the reconstructed walls were not only accurate in terms of their placement but also in their architectural style and materials.

7.2.1. Reconstruction Process

- Digital Modeling: Using the data collected, we employed advanced 3D modeling software to digitally reconstruct the missing walls. This involved creating a detailed 3D model that incorporated the precise dimensions and architectural features identified in the maps and historical documents.

- Integration with Existing Models: The newly reconstructed walls were then integrated with the existing 3D model of the Durrës fortress. This integration was carefully managed to ensure that the new additions seamlessly matched the current structure in terms of scale, texture, and architectural style.
- Texture Mapping: Texture maps were generated from the scans to accurately reproduce the surface details of the walls. These maps were applied to the digital assets, ensuring that the color, texture, and weathering of the new sections matched the original walls as closely as possible, as shown in Figure 31.
- Validation and Adjustment: Finally, the reconstructed model was validated against the historical and architectural data to ensure its accuracy. Any discrepancies identified during this validation process were addressed, and the model was adjusted accordingly to achieve the highest level of precision.

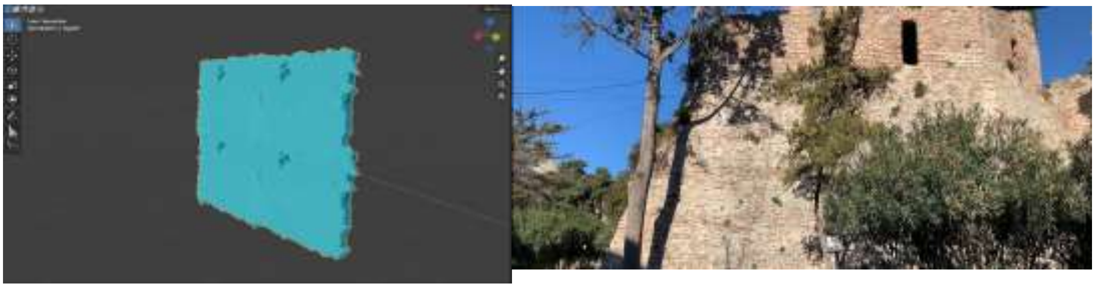


Figure 31 *Image showing the texture of the generated wall and the existing reference*

7.3 Model Presentation

Historical records and analysis of the existing walls were used to match the materials used in the new wall parts. This included sourcing similar types of stone and mortar to maintain consistency in the construction materials.

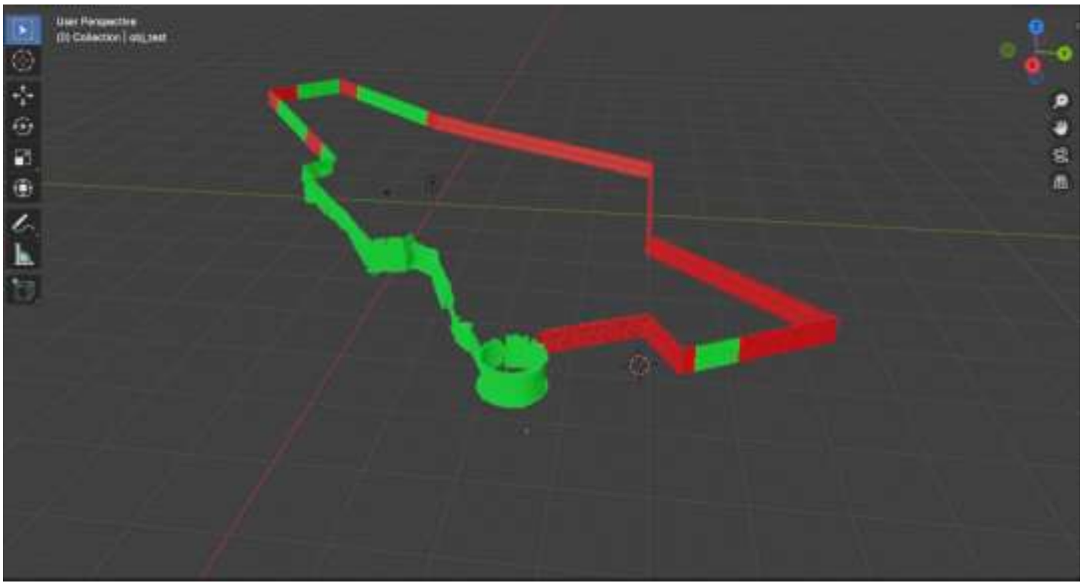


Figure 32 *Image showing the existing (green) vs missing parts (red)*

Figure 33 *Image showing how the whole fortress would be placed*

Figure 34 *Image showing how the fortress will look in the AR experience*

CHAPTER 8

DISCUSSION

This section explores the aspects and consequences of utilizing digital technologies to conserve historical sites with a specific focus, on the Venetian Tower in Durrës. By reconstructing the tower using techniques like Structure from Motion and Multi-View Stereo to generate a model this project incorporates point cloud data into an Augmented Reality application project. The discussions here will touch upon understanding the findings, the pros and cons of AR, the importance of point cloud data, user experience and design interaction as the wider implications.

