جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم ظهر المهراز فاس
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

Ordre number :

Center of Studies: ***Sciences and Techniques and Medical Sciences***

Doctoral training: ***STIC***

Discipline: ***Computer Science***
Specialisation: ***Computer Science***
Laboratory: ***Informatics, Signals, Automatics and Cognitivism (LISAC)***

DOCTORAL THESIS

Title of thesis

Virtual Reality, 3D Human Pose Estimation and Close-Range Photogrammetry for Accessible Heritage Conservation and Immersive Cultural Tourism

Presented by:

KHALLOUFI HAMZA

Defended on July 20, 2024 in front of the jury:

Full Name	Grade	Etablissement	Quality
Pr EL BEQQALI Omar	PES	Faculty of Sciences Dhar El Mahraz, Fez	Chair
Pr NFAOUI El Habib	PES	Faculty of Sciences Dhar El Mahraz, Fez	Reviewer & Examiner
Pr KERZAZI Noureddine	MCH	National Higher School For Computer Science and Systems Analysis, Rabat	Reviewer & Examiner
Pr BERRADA Ismail	MCH	Mohammed VI Polytechnic University, Ben Guérir	Reviewer & Examiner
Pr MAHRAZ Mohamed Adnane	MCH	Faculty of Sciences Dhar El Mahraz, Fez	Examiner
Pr RIFFI Jamal	MCH	Faculty of Sciences Dhar El Mahraz, Fez	Examiner
Pr AZOUGH Ahmed	MCH	Léonard de Vinci Univeersity Pole, Paris	Invited
Pr TAIRI Hamid	PES	Faculty of Sciences Dhar El Mahraz, Fez	Supervisor

Academic year : 2023-2024

Abstract

In the era of digital transformation, the potential to harness emerging technologies for cultural heritage conservation has become increasingly relevant, especially in regions where resource constraints pose significant challenges. This thesis explores such innovative technological solutions with a dual focus: promoting tourism and ensuring the conservation of cultural heritage through affordability and immersive experiences. By integrating 3D Human Pose Estimation and close-range photogrammetry within the realm of Virtual Reality (VR), we aim to redefine cultural heritage conservation and immersive tourism experiences.

Our initial investigation centers on addressing the financial barriers associated with 3D modeling of historic sites, which is crucial for cultural heritage conservation. smartphone-based photogrammetry not only presents a cost-effective alternative but also offers comparable or even superior accuracy. This democratization of technology serves as a pivotal element for developing countries in promoting accessible and affordable tourism while conserving their cultural heritage.

Subsequently, the research shifts its gaze to the dynamic field of 3D Human Pose Estimation (HPE). Within this context, the accuracy of emerging cloud-based tools, particularly MOTiON by RADiCAL, is rigorously evaluated. Traditional HPE methodologies often rely on expensive sensor-based motion capture systems, which are not feasible for widespread adoption. Our comparative analysis with the CMU dataset suggests that cloud-based solutions can match or even exceed the efficacy of traditional systems. This shift underscores the transformative potential of cloud computing in revolutionizing motion capture techniques, which can significantly enhance virtual tourism experiences by providing realistic and interactive human figures.

The final facet of this multifaceted research delves into 'El-FenaVR,' an immersive VR exploration of Marrakech's iconic Jamaa El-fena Square, with a particular focus on intangible cultural heritage. Beyond mere representation, the study examines user perceptions and engagement with the vibrant, intangible cultural elements such as storytelling, music, and local traditions encapsulated within the VR experience. By utilizing the Unified Theory of Acceptance and Use of Technology as a guiding framework, our findings highlight VR's unparalleled potential in providing immersive and educational cultural experiences. These experiences not only foster a deeper appreciation and understanding of intangible cultural heritage among global audiences but also spotlight the challenges related to VR's accessibility and familiarity. Our research paves the way for future technological and educational initiatives aimed at preserving and promoting intangible cultural heritage through immersive tourism experiences.

Keywords: Virtual Reality, Photogrammetry, Human Pose Estimation, Cultural Heritage Conservation, Tourism Promotion, HCI, Usability, User Experience.

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 32,000 words including appendices, bibliography, footnotes, tables, and equations, and has fewer than 91 figures.

Hamza Khalloufi
July 2024

Acknowledgements

First of all, I would like to praise and thank Allah, the Almighty, for answering my prayers and granting me countless blessings, wisdom, and opportunities that led me to the completion of this thesis.

When I started work on this project, I had no idea of the scale of the effort involved. The project would have died many times over without the spirit of the many people involved.

I would like to express my deep gratitude to my advisors, Professor Hamid Tairi and Professor Ahmed Azough. Their enthusiastic commitment, unwavering support, and invaluable advice have played a decisive role in my journey in the field of scientific research. Professor Azough, in particular, was always open to new ideas and gave me the necessary support to achieve my research goals. Having him as my supervisor was a great honor

My sincere thanks go to my circle of family and friends, whose unconditional love and moral support got me through difficult times. A special mention goes to my wife Manar Hamdane; whose sacrifices and total understanding have been the basis of my support. I am eternally grateful to my parents Rkia and Ssi Abdellah for all they have done for me, and this thesis is a tribute to them. I would also like to extend my heartfelt gratitude to my brother, Omar Khalloufi, for his constant encouragement and invaluable advice, which have been instrumental in my journey. A special nod goes to Mohamed Zaifri, whose help with my experiments and thorough proofreading of this thesis were invaluable. He also gave me the daily inspiration I needed to pursue my dreams.

I am immensely grateful to the entire teaching staff at the Dhar El Mahraz Faculty of Science in Fez. The years I have spent learning from them have been priceless, and their influence has been positive, both academically and personally. Lastly, I would also like to thank all my jury members, Professor El Habib Nfaoui, Professor Noureddine Kerzazi, and Professor Ismail Berrada for agreeing to review my thesis and for their interest in my work. Professor Mohamed Adnane Mahraz and Professor Jamal Riffi for having honored my defense by participating in my jury as the examiners.

Abstract

In the era of digital transformation, the potential to harness emerging technologies for cultural heritage conservation has become increasingly relevant, especially in regions where resource constraints pose significant challenges. This thesis explores such innovative technological solutions with a dual focus: promoting tourism and ensuring the conservation of cultural heritage through affordability and immersive experiences. By integrating 3D Human Pose Estimation and close-range photogrammetry within the realm of Virtual Reality (VR), we aim to redefine cultural heritage conservation and immersive tourism experiences.

Our initial investigation centers on addressing the financial barriers associated with 3D modeling of historic sites, which is crucial for cultural heritage conservation. smartphone-based photogrammetry not only presents a cost-effective alternative but also offers comparable or even superior accuracy. This democratization of technology serves as a pivotal element for developing countries in promoting accessible and affordable tourism while conserving their cultural heritage.

Subsequently, the research shifts its gaze to the dynamic field of 3D Human Pose Estimation (HPE). Within this context, the accuracy of emerging cloud-based tools, particularly MOTiON by RADiCAL, is rigorously evaluated. Traditional HPE methodologies often rely on expensive sensor-based motion capture systems, which are not feasible for widespread adoption. Our comparative analysis with the CMU dataset suggests that cloud-based solutions can match or even exceed the efficacy of traditional systems. This shift underscores the transformative potential of cloud computing in revolutionizing motion capture techniques, which can significantly enhance virtual tourism experiences by providing realistic and interactive human figures.

The final facet of this multifaceted research delves into 'El-FenaVR,' an immersive VR exploration of Marrakech's iconic Jamaa El-fena Square, with a particular focus on intangible cultural heritage. Beyond mere representation, the study examines user perceptions and engagement with the vibrant, intangible cultural elements such as storytelling, music, and local traditions encapsulated within the VR experience. By utilizing the Unified Theory of Acceptance and Use of Technology as a guiding framework, our findings highlight VR's unparalleled potential in providing immersive and educational cultural experiences. These

experiences not only foster a deeper appreciation and understanding of intangible cultural heritage among global audiences but also spotlight the challenges related to VR's accessibility and familiarity. Our research paves the way for future technological and educational initiatives aimed at preserving and promoting intangible cultural heritage through immersive tourism experiences.

Keywords: Virtual Reality, Photogrammetry, Human Pose Estimation, Cultural Heritage Conservation, Tourism Promotion, HCI, Usability, User Experience.

Résumé

À l'ère de la transformation numérique, le potentiel d'exploitation des technologies émergentes pour la conservation du patrimoine culturel est devenu de plus en plus pertinent, en particulier dans les régions où les contraintes de ressources posent des défis importants. Cette thèse explore de telles solutions technologiques innovantes avec un double objectif : promouvoir le tourisme et assurer la conservation du patrimoine culturel grâce à des expériences abordables et immersives. En intégrant l'estimation de la pose humaine en 3D et la photogrammétrie à courte distance dans le domaine de la réalité virtuelle (RV), nous visons à redéfinir la conservation du patrimoine culturel et les expériences touristiques immersives.

Notre recherche initiale est centrée sur la résolution des obstacles financiers associés à la modélisation 3D des sites historiques, qui est cruciale pour la conservation du patrimoine culturel. La photogrammétrie basée sur les smartphones présente non seulement une alternative rentable, mais offre également une précision comparable, voire supérieure. Cette démocratisation de la technologie est un élément essentiel pour les pays en développement, qui peuvent ainsi promouvoir un tourisme accessible et abordable tout en conservant leur patrimoine culturel.

La recherche se tourne ensuite vers le domaine dynamique de l'estimation de la pose humaine en 3D (HPE). Dans ce contexte, la précision des outils émergents basés sur le cloud, en particulier MOTiON de RADiCAL, est rigoureusement évaluée. Les méthodologies HPE traditionnelles s'appuient souvent sur des systèmes de capture de mouvement coûteux basés sur des capteurs, qui ne peuvent pas être adoptés à grande échelle. Notre analyse comparative avec l'ensemble des données de la CMU suggère que les solutions basées sur l'informatique dématérialisée peuvent égaler, voire dépasser, l'efficacité des systèmes traditionnels. Ce changement souligne le potentiel de transformation de l'informatique en nuage pour révolutionner les techniques de capture de mouvements, qui peuvent améliorer de manière significative les expériences de tourisme virtuel en fournissant des figures humaines réalistes et interactives.

La dernière facette de cette recherche à multiples facettes se penche sur "El-FenaVR", une exploration RV immersive de l'emblématique place Jamaa El-fena de Marrakech, avec un accent particulier sur le patrimoine culturel immatériel. Au-delà de la simple représentation, l'étude examine les perceptions et l'engagement des utilisateurs à l'égard des éléments culturels vibrants et immatériels tels que les contes, la musique et les traditions locales encapsulés dans l'expérience de RV. En utilisant la théorie unifiée de l'acceptation et de l'utilisation de la technologie comme cadre d'orientation, nos résultats mettent en évidence le potentiel inégalé de la RV pour offrir des expériences culturelles immersives et éducatives. Ces expériences favorisent non seulement une appréciation et une compréhension plus approfondies

du patrimoine culturel immatériel parmi les publics mondiaux, mais mettent également en lumière les défis liés à l'accessibilité et à la familiarité de la RV. Nos recherches ouvrent la voie à de futures initiatives technologiques et éducatives visant à préserver et à promouvoir le patrimoine culturel immatériel par le biais d'expériences touristiques immersives.

Keywords: Réalité virtuelle, photogrammétrie, estimation de la pose humaine, conservation du patrimoine culturel, promotion du tourisme, HCI, utilisabilité, expérience utilisateur.

Table of contents

List of figures xiv

List of tables xvii

Nomenclature xix

1 Introduction 1
- 1.1 Context of the Thesis . . . 1
  - 1.1.1 Enhancing Tourism in Morocco: A Pillar of the Economy . . . 1
  - 1.1.2 A New Era of Exploration: Virtual Tourism in the Wake of COVID-19 2
  - 1.1.3 Bridging Worlds: The Cultural Lifeline of VR During Social Isolation 3
  - 1.1.4 The Sustainability Imperative: Redefining Travel in the Anthropocene 3
- 1.2 Thesis Approach . . . 3
- 1.3 Research Questions . . . 4
- 1.4 Thesis Outline . . . 5
- 1.5 List of Publications . . . 7

2 Background 9
- 2.1 Introduction . . . 9
- 2.2 3D Modeling . . . 9
  - 2.2.1 Photogrammetry . . . 10
  - 2.2.2 Laser Scanning . . . 14
  - 2.2.3 Accuracy measures and data comparison . . . 17
- 2.3 Human Pose Estimation . . . 19
  - 2.3.1 3D human pose estimation . . . 19
  - 2.3.2 Evaluation Metrics And Datasets . . . 22
- 2.4 Virtual Reality (VR) . . . 23
  - 2.4.1 The Evolution of VR Technology . . . 25
  - 2.4.2 Interactivity and Enhanced Capabilities in Virtual Environments . . 26

2.4.3 Selection and Manipulation Techniques in Virtual Environments . . 28
2.4.4 Navigational Dynamics in Virtual Environments 28
2.4.5 Immersion and Presence in VR 29
2.5 Conclusion 30

3 Literature Review 31
3.1 Introduction 31
3.2 Terrestrial Photogrammetry 31
3.3 3D Human Pose Estimation 33
3.3.1 Datasets for 3D HPE 33
3.3.2 Advancements and Challenges 33
3.4 Immersive Digital Technologies in Cultural Heritage 35
3.4.1 AR Applications 43
3.4.2 VR Applications 43
3.4.3 VR in Intangible Culture Heritage 45
3.4.4 MR Applications 49
3.5 Virtual Reality Evaluation 50
3.6 Conclusion 52

4 Low-cost Terrestrial Photogrammetry for 3D Modeling of Historic Sites 53
4.1 Introduction 53
4.2 Study area and data acquisition 54
4.2.1 Marinid Tombs 54
4.2.2 Taking pictures 54
4.2.3 Reference data 56
4.3 Methodology 57
4.3.1 From images to point clouds 57
4.3.2 Cloud-to-cloud distance and data comparison 58
4.4 Results and Analysis 60
4.4.1 Comparison of the bundle block adjustment for the cameras 60
4.4.2 Relative comparison between Canon and Redmi 63
4.4.3 Bundle block without using CPs 67
4.4.4 Effect of the number of CPs 68
4.5 Discussion & conclusion 68

5 Evaluating the Accuracy of Cloud-Based 3D Human Pose Estimation Tools 72
5.1 Introduction 72

5.2 Methodology . . . 73
5.2.1 Data preprocessing . . . 73
5.2.2 MOTiON by RADiCAL . . . 75
5.2.3 BVH projecting to 3D coordinates . . . 77
5.3 Evaluation . . . 78
5.3.1 Quantitative evaluation . . . 78
5.3.2 Qualitative Evaluation . . . 82
5.4 Discussion . . . 84
5.5 Threats to Validity . . . 85
5.5.1 Internal validity . . . 85
5.5.2 External Validity . . . 85
5.5.3 Construct Validity . . . 85
5.5.4 Conclusion Validity . . . 86
5.6 Conclusion . . . 86

6 Design, Development, and User-Centric Evaluation of El-FnaVR: An Immersive Virtual Reality Representation of Jemaa El-Fna in Marrakech for Intangible Cultural Heritage Experiences 88
6.1 Introduction . . . 88
6.2 CASE STUDY: Jemaa El-Fna . . . 89
6.2.1 Historical Background . . . 89
6.2.2 Activities . . . 90
6.3 El-FnaVR: Design and Implementation . . . 91
6.3.1 Stage 1: Content Requirements & Literature Study . . . 93
6.3.2 Stage 2: Design Features and Characteristics . . . 94
6.3.3 Stage 3: Visual content design . . . 96
6.3.4 Stage 4: Prototype implementation . . . 98
6.3.5 Stage 5: End user evaluation . . . 102
6.4 Results . . . 104
6.4.1 Subjective Evaluation Results . . . 104
6.4.2 Objective Evaluation Results . . . 108
6.5 Discussion . . . 109
6.6 Conclusion . . . 110

7 Conclusion 115
7.1 Summary and Main Results . . . 115
7.2 Implications . . . 116

7.3 Future Directions 117

References **119**

List of figures

1.1 Impact of the COVID-19 Pandemic on Morocco's Tourism Sector: A Comparative Analysis of Foreign and Moroccan Diaspora Tourist Arrivals and Tourism Receipts (2012-2020). 2
1.2 Disciplines covered in this thesis. 5

2.1 General photogrammetric workfow. 11
2.2 General classification of image matching methods. 13
2.3 Examples of TLS available in the market and their characteristics as of February 2020. 16
2.4 Frameworks of 3d single-person pose estimation [250]. (a) This method is done in one stage, i.e., directly from RGB image to 3D pose. (b) The approaches perform 3D HPE using a two-stage approach, i.e., it performs 2D HPE first and then uses the 2D keypoints to get 3D ones. (c) The 3D mesh is obtained using a regression stage on the 3D HPE outputs. 21
2.5 Frameworks of 2D and 3D multi-person pose estimation [137] (edited). (a) The top-down approach uses person detection techniques to determine the number of persons detected in the frame, and then it applies a 2D single-person estimation framework. (b) The bottom-up approach identifies each joint in the image and then associates each one with individuals. 22
2.6 Revisited Milgram's Reality-Virtuality Continuum by [204]. 24
2.7 The History of VR. 25

3.1 A Venn diagram illustrating the relationships between application categories. The assignment of applications to overlapping categories is partially determined by the use of shared technologies (e.g., Markerless Augmented Reality for the overlap between Exhibition Enhancement and Reconstruction) [25]. 37

3.2 Considering the identified application areas of cultural heritage, the diagram above illustrates the technical requirements for augmented, virtual, and mixed reality systems in both indoor and outdoor settings[25]. 38
3.3 Examples of AR applications for archaeological purposes include visualizing ancient architecture in its original location using location-based AR technology. 41
3.4 Some examples of tools designed to facilitate restoration and enhance analysis. These applications reveal features that are not visible to the naked eye. 42
3.5 UTAUT [225]. 52

4.1 Study area from Google Maps. 55
4.2 Marinid tombs ruins ('Les tombeaux des mérinides à Fès', 2016) 55
4.3 Location of the site at the top of the hill overlooking the old medina of Fez . 56
4.4 The position of the used total station. (b): The building facade and the chosen CPs. 57
4.5 Configuration of the 4, 6, 8, and 10 control points (green points) from 34 collected points (red points). Clockwise from top-left 59
4.6 The nearest neighbor distance that cloud-to-cloud (C2C) distancing is based on 60
4.7 Initial using parameters in Agisoft Metashape 61
4.8 Histogram and Q-Q plot of photogrammetry using Canon and Total station residuals with normal distribution 62
4.9 Histogram and Q-Q plot of photogrammetry using Redmi phone and Total station residuals with normal distribution (before outliers removing) 62
4.10 Histogram and Q-Q plot of photogrammetry using Redmi phone and Total station residuals with the normal distribution (after outliers removing) . . . 64
4.11 Histograms of C2C distances for Canon model (left) and Redmi one (right) 64
4.12 Histograms of C2C distances of Canon camera using Redmi phone camera as a reference model 65
4.13 C2C distances of Canon and Redmi relative comparison. Blue indicates low distances 66
4.14 Histograms of C2C distances for Canon model (left) and Redmi one (right) 67
4.15 Graph of mean variation as a function of CPs number 69
4.16 Graph of mean variation as a function of CPs number 70

5.1 Overview of the suggested approach. 74

5.2 (A): Skeleton model hierarchy of CMU. (B): Skeleton model hierarchy of RADiCAL. The red ones are the chosen joints to perform the quantitative evaluation. . . . 76
5.3 Rendering process of CMU and RADiCAL skeletons. The purpose is to obtain the 3D pose coordinate of joints from angles. . . . 78
5.4 The mean error results by joints. . . . 80
5.5 The results of MPJPE by scenarios. . . . 81
5.6 The thresholds and the corresponding PCK 82
5.7 The results of MPJPE by gender. . . . 83
5.8 The MPJPE results of comparison of joints by gender. . . . 83
5.9 The qualitative results of each scenario. . . . 87

6.1 Jemaa El-Fna, Marrakech, Morocco. . . . 90
6.2 Some Jemaa El-Fna activities and traditional street foods. (a): Halqa of snakes from Jemaa El Fna square. (b): Halqa of monkeys from Jemaa El Fna square. (c): Horse carriage. (d): Gnawa dancers group performing traditional dance and songs. (e): Orange Jus at the left. (f): The grilled meats and Harira at right. . . . 92
6.3 Sequential Stages of El-FnaVR Design and Implementation approach. . . . 93
6.4 Jamaa El-Fna mapping and tour design. . . . 97
6.5 The layers-based application structure. . . . 98
6.6 Visual Content Design of El-FnaVR. . . . 99
6.7 Development pipeline of El-FnaVR. . . . 100
6.8 Aid-navigation tools. (a): a point of interest placed beside the Halqa of snakes. (b): collectible coin. . . . 101
6.9 Movement using Point & Teleport Locomotion Technique principle [34]. . . 101
6.10 El-FnaVR system scenario. . . . 102
6.11 Customizing hand model from suggested models. . . . 102
6.12 Participants during the El-FnaVR experience. . . . 103
6.13 Depiction of Activities in the El-FnaVR Experience. (a) shows the horse carriage, which visitors can virtually ride and steer. (b) features the Gnawa Group engaged in song and dance. (c) showcases animated monkeys in various poses. (d) displays the option to handle animated snakes. (e) highlights street-side orange juice vendors. (f) illustrates some traditional street food offerings. . . . 112
6.14 VR evaluation test analysis. . . . 113
6.15 Video after VR evaluation test analysis. . . . 114

List of tables

3.1 The purposes and enabling technologies utilized by AR applications in cultural heritage. . . . 46
3.2 The purposes and enabling technologies utilized by VR applications in cultural heritage. . . . 47
3.3 Comparative Analysis of Cultural Heritage Projects Utilizing VR 48
3.4 The purposes and enabling technologies utilized by MR applications in cultural heritage. . . . 51

4.1 Characteristics of utilized DSLR and mobile phone cameras. 56
4.2 Obtained statistical parameters of the Canon and Redmi 7 cameras results . 63
4.3 Achieved statistical results of C2C distances between compared models and point clouds of total station as a reference 65
4.4 Achieved statistical results of C2C distances for relative comparison between Redmi as a reference and Canon as a compared model 65
4.5 obtained statistical results of C2C distances between compared models (without using CPs) and point clouds of the total station as a reference 67
4.6 Statistical results of the models as a function of CPs number 68

5.1 CMU dataset component after cleaning, filtering, and classification 76
5.2 Mean error by joints 79
5.3 MPJPE by scenarios 81
5.4 MPJPE by gender 82

6.1 The statistics of the demographic data for the participants in the study. . . . 105
6.2 Cronbach's Alpha Measures for Survey Consistency Across Different Groups. 106
6.3 Statistical results of the models as a function of CPs number 107
6.4 Chi-Square Test Results and Cramer's V Values for UTAUT Constructs Across Study Conditions. . . . 108
6.5 Summary of Hypotheses and Outcomes. . . . 108

6.6 Objective Evaluation Metrics: Mean Coin Collection in the VR Environment. 109

Nomenclature

Acronyms / Abbreviations

ABM Area-Based Matching

ALU Arithmetic Logic Unit

BA Bundle block adjustments

BVH BioVision Motion Capture

C2C Cloud-to-Cloud

CP Control Point

DSLR Digital Single-Lens Reflex

FBM Feature-Based Matching

FBX Filmbox

FPS Frame Per Second

GSD Ground Sampling Distance

HBIM Historic Building Information Modeling

HPE Human Pose Estimation

ICH Intangible Cultural Heritage

ICP Iterative Closest Point

LiDAR Light Detection and Ranging

MPJPE Mean Per Joint Position Error

PCK Percentage of Correct Keypoints

Q-Q quantile-quantile

RMSE Root Mean Square Error

SGM Semi-global matching

STD Standard Deviation

TLS Terrestrial Laser Scanning

UAV Unmanned Aerial Vehicles

UTAUT Unified Theory of Acceptance and Use of Technology

VR Virtual Reality

Chapter 1

Introduction

1.1 Context of the Thesis

In the rapidly evolving digital age, the intersection of Virtual Reality (VR) with cultural heritage preservation has opened new horizons for experiencing and safeguarding our collective history. The immersive nature of VR technology offers unparalleled opportunities for enhancing accessibility, understanding, and engagement with cultural heritage, both tangible and intangible. This thesis is situated within this transformative landscape, where VR emerges as a powerful tool against the backdrop of contemporary global challenges, including the COVID-19 pandemic and the pressing need for environmentally sustainable practices in tourism and heritage conservation.