8.1 Interpretation of Findings

The utilization of Structure, from Motion and Multi-View Stereo methods has allowed for the development of an 3D model depicting the Venetian Tower. The capacity to complete sections of the edifice has resulted in a digital portrayal, essential for in depth analysis and upcoming restoration endeavors.

The Augmented Reality application enriches the visualization of the model enabling users to interact with the depiction of the tower. This interactive display not assists in grasping the complexities but also captivates a wider audience.

The intricate 3D model acts as an asset providing also a guide for restoration in the future, guaranteeing that any modifications remain true to the design. This is especially critical for upholding the authenticity of the tower.

The AR application holds promise for community engagement initiatives. By delivering an experience it can raise consciousness regarding heritage preservation and involve younger individuals in cultural heritage appreciation.

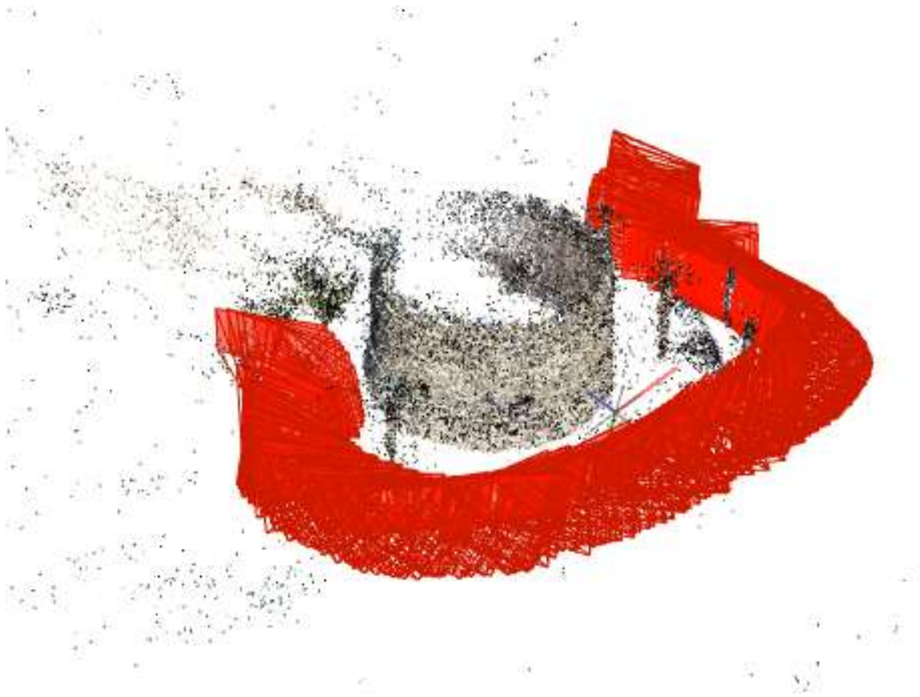


Figure 35 Sample of the point cloud structure of the Venetian tower

8.2 Advantages and Limitations of AR

Augmented reality provides an opportunity to explore sites in an interactive way. By immersing users in the environment of the Venetian Tower it enriches their understanding and appreciation of these locations. AR offers a means for individuals who're unable to visit heritage sites to still engage with them proving particularly valuable, for educational purposes and reaching a global audience. Through AR users can access information like context, restoration processes and structural details that enhance their overall experience.

However, creating and utilizing AR applications necessitates expertise and resources. The quality of AR experiences hinges on hardware and software which can pose challenges, for some heritage organizations.

Dealing with extensive point cloud data sets presents its set of difficulties. Ensuring the accuracy and integrity of data when transitioning from 3D models to AR platforms demands attention to detail. While AR enhances accessibility it also requires users to possess devices and a certain level of proficiency potentially limiting its widespread adoption.

Figure 36 Sample of AR experience failing on detecting coordinates from user's device

8.3 User Experience and Interaction Design

The augmented reality experience includes navigation features to help users easily move around the model of the Venetian Tower, which is pre-processed in Blender to also show the missing parts of the tower. It's essential to offer instructions and user-friendly interfaces comes from incorporating functionalities such as rotating and moving around the AR structure.