1.1.1 Enhancing Tourism in Morocco: A Pillar of the Economy

Tourism plays a crucial role in the Moroccan economy, significantly contributing to its GDP and employment. Renowned cities such as Marrakech, Fez, and Casablanca, which are celebrated for their rich cultural heritage, annually attract millions of tourists. Nonetheless, the COVID-19 pandemic has highlighted the tourism sector's fragility in the face of global crises, pointing to the pressing need for innovative tourism and cultural heritage preservation methods, as shown in figure 1.1. VR technology offers a compelling avenue to revitalize Morocco's tourism industry by providing immersive experiences that enhance the country's attractiveness as a destination while promoting sustainable tourism practices. By integrating VR into Morocco's tourism strategy, the nation can present its cultural and natural heritage to a global audience, fostering economic growth and cultural exchange in a post-pandemic world. This assertion is supported by Doe and Smith (2023), who argue that VR technology not only mitigates the impact of travel restrictions but also opens new pathways for engaging

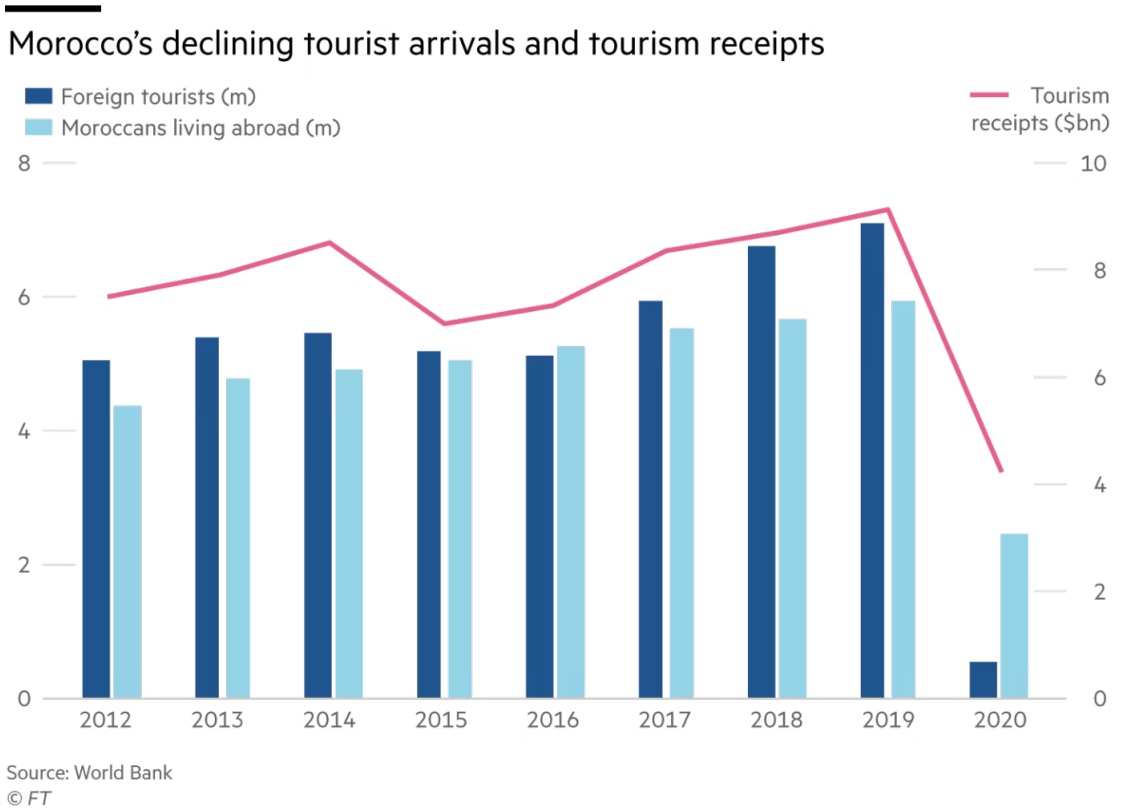


Fig. 1.1 Impact of the COVID-19 Pandemic on Morocco's Tourism Sector: A Comparative Analysis of Foreign and Moroccan Diaspora Tourist Arrivals and Tourism Receipts (2012-2020).

and educating tourists remotely, offering a resilient and adaptable model for future tourism (Doe & Smith, 2023).

1.1.2 A New Era of Exploration: Virtual Tourism in the Wake of COVID-19

The global outbreak of COVID-19 has precipitated an unprecedented crisis in the tourism sector, leading to a swift reimagining of travel and exploration. As international borders closed and social distancing became the new norm, virtual tourism, facilitated by VR, emerged as a vital bridge connecting people to the world's cultural treasures without leaving their homes. Research by Bec et al. (2021) underscores the resilience of VR in providing immersive experiences that compensate for the loss of physical travel, enabling virtual visitors to explore heritage sites in ways that were previously unimaginable. Lin et al. (2021) further illuminate the capacity of VR to enhance the visual aesthetic experience of cultural

landscapes, suggesting a paradigm shift in how we perceive and engage with heritage in the digital age.

1.1.3 Bridging Worlds: The Cultural Lifeline of VR During Social Isolation

The pandemic-induced isolation has heightened the importance of digital connectivity, with VR standing out for its ability to facilitate meaningful cultural exchanges amidst physical separation. Through immersive simulations, VR has enabled individuals to experience and learn about intangible cultural heritage (ICH) from diverse cultures, from traditional crafts to oral traditions and musical heritage, fostering a deeper global cultural understanding. Studies on the application of VR in preserving ICH, such as the African oral traditions and the immersive representation of the Dunhuang Caves, highlight VR's role in transcending geographical and temporal barriers, and promoting cultural literacy and empathy across diverse audiences.

1.1.4 The Sustainability Imperative: Redefining Travel in the Anthropocene

As humanity grapples with the dual challenges of recovering from a global health crisis and addressing the existential threat of climate change, the sustainability of travel and cultural engagement has come under scrutiny. VR presents an eco-friendly alternative to traditional tourism, offering a pathway to explore the world's cultural and natural wonders without the carbon footprint associated with physical travel. By digitizing ICH and providing virtual access to heritage sites, VR not only conserves resources but also opens up new avenues for sustainable tourism and heritage preservation, as evidenced by the research on VR-based digital display technologies and the virtual visualization of heritage sites.

1.2 Thesis Approach

The thesis adopts an interdisciplinary research approach that integrates methodologies from cultural studies, digital technology, and environmental sustainability, aiming to propose a comprehensive, innovative strategy for cultural heritage preservation and enhanced experience. By leveraging the precision of low-cost photogrammetry, the immersive potential of VR, and the interactive capabilities of advanced HPE tools, this approach seeks to create accessible, engaging, and educational cultural heritage experiences. This multidisciplinary

method synthesizes insights from a broad range of sources, including recent case studies and empirical research, contributing both to the academic field of cultural heritage preservation and offering practical insights and solutions for heritage practitioners and technologists. The goal is to provide a comprehensive understanding of the potentialities and challenges associated with the use of VR in the cultural heritage domain, thus fostering a sustainable and enriching engagement with cultural heritage.

1.3 Research Questions

This thesis explores the application of digital technologies in cultural heritage preservation. It examines methods to optimize terrestrial photogrammetry for 3D reconstruction of heritage sites, develops a VR application for Jemaa El Fna to provide immersive cultural experiences, and evaluates the performance of the MOTiON by RADiCAL cloud-based 3D Human Pose Estimation tool against traditional motion capture systems. Each section seeks to assess these technologies' effectiveness, limitations, and potential improvements in their respective fields.

1. In what ways can terrestrial photogrammetry be optimized for the 3D reconstruction of cultural heritage sites?
2. How do these methods impact the sustainability and accuracy of heritage preservation?
3. How accurate are cloud-based 3D HPE tools in estimating human poses compared to ground truth data?
4. Can cloud-based 3D HPE tools serve as a feasible alternative to traditional motion capture systems in various scenarios?
5. What improvements are necessary for cloud-based 3D HPE tools to extend their applicability to high-precision tasks, such as advanced biomechanical analysis, sports performance analysis, or precise VR interaction?
6. How does VR technology enhance the experience of engaging with intangible cultural heritage, specifically through the use of immersive environments like El-FnaVR?
7. What are the impacts of VR experiences on user engagement and perception when compared to traditional media in the context of cultural heritage?
8. How does the sequence of media exposure (VR followed by traditional media vs. traditional media alone) affect the engagement levels and enjoyment of users in cultural heritage experiences?

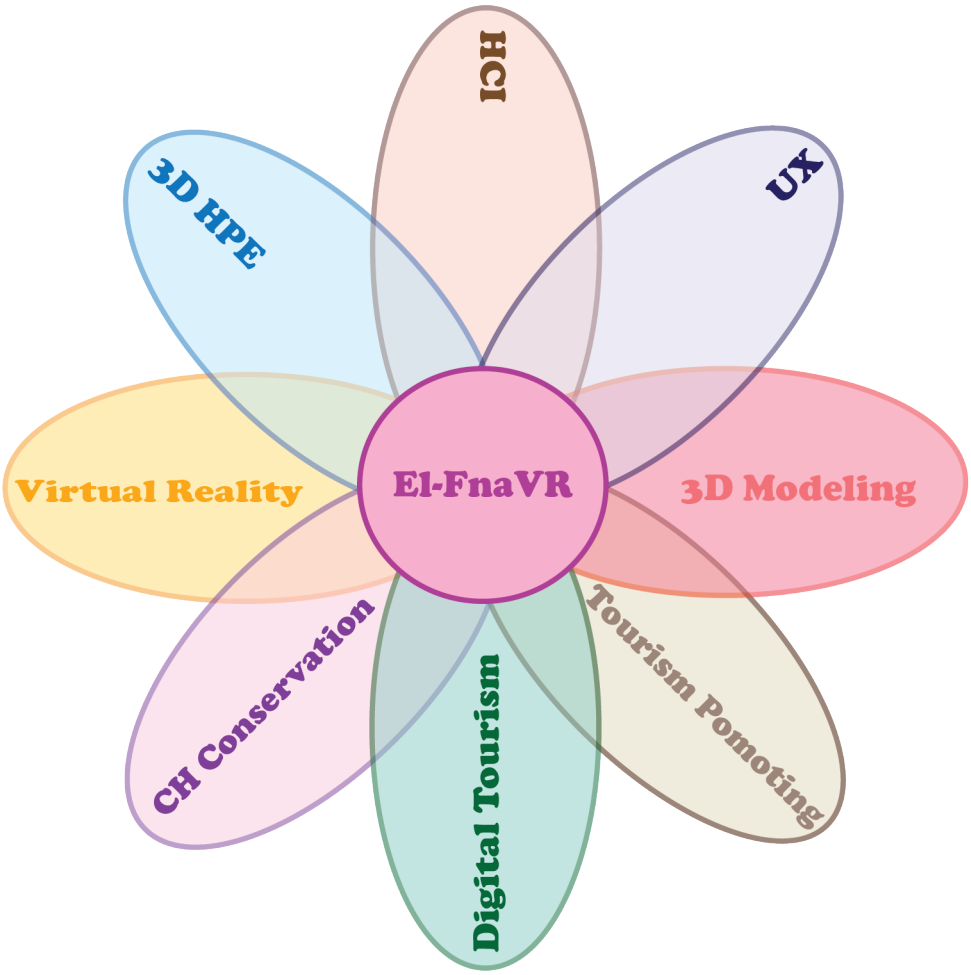


Fig. 1.2 Disciplines covered in this thesis.

The questions addressed in this work fall into several disciplines, including VR, 3D modeling, 3D human pose estimation (3D HPE), user experience (UX), digital tourism, tourism promoting, tangible and intangible cultural heritage conservation, human-computer interaction (HCI), as illustrated in Figure 1.2.

1.4 Thesis Outline

The current chapter sets the stage for the thesis, outlining the research context, motivation, and research questions. It introduces the central role of emerging technologies in cultural heritage preservation and the innovative approach this thesis proposes to enhance cultural heritage experiences through VR, photogrammetry, and 3D HPE.

Delving into the technical and theoretical foundations, **the second chapter** provides a comprehensive overview of photogrammetry (including terminology, the use of laser

scanners, and accuracy considerations), 3D HPE (covering types, datasets, and applications), and VR (exploring its development, immersive capabilities, and role in cultural heritage). This background establishes the necessary context for understanding the subsequent research and analyses.

The third chapter reviews existing literature relevant to the thesis topics, critically analyzing studies that have explored the use of photogrammetry, 3D HPE, and VR in cultural heritage contexts. It identifies gaps in the current research landscape and positions the thesis within the broader academic discourse, highlighting its unique contributions.

The fourth chapter introduces the inaugural contribution of this research, exploring the application of low-cost terrestrial photogrammetry for historic site modeling. It elaborates on the methodology, execution, and outcomes of utilizing an entry-level smartphone camera for close-range photogrammetry, presenting it as a cost-effective and widely available tool. This chapter aims to assess the precision, viability, and potential of such technology in detailed heritage documentation. For this purpose, the Marinid Royal Necropolis, located in the UNESCO-listed city of Fez, Morocco, was documented and transformed into a 3D model using both a smartphone and a DSLR camera. The findings were then benchmarked against control points (CPs) gathered through a total station, with their accuracy scrutinized via statistical metrics. Furthermore, to examine the influence of various factors, Cloud-to-Cloud (C2C) distance measurements were employed to assess each 3D model's fidelity.