Users should have the ability to explore parts of the tower and access information at their convenience, learning about the significance of the site and better understand conservation efforts, which can deepen their connection, with it. Collecting feedback from users through ratings and comments also aids in enhancing the project base

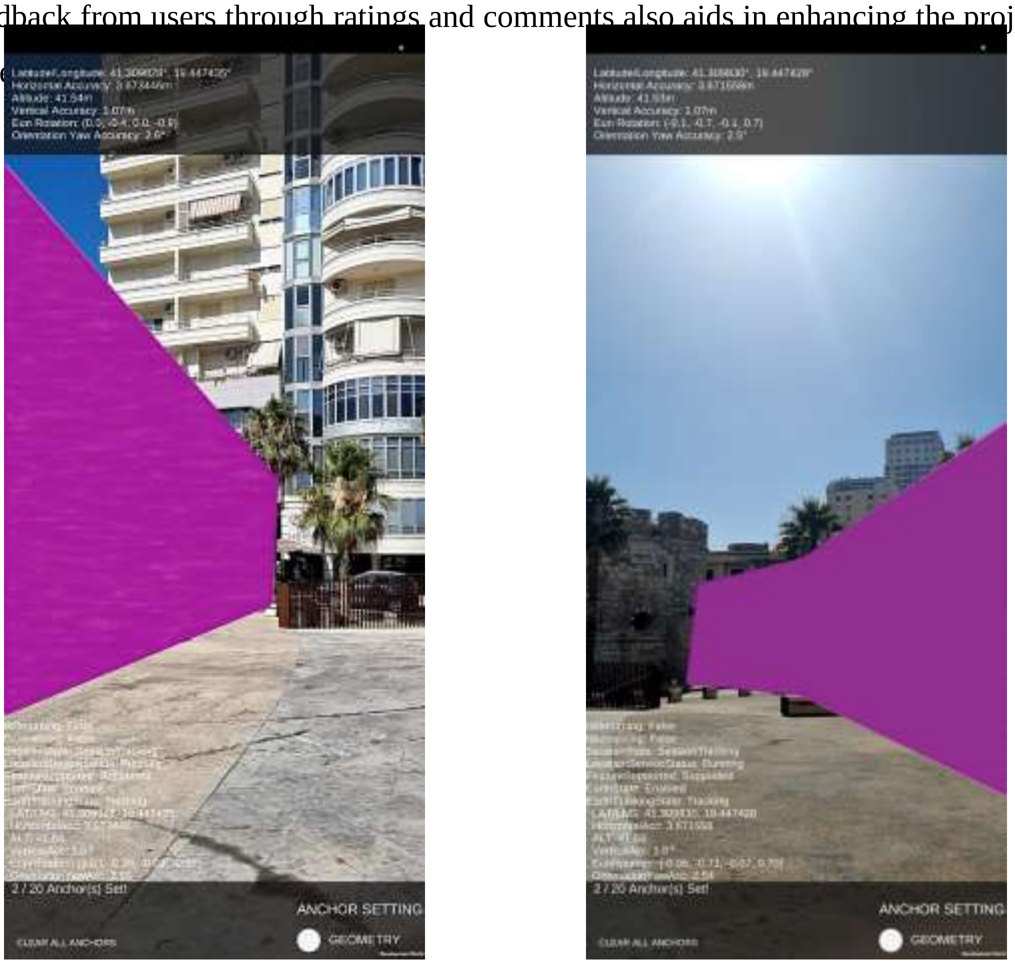


Figure 37 *Images showing the blueprint-wall of the missing part of the fortress in the AR experience*

CHAPTER 9

CONCLUSIONS

9.1 Conclusions

The growing importance of safeguarding and conserving cultural heritage sites and artifacts is more crucial now, than before. This sense of urgency arises from the impact of both activities and natural elements which pose serious threats to these priceless remnants of our history. As a result, experts and conservationists are reassessing their approaches realizing the need to develop techniques to tackle the challenges faced by cultural heritage sites.

Among the advancements tools like 3D point cloud data, laser scanning and augmented reality have emerged as key assets in preservation endeavors. These technologies go beyond reacting to issues; they embody proactive measures that address the obstacles presented by an ever-evolving environment. By utilizing 3D point cloud data conservationists can craft replicas of archaeological sites and structures. This facilitates the documentation and conservation of details that may otherwise be lost due to deterioration or human interventions.