The fifth chapter delivers a comprehensive scientific evaluation of MOTiON by RADiCAL, a pioneering cloud-based 3D Human Pose Estimation (HPE) tool leveraging deep learning and cloud computing technologies. The assessment focuses on its accuracy and utility in capturing human motion. Utilizing the cleaned and filtered CMU dataset for this analysis, the tool's performance was meticulously compared to the ground truth, employing two critical metrics: the Mean Per Joint Position Error (MPJPE) and the Percentage of Correct Keypoints (PCK). Additionally, this evaluation explores the tool's potential to significantly enhance VR experiences' interactivity and realism, underlining its implications for cultural heritage preservation and interpretation.

The sixth chapter delves into the El-FnaVR project, a pioneering VR application designed to encapsulate the dynamic spirit of Marrakech's famous UNESCO-listed Jemaa El-Fna square, Morocco. This section explores the intricate development and evaluative processes behind creating an immersive VR experience that not only integrates photogrammetric models for a realistic representation of the square but also enriches this virtual environment with lifelike animated characters and spatial 3D audio. These elements work in tandem to authentically recreate the square's unique ambiance, from the bustling activities to the captivating performances. The chapter outlines the comprehensive five-stage design and

development journey of El-FnaVR, emphasizing its innovative features and the meticulous attention to detail that facilitates an engaging, user-friendly, and socially appealing VR experience. A conducted user study, with feedback from participants, underscores the application's success in delivering an immersive and educational experience, highlighting its effectiveness in engaging users and enhancing their understanding of cultural heritage sites.

Concluding the thesis, **the final chapter** synthesizes the findings from the three studies, reflecting on the implications for cultural heritage preservation and VR-based tourism. It discusses the theoretical and practical contributions of the research, limitations encountered, and suggestions for future work in the field.

1.5 List of Publications

Most sections of this thesis have been published in various scientific journals and conferences, a brief overview of which is provided below. Other sections are currently undergoing a peer-review process.

Journal Articles

- H. Khalloufi, M. Zaifri, M. Kadri, A. Benlahbib, F. Z. Kaghat and A. Azough, "El-FnaVR: An Immersive Virtual Reality Representation of Jemaa El-Fna in Marrakech for Intangible Cultural Heritage Experiences," in IEEE Access, vol. 12, pp. 9331-9349, 2024, doi: 10.1109/ACCESS.2023.3347195.

- H. Khalloufi, M. Zaifri, M. Kadri, A. Benlahbib, F. Z. Kaghat and A. Azough, "Evaluating the Accuracy of Cloud-Based 3D Human Pose Estimation Tools: A Case Study of MOTiON by RADiCAL" in International Journal of Advanced Computer Science and Applications (IJACSA), doi: 10.14569/IJACSA.2024.0150401.

- H. Khalloufi, F. Z. Kaghat, A. Azough, and N. Ennahnahi "Low-Cost Terrestrial Photogrammetry For 3D Modeling Of Historic Sites: A Case Study Of The Marinids' Royal Necropolis City Of Fez, Morocco," in Mediterranean Archaeology and Archaeometry Vol. 20, No 3, (2020), pp. 257-272, doi: 10.5281/zenodo.3930412.

- M., Zaifri., H., Khalloufi, F.Z., Kaghat, A., Benlahbib, A., Azough, K.A., Zidani,. "Enhancing Tourist Experiences in Crowded Destinations through Mobile Augmented Reality: A Comparative Field Study". International Journal of Interactive Mobile Technologies (iJIM), 17(20), (2023), pp. 1–22, doi: 10.3991/ijim.v17i20.42273

- M., Zaifri., H., Khalloufi, F.Z., Kaghat, A., Benlahbib, A., Azough, K.A., Zidani,. "From Earlier Exploration to Advanced Applications: Bibliometric and Systematic Review of Augmented Reality in the Tourism Industry (2002–2022)." Multimodal Technologies and Interaction (MTI) 7(7): 64. doi: 10.3390/mti7070064

Conference Paper

- M. Kadri, H. Khalloufi and A. Azough, "V-Museum: A Virtual Museum Based on Augmented and Virtual Realities for Cultural Heritage Mediation," 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2020, pp. 1-5, doi: 10.1109/ISCV49265.2020.9204253.

Chapter 2

Background

2.1 Introduction

This chapter delves into the foundational concepts and technological advancements that underpin the fields of 3D modeling, 3D HPE, and VR, which are crucial for the understanding and application of these technologies in various domains. Beginning with 3D modeling, the chapter explores its evolution and its pivotal role in digital documentation and preservation, especially concerning cultural heritage. It then transitions into human pose estimation, discussing its significance, methodologies, and challenges, particularly in translating 2D representations into 3D reconstructions. The chapter culminates with an examination of VR, detailing its historical development, immersive experiences, and the impact on user interaction and perception within virtual environments. This background provides a comprehensive overview essential for grasping the current state and potential future directions of these dynamic and interrelated fields.

2.2 3D Modeling

In the context of digital transformation, the documentation of cultural heritage has evolved significantly, guided by the diverse range of 3D sensing technologies available today. This evolution is characterized by the ongoing challenge of balancing cost, effort, and precision. Technological advancements have made previously cost-prohibitive methods more accessible, such as transitioning from analog metric cameras to digital non-metric cameras in photogrammetry. Despite these advancements, laser scanning technology has increasingly become the go-to solution for heritage documentation, a trend supported by references to its establishment as a revolutionary 3D mapping tool since the 1980s. Contrary to traditional

total stations, Terrestrial Laser Scanning (TLS) technology offers the capability to record environments quickly and with relative accuracy, presenting an ease of use that has favored its adoption [48, 104, 127, 161].

Photogrammetry, on the other hand, dates back to the analog era but has seen a resurgence in the past two decades. This resurgence is attributed to significant advancements in image processing algorithms and the availability of more powerful computing resources. Recognized as a cost-effective alternative to laser scanning, photogrammetry requires only 2D images and dedicated processing software for its application. This method has been extensively utilized in architectural documentation, including the extraction of orthophotos, creation of vector models for façades, and generation of 3D models, as well as providing inputs for Historic Building Information Modeling (HBIM) [23, 47].

The study of sensor integration, reflecting the complementary nature of photogrammetry and laser scanning, has emerged as a significant area of research. This integration, along with the exploration of alternative low-cost sensors such as RGB-D cameras, fisheye lenses, panoramic 360° cameras, and even smartphones, has broadened the scope of tools available for heritage documentation [129, 184, 163, 161, 21].

These methods have been deployed in various research endeavors to document architectural heritage, demonstrating their effectiveness in producing detailed representations of historical sites. The employment of both photogrammetry and laser scanning in documenting various types of vernacular buildings underscores their vital role in cultural heritage preservation, facilitating a nuanced understanding of historical contexts [47]. This narrative on the evolution of heritage documentation technologies reflects the dynamic interplay between traditional methods and cutting-edge, low-cost solutions, highlighting their collective contribution to safeguarding cultural legacies in the digital era.

2.2.1 Photogrammetry

In the past decade, photogrammetry has experienced significant advancements and a resurgence of interest, driven largely by innovations in sensor technology [162]. This resurgence is further supported by advancements in dense matching algorithms [187] and the widespread adoption of UAVs or drones [185], enhancing its applicability in various fields, including heritage documentation. For heritage conservation, close-range photogrammetry is particularly favored, allowing for the design of a ground sampling distance (GSD) that meets specific project needs, thus facilitating the capture of detailed architectural features. The technique's ability to derive photo-realistic textures from captured images offers a distinct advantage over alternative methods. Moreover, the introduction of UAVs has expanded the perspectives available to photogrammetry, incorporating aerial views for comprehensive

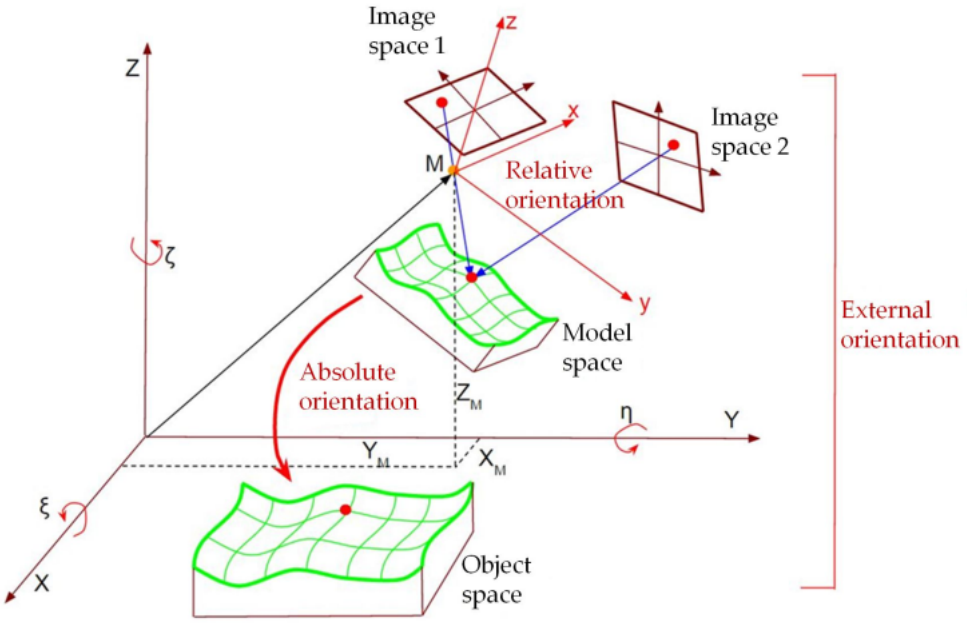


Fig. 2.1 General photogrammetric workfow.

documentation. However, processing photogrammetric data demands considerable time, expertise, and resources.

Traditionally, photogrammetry necessitated the use of well-calibrated, often metric, sensors. Currently, advancements in computing power and algorithmic efficiency have broadened the range of sensors suitable for photogrammetric analysis, including DSLR cameras, which are preferred for standard heritage documentation projects due to their stability and customizable settings [186]. Additionally, the quality and applicability of lower-end sensors, such as smartphone cameras and spherical panoramic cameras, for documenting heritage sites have significantly improved.

The workflow of photogrammetry is well-defined and illustrated in Figure 2.1 The conventional process starts with interior orientation and camera calibration, proceeding to relative and absolute orientations. The block bundle adjustment technique is pivotal in this process, as it computes the projection of points from the image space to the object space in a unified step, effectively solving the external orientation problem (detailed further in Chapter 2). External orientation is essential for establishing the relationship between the image space and the object space, determined by the camera's position and orientation within the object space coordinate system, described mathematically by three rotation and three translation parameters [198].

A fundamental set of equations in photogrammetry, known as the collinearity equations, are foundational to the field [93, 141]:

$$\begin{cases} x' = x_0 - f \cdot \dfrac{r_{11}(X - X_s) + r_{12}(Y - Y_s) + r_{13}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \\ y' = y_0 - f \cdot \dfrac{r_{21}(X - X_s) + r_{22}(Y - Y_s) + r_{23}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \end{cases} \tag{2.1}$$

Where: f: focal length x_0, y_0: coordinates in the image space r: rotation matrix X, Y, Z: coordinates in the object space X_s, Y_s, Z_s: coordinates of camera projection centre

Equation 2.1 is pivotal in the bundle adjustment procedure, delineating the essential relationship between the image and object space. This equation can be further augmented to encompass image distortion parameters, thereby facilitating simultaneous self-calibration and block bundle adjustment computations [156]:

$$\begin{cases} x' = x_0 + f \cdot \dfrac{r_{11}(X - X_s) + r_{12}(Y - Y_s) + r_{13}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} + \delta x_r + \delta x_t \\ y' = y_0 - f \cdot \dfrac{r_{21}(X - X_s) + r_{22}(Y - Y_s) + r_{23}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} + \delta y_r + \delta y_t \end{cases} \tag{2.2}$$

With:

$$\begin{cases} \delta x_r = x(K_1 r^2 + K_2 r^4 + K_3 r^6) \\ \delta y_r = y(K_1 r^2 + K_2 r^4 + K_3 r^6) \end{cases} \tag{2.3}$$

$$\begin{cases} \delta x_t = P_1(2x^2 + r^2) + 2P_2 xy \\ \delta y_t = 2P_1 xy + P_2(2y^2 + r^2) \end{cases} \tag{2.4}$$

The parameters K_1, K_2, and K_3 are associated with radial distortion, and $P1$ and $P2$ with tangential distortion. Concurrently, the field of computer vision has independently evolved solutions to the image orientation challenge. The introduction of Structure from Motion (SfM) marked a significant milestone in photogrammetry, streamlining many manual tasks, including the marking of tie points. While similar automation efforts have historical precedents in photogrammetry [92], the contributions from the field of computer vision have been substantial in advancing 3D data generation algorithms. The advent of dense matching marked another turning point for photogrammetry, enabling the creation of dense point clouds from images, a task that was once limited to generating only sparse points. This resurgence established photogrammetry as a formidable alternative to laser scanning for generating dense point clouds.

The majority of 3D reconstruction programs that utilize imagery employ proprietary algorithms for dense point cloud generation. Remondino [2014] attempted to categorize the various approaches to dense matching. Fundamentally, these methods are distinguished by

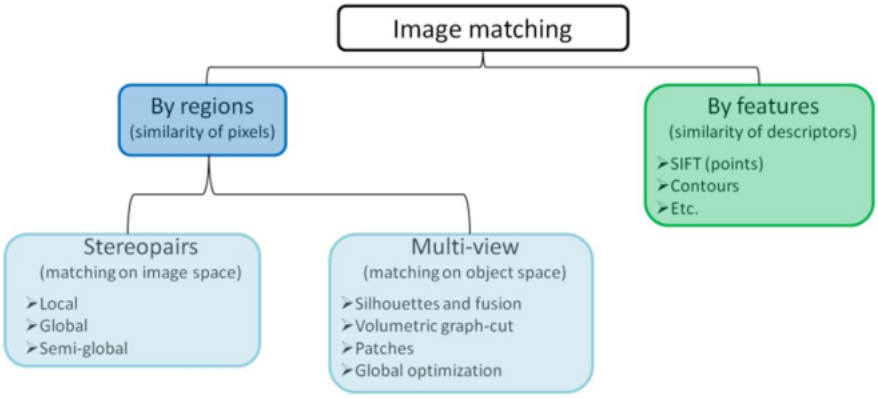


Fig. 2.2 General classification of image matching methods.

either feature matching, which involves comparing descriptors, or area-based matching, which compares grayscale values within a predefined search window. Following correspondence identification, a straightforward mathematical operation calculates the spatial coordinates of the object in object space. Feature matching is commonly referred to as feature-based matching (FBM), while the latter method is known as area-based matching (ABM). Figure 2.2 in the text provides a simplified diagram of this categorization.

Algorithms categorized under Area-Based Matching (ABM) boast remarkable potency, with the capability to achieve matching accuracy as precise as 1/50th of a pixel. ABM, however, relies on the presence of texture in the objects and requires initial approximate values. Conversely, Feature-Based Matching (FBM) algorithms are less reliant on texture presence, but they typically produce less densely populated point clouds. Notable FBM techniques include the SIFT [140], SURF [24], and Harris detectors [98].

In modern practices, FBM algorithms are often employed initially to construct a sparse point cloud [218]. This preliminary point cloud then serves as a reference for the subsequent dense matching stage, which utilizes ABM techniques. Regarding ABM, there are two distinct sub-categories based on the spatial domain of the matching process. The first conducts matching along epipolar lines within the image space, further subdivided into local and global methods by [218]. Local methods, which operate on a per-pixel basis within a search window, inherently smooth object edges [107]. Global methods, in contrast, involve energy minimization across the entire image, incorporating explicit smoothing techniques. Semi-global matching (SGM), as proposed by [107], is an approach that computes energy minimization along epipolar lines and additional directions surrounding the pixel.

The second sub-category of ABM undertakes the matching process within the object space. For this method, an initial sparse point cloud—potentially sourced from an FBM process—is essential as a foundational dataset. With this sparse cloud in hand, the algorithm

organizes images based on their relative positioning to the object under reconstruction. Dense matching then occurs around a 3D point in the sparse cloud, observed by a specific image cluster. This area of focus incrementally expands to adjacent pixels within the image space, incorporating visibility constraints to refine the matching results [83].

In the realm of heritage documentation, photogrammetry is typically applied in a close-range setup, using either terrestrial or aerial platforms. Terrestrial photogrammetry frequently involves DSLR cameras, which have been widely recognized for yielding superior results among non-metric cameras [Bedford, 2017] and have also been utilized for metrological and referencing tasks [154, 38]. However, the terrestrial approach is somewhat constrained when documenting tall structures due to its limited vantage points. Historically, this limitation has been addressed using various techniques, such as cranes [4, 77] or passive platforms like kites and balloons [226], but the current trend leans more towards drone utilization [91, 97, 162]. The use of drones in close-range photogrammetry introduces a trade-off between the payload and image quality. While high-quality cameras, such as DSLRs, necessitate larger and more costly drones, smaller drones are limited to supporting lower-quality cameras. Nevertheless, UAVs inherently offer an advantage in capturing aerial images and accessing perspectives that are unattainable from the ground.

While the discussion thus far has primarily focused on sensors akin to the classical pin-hole camera, alternative lens types exist, each with unique benefits and drawbacks. For instance, fish-eye lenses, despite their significant image distortion, afford a much broader field of view, which can be advantageous for surveying confined spaces [78]. An expanded form of this is spherical photogrammetry, which often employs 360° equirectangular panoramic imaging [20]. The key appeal of these sensors lies in their cost-effectiveness and the simplicity of data capture, despite the trade-off in image quality. However, recent advancements have significantly enhanced the quality and precision of these sensors [20].

2.2.2 Laser Scanning

Since its inception in the 1980s, laser scanning technology has revolutionized the field of 3D spatial data acquisition. Terrestrial Laser Scanning (TLS) systems offer rapid and relatively precise environmental recordings, surpassing traditional total stations in both ease of operation and data collection speed. However, the voluminous data produced, particularly when documenting complex heritage structures, results in substantial point cloud files [23], and may lead to occlusions, leaving gaps in the data [126]. TLS also faces challenges in accurately capturing colors and textures, despite the integration of cameras [99].

Laser scanning utilizes laser beams and rotating mirrors or prisms to systematically measure distances, enabling swift and dense object measurements [94]. As an active sensor

system, it operates independently of ambient lighting conditions, emitting laser pulses and detecting their reflections, which permits operation under nearly any environmental conditions, including darkness [189]. Unlike photogrammetric models, well-calibrated TLS systems can provide scaled measurements without additional processing. Moreover, TLS can record additional data, such as the intensity of returned laser pulses, which aids in material identification, and when coupled with RGB cameras, provides colorized point clouds.

Laser scanning technologies are categorized by application into aerial (ALS) and terrestrial (TLS) systems, with 'LIDAR' commonly referring to the former [229]. Less frequently, the term 'LADAR' is encountered in literature. [211] distinguished laser scanners by their field of view:

- Panoramic scanners capture an entire hemisphere.
- Camera scanners have restricted horizontal and vertical fields of view.
- Hybrid scanners combine unrestricted rotation on one axis with limitations on another.

Furthermore, scanners are classified by their distance-measuring techniques [94]:

- Pulse-based scanners calculate distances by measuring the time it takes for a laser pulse to return to the sensor, analogous to GNSS's absolute positioning. These are suited for longer distances where precision is less critical.
- Phase-based scanners emit modulated waves and measure the phase difference upon return, similar to GNSS's relative positioning. They offer higher precision but are limited to mid-range distances due to phase ambiguity risks.
- Triangulation-based scanners determine distances by forming and solving mathematical triangles between the laser emitter, object, and receiver. They are highly precise but limited to short-range applications.

Figure 2.3 offers a comparative overview of TLS technologies available as of 2020. The accuracy of the resulting point cloud also depends on the registration of multiple scan stations, ensuring geometric correctness. Registration aligns scans within the same coordinate system while georeferencing assigns scans to a geodetic system [188]. Georeferencing can be direct—by placing the scanner on pre-measured control points—or indirect, requiring scan overlap and involving initial coarse alignment and subsequent fine registration[128, 74]. Typically, artificial targets, such as planar or spherical shapes, are used to facilitate this alignment [94]. The mathematical underpinnings of this transformation are detailed in the subsequent section:



Fig. 2.3 Examples of TLS available in the market and their characteristics as of February 2020.

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + m \cdot \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} \tag{2.5}$$

Where:

- X_1, Y_1, Z_1: coordinates in the first (or reference) system
- X_2, Y_2, Z_2: coordinates in the second (or transformed) system
- t_x, t_y, t_z: 3D translations
- r: rotation matrix containing rotations on the three axes
- m: scaling factor

It should be noted that, in many instances, the inclusion of a scaling factor can be omitted due to the precision of Terrestrial Laser Scanning (TLS) systems that directly measure distances without its necessity, unlike photogrammetric methods that may require scaling.

Should there be evidence indicating the need for a scaling factor, its calculation can be undertaken [229].

Subsequently, a more refined transformation, often employing the Iterative Closest Point (ICP) algorithm [Besl and McKay, 1992], is applied to finely adjust the alignment of the individual point clouds. This refinement is iterative and presupposes an accurate initial alignment of the point clouds, hence the preliminary two-step process as previously outlined. While there are other iterative fine registration algorithms available in scholarly works, they fall beyond the ambit of this thesis. Readers with a keen interest in these methods are directed to consult [229] for a comprehensive discussion on the topic.

2.2.3 Accuracy measures and data comparison

Cloud-to-cloud comparison

CloudCompare is a pivotal open-source software in the photogrammetry field, extensively used for analyzing 3D point clouds and meshes. It plays a crucial role in verifying the accuracy of 3D models derived from photographic data through various processes, including noise reduction and point cloud comparison [66].

One of the significant functionalities of CloudCompare is its nearest neighbor calculation. This feature is vital for registering point clouds and measuring distances. The algorithm often utilizes a k-D tree structure to efficiently identify the closest points between datasets, which is an essential step in assessing the quality of photogrammetric models [195].

Integrating CloudCompare into photogrammetry workflows allows for precise model verification, thereby improving the quality and dependability of photogrammetric data. This is particularly valuable in applications such as cultural heritage preservation, where accuracy is paramount [183].

Statistical analysis

Accuracy measures are derived from coordinate differences of some points obtained from a photogrammetric solution and another reference one. The reference values should be another method that can produce reference values with an accuracy that is at least 3 times higher [87, 1]. Such points which can be determined in both systems are named checkpoints or control points. These differences are called ‘discrepancies or errors’ as shown in Equations:

$$E_x = (x_{pho} - x_{tot}), E_y = (y_{pho} - y_{tot}), E_z = (z_{pho} - z_{tot}) \tag{2.6}$$

$$D = \sqrt{(E_x)^2 + (E_y)^2 + (E_z)^2} \tag{2.7}$$

Where:

- E_x, E_y, E_z = Errors or residuals in x, y and z directions.
- $x_{tot}, y_{tot}, z_{tot}$ = reference coordinates from total station

$x_{pho}, y_{pho}, z_{pho}$ = estimated coordinates from image measurements
D = error length based on coordinate residuals

From the differences, the Root Mean Square Error (RMSE), Mean value, and Standard Deviation (σ) are calculated by using Equations (3), (4) and (5).

$$RMSE_x = \sqrt{\frac{\sum(E_x)^2}{n}}, RMSE_y = \sqrt{\frac{\sum(E_y)^2}{n}}, RMSE_z = \sqrt{\frac{\sum(E_z)^2}{n}} \tag{2.8}$$

$$RMSE_{total} = \sqrt{(RMSE_x)^2 + (RMSE_y)^2 + (RMSE_z)^2} \tag{2.9}$$

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} \tag{2.10}$$

Where:

- $RMSE_x, RMSE_y, RMSE_z$ = root mean square error for X, Y, Z coordinates
- $RMSE_{total}$ = total root mean square error for a point
- n = number of observations in the sample.
- x = the observed values of the sample items
- $\bar{x}$ = the mean value of these observations
- σ = the sample standard deviation

After calculating the errors, a histogram of the errors will inform whether the distribution of the errors is normal or not. The curve for a normal distribution can be superimposed. If the curve does not match the data very well, then the distribution is not normal. A better diagnostic plot in order to detect a deviation from the normal distribution is the so-called quantile-quantile (Q-Q) plot. Also, the Skewness value gives an indication of departure from symmetry in a distribution and is expressed as:

$$Skewness = \frac{\sum_{i=1}^{n}(x_i - \mu)^3}{(n-1)\sigma^3} \tag{2.11}$$

The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. The Kurtosis is a measure of the data are peaked or flat relative to a normal distribution and is expressed as:

$$Kurtosis = \frac{\sum_{i=1}^{n}(x_i - \mu)^4}{(n-1)\sigma^4} \tag{2.12}$$

If the distribution is perfectly normal, Skewness and Kurtosis values of zero and three are obtained, respectively.