This study has delved into the history of Durrës with a focus on the importance of the Venetian Tower and initiatives aimed at conserving the citys historic landmarks. Through this exploration we have gained an insight into how modern technology bridges gaps, between the past and present ensuring that Durrës rich history and that of Albanian cities are respected and protected for future generations.

The situation, in Durrës showcases how changes in the environment and human actions impact development significantly. The citys past, influenced by periods of occupation and natural changes emphasizes the importance of preservation plans. The Venetian Tower stands as a symbol of the citys journey facing preservation obstacles due to unregulated construction activities and events like the 2019 earthquake. These challenges not jeopardize the tower. Also put at risk the iconic Amphitheater underlining the pressing need to protect and document these historic sites.

The incorporation of computer vision methods, Structure from Motion into preservation strategies has been incredibly beneficial. SfMs capability to create three environments from two images or video frames enhances computational systems understanding of spatial complexities proving to be an essential tool for heritage conservation efforts. This approach not advances augmented reality applications. Also extends its advantages to various fields related to heritage protection.

In reality applications precise object alignment is crucial. This process involves placing objects within real world settings in a way that seamlessly blends with their surroundings. Despite challenges like obstruction, poor lighting conditions and limited visual cues computer vision techniques ensure tracking and anchoring of objects, onto physical surfaces.

In situations where automated systems fall short human intervention is crucial to ensure the accuracy and integrity of the representation.

9.2 Recommendations for future research

One promising area for exploration and innovation involves developing web-based platforms dedicated to cultural heritage preservation. These platforms would enable access and management of 3D models fostering collaboration among experts, conservationists and the public on a scale. By harnessing computing capabilities such platforms can provide storage solutions and scalable processing power to facilitate the handling of extensive datasets related to detailed 3D point cloud models and high-resolution scans. Moreover, web-based platforms have the potential to democratize access to heritage sites by offering tours and educational experiences that can raise awareness and garner support for preservation initiatives.

When objects or parts are hidden from view it presents a challenge, in reconstruction and augmented reality tasks. Future work could center on improving algorithms to handle occlusions better. This might involve enhancing computer vision methods to infer the shape of obstructed elements using machine learning models trained on datasets.

REFERENCES

[1] J. Yu, L. Sun, J. Wang and Q. Zhu, "A Multi-layer Residual Architecture for Point Cloud Geometry Compression," in *2022 International Conference on Virtual Reality, Human-Computer Interaction and Artificial Intelligence (VRHCIAI)*, 2022.

[2] Google, "ARCore Documentation," 2022. [Online]. Available: https://developers.google.com/ar/develop.

[3] A. Aditya, T. V. Prashant, R. Vishal and K. S. Yashas, *Multiview 3D Reconstruction*, 2022.

[4] E. Alexiou, E. Upenik and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, 2017.

[5] J. E. Aparicio Carranza, S. R. Zaidi and H. Carranza, "Plane Detection Based Object Recognition for Augmented Reality".

[6] J. E. Aparicio Carranza, S. R. Zaidi and H. Carranza, "Plane Detection Based Object Recognition for Augmented Reality".

[7] A. U. Batmaz and W. Stuerzlinger, "When Anchoring Fails: Interactive Alignment of Large Virtual Objects in Occasionally Failing AR Systems," in *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 1*, 2022.

[8] M. Batur, O. Yilmaz and H. Ozener, "A case study of deformation measurements of Istanbul land walls via terrestrial laser scanning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, p. 6362–6371, 2020.

[9] W. J. Beksi and N. Papanikolopoulos, "A topology-based descriptor for 3D point cloud modeling: Theory and experiments," *Image and Vision Computing*, vol. 88, p. 84–95, 2019.

[10] R. Bergquist and N. Stenbeck, *Using augmented reality to measure vertical surfaces*, 2018.

[11] G. J. Brostow, J. Shotton, J. Fauqueur and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10,* 2008.

[12] M. Bui, L.-C. Chang, H. Liu, Q. Zhao and G. Chen, "Comparative study of 3d point cloud compression methods," in *2021 IEEE International Conference on Big Data (Big Data),* 2021.

[13] J. Carmigniani and B. Furht, "Augmented reality: an overview," *Handbook of augmented reality,* p. 3–46, 2011.

[14] G. Caroti, A. Piemonte and others, "An integrated survey for knowledge and preservation of a cultural heritage: The Albanian fortified citadel of Elbasan," *INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES,* vol. 37, p. 373–378, 2008.