2.3 Human Pose Estimation

2.3.1 3D human pose estimation

Over the past few years, there has been a growing interest in 3D HPE due to its ability to provide accurate information about the 3D structure of the human body. 3D HPE seeks to predict the location of body joints in 3D space. It can be applied to a variety of situations (e.g., 3D animation movies, extended realities, and cloud-based 3D action estimation). Even though 2D HPE has recently seen significant advancements, 3D HPE is still a challenging task to complete. Recent research in the field of computer vision has been focused on the extraction of 3D human pose estimation (HPE) from monocular images or videos. Those are a 2D representation of a 3D scene, resulting in the loss of one dimension. As a result, researchers have been working on developing algorithms and techniques to accurately estimate the 3D pose of human subjects from these 2D images. 3D human pose estimation can be a well-defined problem that is solvable using information fusion methods if there are multiple perspectives or additional sensors such as IMU and LiDAR available. However, one drawback of using deep learning models for this task is their high data dependence and sensitivity to data collection circumstances. Extensive amounts of annotated data are necessary for these models to learn accurate representations of input and output spaces, and factors such as lighting conditions, camera positions, and background can influence their performance negatively.

While obtaining accurate two-dimensional posture annotations for human datasets is relatively straightforward, obtaining accurate three-dimensional pose annotations is considerably more challenging and cannot be done manually. Furthermore, datasets are often collected in

controlled indoor settings that focus on specific activities, making them biased towards these scenarios. Recent studies have shown that models trained on such biased datasets tend to perform poorly when applied to other datasets, as demonstrated by cross-dataset inference [231, 244].

Single-person 3D HPE

The strategies of Single-person 3D HPE can be categorized as model-free or model-based methods. The first one can be divided into two categories:

- **Direct estimate techniques:** instead of first estimating the 2D pose representation, some algorithms in 3D human pose estimation employ direct estimation techniques, as seen in [158, 168], to directly infer the 3D human position from 2D images.
- **2D to 3D lifting techniques:** The process of inferring 3D poses from intermediate 2D pose pairings is inefficient because it requires multiple network inferences. Human body models are not used in the model-free approaches for recreating 3D human representations. Standard 2D HPE models are used in the first stage to estimate the 2D posture, and 2D to 3D lifting is used in the second stage to construct the 3D pose, such as [45, 147], and [168]. 2D heatmaps rather than 2D poses were used as intermediate representations to estimate 3D posture [219, 251]. Through distance matrix regression, Moreno-Noguer [159] deduced the 3D human position from the distances between the joints in the 2D and 3D body (EDMs). When normalization techniques are used, EDMs are invariant to scaling invariance as well as in-plane image rotations and translations. A Paired Ranking Convolutional Neural Network (PRCNN) was created to predict the depth ranking of pairwise human joints [234]. The 3D pose was then regressed from the 2D joints and the depth ranking matrix using a coarse-to-fine pose estimator. [132, 202, 115] were the first to develop numerous, different 3D pose hypotheses.

Parametric body models, such as kinematic and volumetric models, are utilized by model-based methods to estimate human position, as illustrated in Figure 2.4. The kinematic model represents the body as a series of joints and articulating bones, and in recent years, it has garnered increasing attention in the field of 3D human pose estimation. A temporal convolution network was suggested for estimating 3D posture from sequential 2D sequences using 2D key points [169]. A Short-Term Long Memory (LSTM) unit and shortcut connections were employed in a recurrent neural network to leverage temporal information from human pose data [108]. The Skinned Multi-Person Linear (SMPL) model is among the most commonly utilized volumetric models in the field of 3D HPE, as evidenced by its implementation in works such as [242, 249].

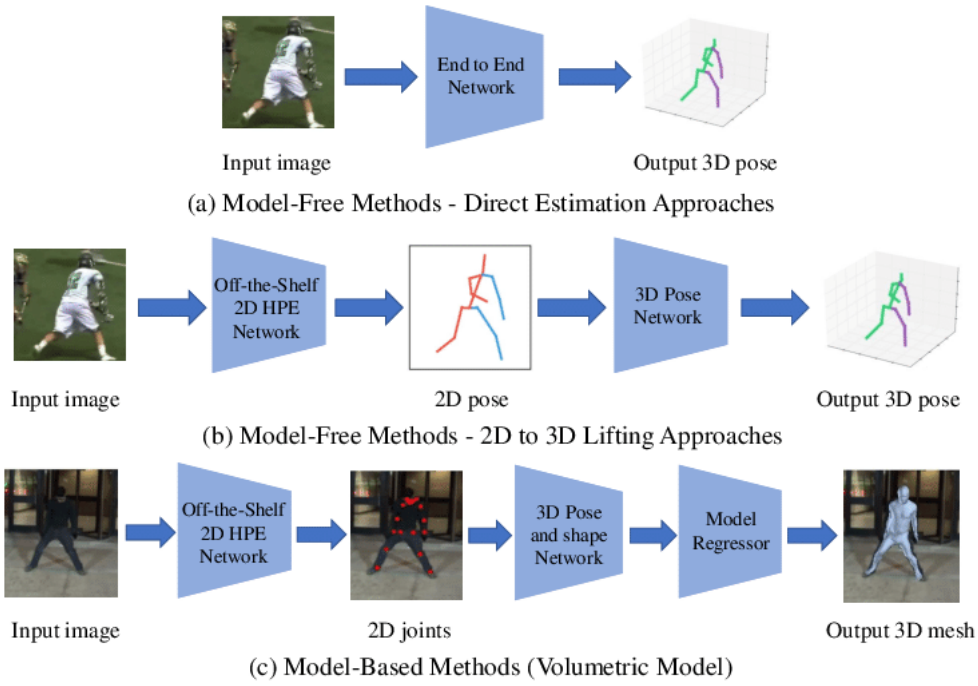


Fig. 2.4 Frameworks of 3d single-person pose estimation [250]. (a) This method is done in one stage, i.e., directly from RGB image to 3D pose. (b) The approaches perform 3D HPE using a two-stage approach, i.e., it performs 2D HPE first and then uses the 2D keypoints to get 3D ones. (c) The 3D mesh is obtained using a regression stage on the 3D HPE outputs.

Multi-person 3D HPE

There are two approaches for 3D multi-person human pose estimation from monocular RGB images or videos, which are classified into top-down and bottom-up categories. These approaches are illustrated in Figure 2.5.

- **Top-down approaches:** They use human detection to estimate each person's position. Each time a person is identified, 3D pose networks estimate their root (the human body's central joint) coordinate and their 3D root-relative posture. Candidate areas of each individual were targeted to produce prospective postures and then utilized a regressor to improve the pose suggestions jointly [193]. The LCR-Net technique, which involves localization, classification, and regression, performed well on datasets collected in controlled environments, but not on images captured in natural settings. To address this limitation, LCR-Net++ was introduced, which utilizes synthetic data augmentation during training to improve performance [194]. The 3D multi-person HPE module was enhanced with scene constraints and semantic segmentation [245]. The 3D temporal assignment problem was also tackled by the Hungarian matching

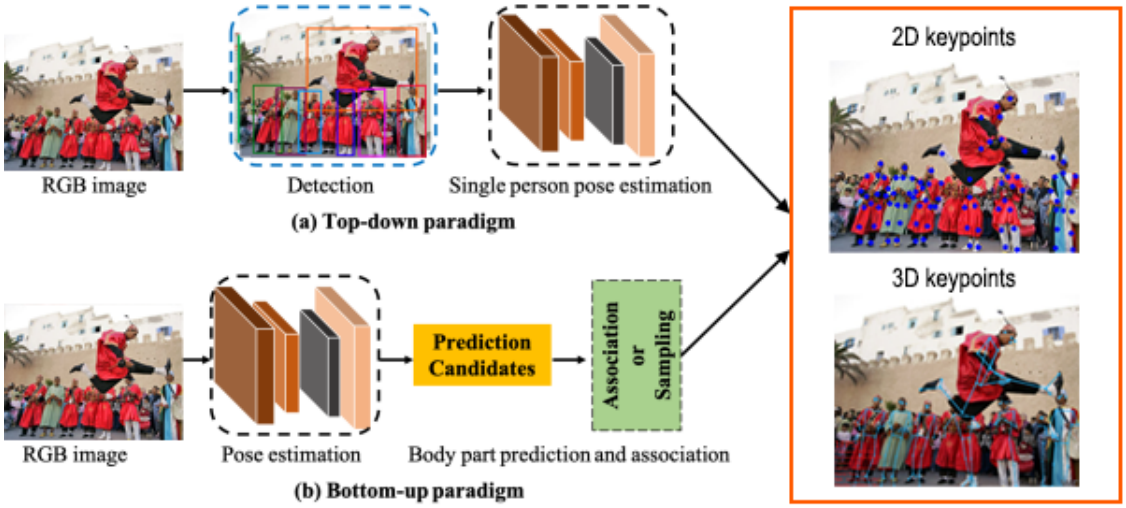


Fig. 2.5 Frameworks of 2D and 3D multi-person pose estimation [137] (edited). (a) The top-down approach uses person detection techniques to determine the number of persons detected in the frame, and then it applies a 2D single-person estimation framework. (b) The bottom-up approach identifies each joint in the image and then associates each one with individuals.

approach for video-based multi-person 3D HPE, which achieved impressive results in [27, 232].

- **Bottom-up approaches:** First, generate joint positions and depth maps for all body joints. They then assign body parts to each individual based on the root depth and relative depth of the body component [75, 165]. How to categorize human body joints is a fundamental difficulty for these techniques. Methods at a lower level exploit the common latent space between two distinct modalities.

2.3.2 Evaluation Metrics And Datasets

Skeleton Scaling

Since the skeletons of CMU and Radical are not similar; the Procrustes analysis was used to determine the scale [90], rotation, and translation. Given correspondences of points $A_j \in \mathbb{R}^3$ and $B_j \in \mathbb{R}^3$ of the joint j find scaling, rotation, and translation transformation, called similitude transformation, that satisfies:

$$A_{\mathrm{i}} = sRB_{\mathrm{i}} + T \tag{2.13}$$

for $R \in SO(3)$, $T \in \mathbb{R}$, and $s \in \mathbb{R}^+$.

Evaluation Metrics

The first is the mean per-joint position error (MPJPE [114]) between the ground-truth 3D pose and the predicted 3D pose, which is calculated using the equation (2): Then, we calculate the mean error across all poses and actions in the dataset.

$$MPJPE = \frac{1}{m}\sum_{i=1}^{m} \|p_i^3 - \bar{p_i^3}\|_2 \tag{2.14}$$

For a given skeleton comprising m joints, p_i^3 denotes the actual 3D pose of $jointi$, whereas $\bar{p_i^3}$ signifies the predicted 3D pose of the same joint.

The second metric is the Percentage of Correct key points for 3D Pose Estimation (PCK3D) [152], a 3D version of the PCK utilized for 2D pose estimation [220]. If the estimated joint location is within a reasonable distance of the ground-truth joint, it is considered to be accurately estimated. Then, the proportion of accurately calculated joints is computed. As in earlier research, the neighborhood threshold is chosen at 150mm [152], corresponding to about half the head size. This statistic is more expressive and robust than MPJPE, highlighting joint mispredictions more clearly.

2.4 Virtual Reality (VR)

The term VR designates a reality that is "virtual", i.e. "created by computer technology and appearing to exist but not existing in the physical world". The genesis of VR can be traced back to the latter half of the 20th century, denoting a digitally fabricated experience that appears real yet lacks a physical presence. The late 1980s saw the popularization of the term, although the concept of emulating environments predates this era, with early instances like Morton Heilig's Sensorama simulator in 1961 [103], which simulated a motorcycle journey through multi-sensory stimulation. VR technology has since undergone significant development. Plotted on one extreme of Milgram's Reality-Virtuality Continuum [123], VR stands in contrast to the real environment, offering a completely simulated experience. This spectrum also includes mixed reality, which blends actual and virtual elements. Recent discourse [204] has revisited this continuum, critiquing its visual-centric view and suggesting that complete VR is an aspiration not yet realized, with most contemporary VR systems being variants of mixed reality (refer to Figure 2.6).

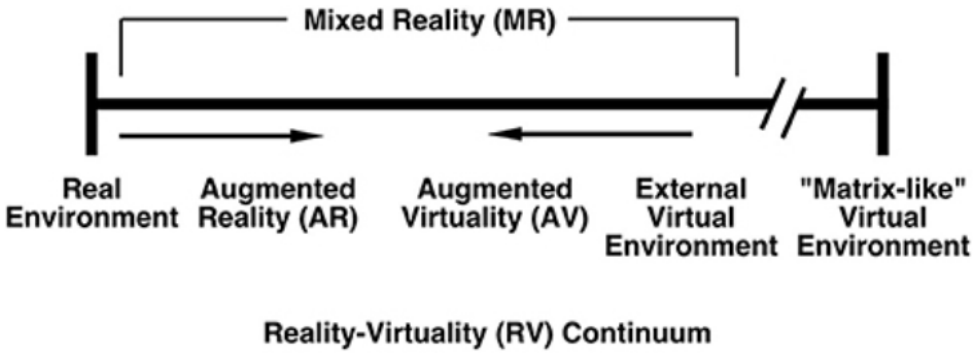


Fig. 2.6 Revisited Milgram's Reality-Virtuality Continuum by [204].

Technical definition of VR

Virtual reality is a scientific and technical field that uses computer science and behavioral interfaces to simulate the behavior of 3D entities in a virtual world, which interacts in real-time with each other and with one or more users in pseudo-natural immersion through sensory-motor channels [82].

Different interpretations of VR have been suggested, with [214] defining it as an environment, real or synthetic, where one can experience telepresence. Another more technical definition was introduced in 2003.

VR engulfs users in an entirely contrived universe, modulated in every aspect by computers, offering control over what is seen, heard, felt, and even smelled, thereby unlocking boundless prospects. These can range from recreating the tangible world for safe, tailored experiences to conjuring fantastical realms that ignite the imagination. To achieve such immersion, various devices are utilized, with a particular emphasis in this thesis on Head-Mounted Displays (HMDs). Regarded as the pioneer, Ivan Sutherland's team developed the earliest HMD in 1968 [216]. Contemporary HMDs have evolved to be more lightweight and even wireless, providing users with the freedom to move unencumbered. However, alternative VR equipment exists, like the Cave Automatic Virtual Environment (CAVE), where users are enveloped within a room-sized cube with projections of the virtual environment on its walls [55].

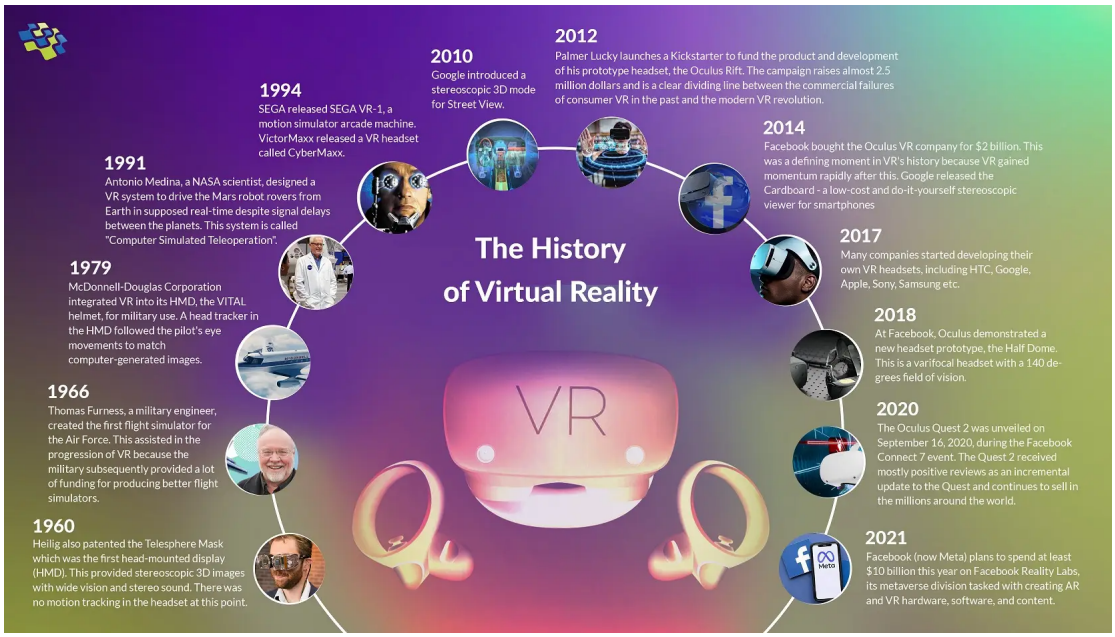


Fig. 2.7 The History of VR.

2.4.1 The Evolution of VR Technology

The historical progression of VR, as shown in Figure 2.7, from its inception to its current state reveals a trajectory marked by significant technological breakthroughs and increasing integration into various sectors [28]:

- **The Pioneering Era (1960-1979):** The journey began in 1960 when Morton Heilig patented the Telesphere Mask, the first head-mounted display (HMD), providing stereoscopic 3D images with expansive vision and stereo sound. Although it lacked motion tracking, it laid the foundation for immersive visual experiences. The technological advancement continued in 1966 when Thomas Furness, a military engineer, developed the first flight simulator for the Air Force, significantly contributing to the VR field due to substantial military funding for enhanced simulators. In 1979, McDonnell-Douglas Corporation's integration of VR into the VITAL helmet for military applications introduced a head tracker that followed the pilot's eye movements to align with computer-generated imagery.

- **Technological Breakthroughs (1991-1994):** A leap forward occurred in 1991 when NASA scientist Antonio Medina designed a VR system to remotely operate Mars rovers, termed "Computer Simulated Teleoperation," despite interplanetary signal delays. In 1994, the commercial space saw SEGA releasing the SEGA VR-1, an

arcade machine with motion simulation capabilities, and VictorMaxx launching the CyberMaxx VR headset.

- **Mainstream Emergence (2010-2014):** The turn of the decade in 2010 witnessed Google introducing a stereoscopic 3D mode for Street View, pushing VR towards mainstream applications. This momentum was catalyzed in 2012 when Palmer Luckey's Kickstarter campaign for the Oculus Rift signified a turning point, raising substantial funds and bridging past commercial VR failures with the modern VR revolution. The acquisition of Oculus by Facebook for $2 billion in 2014 was another pivotal moment, as it underscored the growing momentum of VR technology and was followed by the release of Google Cardboard, a low-cost VR viewer for smartphones.

- **Rapid Expansion and Diversification (2017-2021):** By 2017, an array of companies, including HTC, Google, Apple, Sony, and Samsung, had commenced the development of their proprietary VR headsets, indicating the technology's widespread adoption. In 2018, Oculus showcased the Half Dome prototype, a varifocal headset with an impressive 140-degree field of vision. The Oculus Quest 2, launched in 2020, enjoyed critical acclaim and strong sales globally, reflecting VR's deepening penetration into consumer markets. In 2021, Meta (formerly Facebook) announced significant investment plans for its Reality Labs division, emphasizing its commitment to advancing AR and VR technologies.

- **Current Innovations (2023-2024):** The unveiling of Meta's Quest 3 in 2023 continued the trajectory of VR's advancement, offering enhanced capabilities over its predecessor. In a groundbreaking move on January 8, 2024, Apple announced the upcoming release of the Apple Vision Pro, set to debut on February 2, 2024. This spatial computer, expected to ship in the tens of thousands, promises to integrate digital content with the physical world, navigable through eye movements, hand gestures, and voice commands.

2.4.2 Interactivity and Enhanced Capabilities in Virtual Environments

Within the virtual domain, users can engage with the Virtual Environment (VE) by performing actions that prompt responsive feedback from the system. The range of potential user actions in a VE mirrors real-world actions such as walking, jumping, or manipulating objects through grabbing or pinching. Efforts to replicate real-world knowledge in virtual interactions have been organized under the Reality-Based Interaction framework [Jacob et al. 2008], segmented into four knowledge realms: Naive Physics, Body Awareness & Skills,

Environment Awareness & Skills, and Social Awareness & Skills. Techniques informed by this framework represent a new wave of interactive modalities, succeeding in the era of desktop interfaces. These interaction techniques are noted for their high fidelity to real-world interactions [150], effectively emulating everyday human activities. Yet, VR transcends physical limitations, empowering users with 'superpowers'—the ability to interact with the environment in ways that defy real-world physics, such as reaching distant objects or phasing through walls. These enhanced capabilities can either substitute for or augment realistic interactions, introducing novel functionalities within the VE.

The process of interaction in a VE is encapsulated by the action-perception loop. Input from users can be diverse, encompassing motion, voice, or button press, and is relayed to the application via input devices. The VE processes this input and provides feedback through output devices. Feedback can directly correspond with the action, as seen with one-to-one mapping or isomorphic interactions, where a user's virtual representation mimics their actual movements. However, feedback may also result from a transformation of input, yielding a different or exaggerated response—referred to as anisomorphic interaction. The Control-Display (CD) gain describes the modification level applied to user input to determine the virtual representation's position. This concept has been explored both in two-dimensional [41] and three-dimensional spaces [15]. The CD gain can be static or dynamic, changing based on certain conditions. A CD gain below one indicates that the virtual movement is less extensive than the actual movement, whereas a gain above one amplifies the user's motion.

Following input interpretation, the VE issues feedback to the user. This feedback can target any of the five senses—visual, auditory, haptic, and more infrequently, olfactory and gustatory—with the first three being the most common. Feedback can either affirm the result of an action or simply convey the application's current status. Efficient interaction with the VE is hampered without such feedback [238].

Interactions within the VE are varied. A foundational classification by [32] sorts interactions into four categories: navigation, selection, manipulation, and system input. Navigation involves the user's movement translating into camera movement. Selection and manipulation enable users to pick and adjust objects, such as moving, rotating, or scaling them. System input involves changing application settings, navigating menus, or entering text. This section will delve into past studies on selection, manipulation, and navigation, while system input will not be examined in depth, given that its techniques overlap significantly with those used for selection and manipulation.

2.4.3 Selection and Manipulation Techniques in Virtual Environments

The processes of selection and manipulation within virtual environments may be treated as distinct activities or as a unified interaction paradigm. Selection is the process through which a user identifies and chooses an object for interaction. This typically involves two steps: initially, the user positions a selection tool, like a virtual hand or a ray, over the intended target; subsequently, the selection is confirmed, often by clicking a button. Various selection tools and methodologies exist, which can be categorized by criteria such as the number of degrees of freedom they allow [15].

Manipulation encompasses the array of modifications that can be applied to a chosen virtual object. This process is traditionally broken down into four fundamental tasks: positioning, rotating, scaling, and, in some frameworks, selection as an integral part of manipulation [33]. To address the intricacies of 3D manipulation, a multitude of techniques have been developed [153].

Both selection and manipulation may employ anisomorphic interactions. A Control-Display (CD) gain below one can be used to increase control precision during tasks like detailed positioning [81]. Conversely, a CD gain above one may be employed to enhance the speed of manipulation when fine accuracy is less essential or to extend the user's reach to distant objects [179]. Altering the CD gain can also facilitate haptic feedback by guiding the user towards a physical object. This is achieved by introducing an offset to the virtual representation, leading users to physically touch a tangible item within the virtual space [124, 18].

2.4.4 Navigational Dynamics in Virtual Environments

In virtual environments, locomotion techniques are instrumental in enabling users to move and orient themselves. While the term 'navigation' often pertains to tasks that include wayfinding, locomotion is a broader concept that has garnered considerable attention in VR research [9]. A variety of methods have been developed to address the constraints of physical space, such as walking-in-place [209] or teleportation [34]. These methods range from virtual movements, like virtual steering, to physical actions that do not fully replicate real-world movements, such as walking-in-place, leading to a lower fidelity of interaction [151].

Similar to manipulation, navigation can also feature anisomorphic elements. For example, a Control-Display (CD) gain larger than one enables users to cover greater distances more rapidly [112] or to increase rotational movements to aid in reorienting the user towards a specific direction [213].

The classification of locomotion techniques has evolved over time. Initial taxonomies focused on implementation details, such as velocity and target selection [Bowman et al. 1997]. As expertise in navigation algorithms improved, classifications began to emphasize the underlying paradigms. [17] differentiated between physical and virtual navigation. More recent classifications have divided locomotion techniques into four main groups: room-scale-based, motion-based, teleport-based, and controller-based [30]. The current trend in research places a greater emphasis on user perception and experience of locomotion within virtual spaces [10].

2.4.5 Immersion and Presence in VR

VR offers a multitude of user experience dimensions, including enjoyment and comfort. This section delves into two pivotal aspects frequently discussed in VR scholarship: immersion and presence. Immersion quantifies the degree to which a virtual environment provides realistic sensory experiences [206], shaped by the combination of hardware output devices and software capabilities. A highly immersive application will typically feature elaborate visual graphics, nuanced audio, and tactile feedback, enhancing the user's sensory engagement with the virtual world. This technological envelopment can engender in users a strong illusion of being transported into an alternate reality, an experience known as presence.