[15] R. Comes, C. Neamţu, Z. Buna, I. Badiu and P. Pupeză, "METHODOLOGY TO CREATE 3D MODELS FOR AUGMENTED REALITY APPLICATIONS USING SCANNED POINT CLOUDS.," *Mediterranean Archaeology & Archaeometry,* vol. 14, 2014.

[16] M. De Fino, C. Ceppi and F. Fatiguso, "Virtual tours and informational models for improving territorial attractiveness and the smart management of architectural heritage: The 3d-imp-act project," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 44, p. 473–480, 2020.

[17] L. Dipasquale, M. Carta, S. Galassi and A. Merlo, "The Vernacular Heritage of Gjirokastra (albania): Analysis of Urban and Constructive Features, Threats and Conservation Strategies," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 44, p. 33–40, 2020.

[18] S. Gupta and B. Lohani, "Augmented reality system using lidar point cloud data for displaying dimensional information of objects on mobile phones," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 2, p. 153–159, 2014.

[19] M. Habbecke and L. Kobbelt, "A surface-growing approach to multi-view stereo reconstruction," in *2007 IEEE Conference on Computer Vision and Pattern Recognition,* 2007.

[20] A. Hajirasouli, S. Banihashemi, A. Kumarasuriyar, S. Talebi and A. Tabadkani, "Virtual reality-based digitisation for endangered heritage sites: Theoretical framework and application," *Journal of Cultural Heritage,* vol. 49, p. 140–151, 2021.

[21] Y. He, Y. H. Ma and X. R. Zhang, "“DIGITAL HERITAGE” THEORY AND INNOVATIVE PRACTICE," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, p. 335–342, 2017.

[22] A. Holynski and J. Kopf, "Fast depth densification for occlusion-aware augmented reality," *ACM Transactions on Graphics (ToG),* vol. 37, p. 1–11, 2018.

[23] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[24] V. Lepetit and M.-O. Berger, "A semi-automatic method for resolving occlusion in augmented reality," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, 2000.

[25] S. Lim, M. Shin and J. Paik, "Point cloud generation using deep local features for augmented and mixed reality contents," in *2020 IEEE International Conference on Consumer Electronics (ICCE),* 2020.

[26] C. Lv, W. Lin and B. Zhao, "Approximate intrinsic voxel structure for point cloud simplification," *IEEE Transactions on Image Processing,* vol. 30, p. 7241–7255, 2021.

[27] E. H. Makiyah and N. N. Khamiss, "Point Cloud Compression for Augmented Reality-Based Campus Navigation Application," in *2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS),* 2023.

[28] O. Martinez-Rubi, S. Verhoeven, M. Van Meersbergen, P. Van Oosterom, R. GonÁalves, T. Tijssen and others, "Taming the beast: Free and open-source massive point cloud web visualization," in *Capturing Reality Forum*, 2015.

[29] A. P. Placitelli and L. Gallo, "Low-cost augmented reality systems via 3D point cloud sensors," in *2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems*, 2011.

[30] S. Santoro, A. Monti and others, "The city of Durres and its history, between environmental evolution and cultural transformations," 2011.

[31] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06),* 2006.

[32] M. P. Semini, L. Çapeli and F. Nepravishta, "Overview of The Italian Architecture in Durrës From 1920 to 1944".

[33] Y. Tian, T. Guan and C. Wang, "Real-time occlusion handling in augmented reality based on an object tracking approach," *Sensors,* vol. 10, p. 2885–2900, 2010.

[34] H. Turku, "Cultural property, human rights, and sustainable development: the case of the ancient city of durrës," *Hastings Envt'l LJ,* vol. 25, p. 121, 2019.

[35] X. Wei, Y. Zhang, Z. Li, Y. Fu and X. Xue, "Deepsfm: Structure from motion via deep bundle adjustment," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16,* 2020.

[36] Y. YARDIM and E. MUSTAFARAJ, "Structural Assessment and Restoration of the Leaden Mosque in Berat, Albania," *Kahramanmaras Sutcu Imam University Journal of Engineering Sciences,* vol. 15, 2012.

[37] L. Zhang, P. van Oosterom and H. Liu, "Visualization of point cloud models in mobile augmented reality using continuous level of detail method," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 44, p. 167–170, 2020.

[38] Y. Zhang and H. Bao, "Point clouds-image fusion by convolutional method for data integrity using neural network," in *2019 15th International Conference on Computational Intelligence and Security (CIS)*, 2019.

[39] J. Zhou, N. Li, Y. Liu, S. Yao and S. Chen, "Exploring spherical autoencoder for spherical video content processing," in *MM'22: Proceedings of the 30th ACM International Conference on Multimedia*, 2022.