The notion of presence, or the feeling of 'being there,' has been the subject of numerous studies identifying the elements that heighten this sensation. One such element is the embodiment of a virtual avatar, which has been shown to amplify the sense of presence [208]. Unlike immersion, which is more technology-driven, presence is a subjective psychological state that manifests within the user's mind and can be influenced by individual mental constructs. It is a response to immersion [Slater 2003], eliciting naturalistic responses to the virtual environment. [207] proposes breaking down the concept of presence into two components: Plausibility (Psi) and Place Illusion (PI). Place Illusion is aligned with the traditional understanding of presence, fostering a sensation akin to being in an actual, physical location. Plausibility, on the other hand, pertains to the perceived realism of the virtual scenario, with higher plausibility leading users to believe that events within the VR are truly happening.

In the context of this thesis, the focus is on the enveloping experience of VR, a synthetic realm where individuals can not only feel genuinely situated but also engage interactively. A crucial factor in enhancing presence and enabling interaction is the introduction of a virtual body, which provides users with a new dimension of self within the virtual domain.

2.5 Conclusion

This chapter has provided a thorough examination of the foundational elements and technological progress within the realms of 3D modeling, human pose estimation, and VR. Through exploring the intricacies and evolution of these fields, we have gained insights into their significant impact on various sectors, particularly in enhancing digital documentation, interactive media, and immersive experiences. The chapter underscores the interconnectivity of these technologies and their collective potential to transform traditional approaches, whether in cultural preservation, animation, or immersive simulations. As we advance, it is imperative to continue exploring these technologies' boundaries, addressing existing challenges, and leveraging their synergies to pave the way for innovative applications and deeper user engagement in both virtual and real worlds. The insights gleaned from this background form a solid foundation for understanding the complexities and potential of these technologies as they continue to evolve and reshape our digital and physical landscapes.

Chapter 3

Literature Review

3.1 Introduction

This chapter provides a comprehensive review of existing literature in the domains of terrestrial photogrammetry, 3D human pose estimation, and the use of virtual reality in cultural heritage preservation. The review aims to outline the current state of research, identify the technological advancements, and highlight the gaps and opportunities within these fields. Terrestrial photogrammetry has emerged as a significant tool for documenting and preserving cultural sites, offering a non-invasive means to capture and model historical artifacts accurately. Meanwhile, 3D human pose estimation has become crucial in understanding and re-enacting cultural and historical events, providing insights into past human activities. Finally, the incorporation of virtual reality technology in preserving intangible cultural heritage represents a groundbreaking approach, offering immersive experiences that traditional methods cannot. This literature review sets the groundwork for understanding the interplay between technology and cultural heritage preservation, paving the way for future research and application.

3.2 Terrestrial Photogrammetry

Today, the 3D modeling of cultural and archaeological sites has become a powerful tool. It changes the traditional known workflow and offers many benefits that were widely investigated [221]. It is a significant solution, for many purposes such as preservation, documentation, and keeping a digital record of culture to support maintenance and restoration operations in cases of damage or loss (earthquakes, fires, wars) [131, 215, 68]. The terrestrial laser scanner is the optimal standard for this purpose due to its highly accurate results [14, 125, 201]

and its automated and practical process. It can also be used under most weather and lighting conditions [180]. However, it has some limitations when it comes to complex cultural and historical objects [160], and it remains a costly technique. This could prevent its use in some developing countries. Therefore, the proposal of low-cost and accurate alternative ways is necessary to support and enhance culture preservation in such countries. The digital close-range photogrammetry technique has been known recently as an accessible promising alternative for laser scanners in several fields. It is recognized as a major tool in cultural heritage and an added value to sustainability in the interdisciplinary field of archaeometry for conservation-restoration purposes [135]. It can have been frequently used in conservation, restoration, documentation [172, 80], archaeology [130, 100, 8, 230, 116], construction [136, 3, 247], and extended realities [109, 181, 161]. This success is due to the ease of processing and high accessibility from most archaeological and researchers having limited budgets around the world. However, most of the close-range photogrammetry studies use an expensive digital single-lens reflex camera (DSLR) to take photos of the sites. Therefore, the optimization of cost-effectiveness is necessary. Numerous studies have investigated low-cost image-based digital close-range photogrammetry techniques for cultural heritage. The accuracy of many datasets was evaluated with different complexity degrees using the control points and scale bars as a bundle block adjustment, in order to derive the potential of low-cost automated image-based 3D modeling. The results indicated good accuracy when ground control points were used [186]. Low-cost image-based 3D modeling methods for cultural heritage and archaeological for nontechnical users were also investigated, using cheap or free instruments and software, with variable geometric accuracy level [31, 121].

Several studies have also evaluated the feasibility of low-cost image-based modeling quality of cultural heritage in comparison with a terrestrial laser scanner and/or CPs. A masonry arch bridge was 3D modeled using photogrammetry and dense cloud as low-cost methods. Although the results were evaluated through several parameters, the number of used images and the processing time were very high [12]. The profits of low-cost image-based reconstruction of the historical and cultural heritage of Nepal devastated in an earthquake were investigated, using phone and DSLR cameras. The obtained models were uncontrolled, i.e., the models were not in the real scale [67]. Otherwise, the parameters that impact the accuracy of the obtained results according to the used software, the number of photos, and ground control points using smartphone cameras were analyzed by [11], yet the achieved results were not always steady. The potential of a cheap camera using two solutions for the bundle block adjustment was exploited by [71]. The first solution uses ground control points, whereas the other one, the controlling model employs scale bars. The obtained 3D models were compared to a section of CPs and exhibited high quality, mainly when CPs were used.

Also, the effect of the number of CPs on model quality was analyzed. Nevertheless, the case of study was perfect, i.e., the building had no historical value and presented no architectural challenges.

3.3 3D Human Pose Estimation

3.3.1 Datasets for 3D HPE

Obtaining precise 3D labeling for 3D human pose estimation datasets is a difficult endeavor that necessitates the use of motion capture techniques such as MoCap and wearable IMUs. Since the 3D HPE deep learning-based needs large datasets to train, validate, and test their models, several 3D posture datasets are created due to this need.

- **HumanEva Dataset [203]:** includes seven calibrated video sequences (4 grayscale and three colors) with ground truth 3D annotation taken by a ViconPeak commercial MoCap system. The database comprises four scenarios executing six common actions in a 3m × 2m area: walking, jogging, pointing, throwing and catching a ball, boxing, and combination.

- **Human3.6M [114]:** One of the most commonly used datasets for indoor 3D human pose estimation from monocular images and videos. The dataset features 11 professional actors (six males and five females) performing 17 actions (such as smoking, taking photos, and talking on the phone) in a laboratory environment captured from four different perspectives.

- **The CMU Graphics Lab Motion Capture Database (CMU) [2]:** CMU is one of the most publicly large databases of motion capture data. Numerous researchers within the scientific world have utilized it to develop previous models of human motion. However, the dataset is poorly synced and contains some films unsuitable for HPE due to multiple actors in each scene. The database comprises more than 100 scenarios executing several actions in a 3m × 8m area.

3.3.2 Advancements and Challenges

- **Advanced Evaluation Metrics:** Beyond traditional metrics such as MPJPE and PCK, the community has developed more sophisticated measures to capture the nuances of 3D pose estimation accuracy. The introduction of Procrustes Aligned Mean Per Joint Position Error (P-MPJPE) and Normalized MPJPE (N-MPJPE) offers a way to

evaluate pose estimations with scale and rotation normalization, providing a more refined analysis of model performance. [233] have shown how these metrics can offer a clearer picture of an algorithm's capability to capture human pose dynamics accurately.

- **In-the-Wild Evaluation:** As 3D human pose estimation models are increasingly applied to real-world scenarios, the need for in-the-wild evaluation has grown. Datasets like 3D Poses in the Wild (3DPW) challenge models with complex outdoor scenes, diverse activities, and varying lighting conditions. [228] highlighted the importance of such datasets in testing the robustness and generalizability of pose estimation models.

- **Synthetic Data for Evaluation:** The use of synthetic data has emerged as a valuable tool for evaluating 3D pose estimation models, addressing the scarcity of labeled real-world data. Synthetic datasets allow for the generation of large volumes of annotated images with controlled variability in poses, backgrounds, and lighting conditions. [223] demonstrated how synthetic data could be used to augment real-world datasets, enhancing model training and evaluation processes.

- **Inter-Dataset Variability and Standardization:** The variability across datasets poses a significant challenge for model evaluation, as differences in annotation protocols, camera perspectives, and subject diversity can lead to inconsistent performance metrics. Efforts towards standardizing dataset protocols, as discussed in [203], aim to mitigate these issues by providing a common framework for dataset creation and evaluation.

- **Evaluation through Auxiliary Tasks:** Recent research has explored the evaluation of 3D human pose estimation models through auxiliary tasks, such as action recognition or human-object interaction. This approach, as explored by [244], assesses the utility of pose estimations in downstream applications, providing insights into the practical effectiveness of estimation models.

- **Ethical Considerations in Evaluation:** As the field progresses, ethical considerations regarding data diversity, privacy, and bias in evaluation datasets have come to the forefront. Ensuring that datasets represent a wide range of body types, ethnicities, and environments is crucial for the fair and inclusive evaluation of pose estimation models. Studies like [85] highlight the need for transparency and diversity in dataset collection and evaluation practices.

- **Future Directions in Evaluation:** As we advance, the assessment of 3D human pose estimation is set to incorporate groundbreaking technologies such as VR and AR, providing immersive environments for testing. These innovative platforms can

significantly enhance the evaluation process by enabling real-time, interactive scenarios that mirror real-world applications, aligning with insights from [16]. Furthermore, the development and adoption of unsupervised and self-supervised learning metrics are expected to become increasingly important. These metrics will be crucial for evaluating model performance, particularly in scenarios where large-scale annotated datasets are scarce or unavailable, paving the way for more generalized and adaptable pose estimation solutions.

3.4 Immersive Digital Technologies in Cultural Heritage

Cultural heritage sites and artifacts gain significant added value through digital media enrichment. However, many art curators believe that technology relegates art to the background [39]. This perception seems to stem from cultural or generational sources. Firstly, there is widespread skepticism among those uncomfortable with technology about the benefits of mobile technologies. Secondly, and more importantly, issues arise with the way technology is used, as multimedia applications often focus on the show and glamour of innovation rather than solving specific problems with digital means [175].

Nonetheless, there is general agreement that visual CH tools suitable for users unskilled in multimedia technologies are crucial for CH dissemination [51], especially for younger participants. Museum installations that do not introduce new technologies are often regarded as less interesting and attract fewer visitors [86]. Learning experiences in museums relying solely on labels and descriptions may be informative but lack interactivity [236]. Creating an intelligent environment that dynamically adapts to human presence and supports mobile technology makes visits more appealing, opening new avenues in both tangible and intangible CH [146].

Tangible CH refers to physical artifacts of a society, including artistic creations, built heritage, and other physical products imbued with cultural significance. Conversely, Intangible Cultural Heritage encompasses non-physical practices, representations, expressions, knowledge, and skills recognized as vital components of Cultural Heritage [6].

Immersive reality systems have proven to be viable solutions, allowing navigation, interaction, and discovery in various settings for different purposes. In archaeology, for instance, the dissemination of heritage often involves communicating goods that are seriously damaged or lost. Technologies can function like X-rays to reveal what is concealed underground or augment environments with virtual reconstructions of lost heritage [52].

These considerations highlight the need for classifying Cultural Heritage application areas to understand where AR, VR, and MR can offer successful solutions. We classify

the purposes of immersive reality in CH as education, exhibition enhancement, exploration, reconstruction, and virtual museums.

- Education: Enabling users to learn the historical aspects of tangible and intangible CH.
- Exhibition enhancement: Improving the visitor experience at physical museums and heritage sites, typically through tour guidance.
- Exploration: Supporting users in visualizing and exploring historical and current views of CHs to discover, interpret, and acquire new insights and knowledge.
- Reconstruction: Enabling users to visualize and interact with reconstructed historical views of tangible and intangible CHs. Unlike exploration, it does not solely target experts and the visualization and interaction do not necessarily extend to discovering new insights.
- Virtual museums: Simulating and presenting tangible and intangible CHs in digital museum form to the public.

Some categories overlap. For instance, a reconstruction application might also allow users to learn the history of the reconstructed CH, combining education and reconstruction. Similarly, a virtual museum exhibit might enhance a physical museum, combined with exhibition enhancement. Despite these overlaps, the central objective of a given application generally favors a particular purpose, which we use as the basis for determining the primary category.

The relationship between categories is depicted in the Venn diagram [102] in Figure 3.1, showing the technologies that dominate overlapping categories and a statistical overview of developed systems in the CH domain.

Considering the objectives of these application areas, we next discuss technological suitability and technical requirements. Figure 3.2 summarizes these requirements in tabular form. We also categorize several augmented-, virtual-, and mixed-reality applications based on these themes. See Tables 3, 5, and 4.

1. **Education:** Every research area dealing with the dissemination and diffusion of CH must consider education. However, our focus here is on tools and applications where learning is the primary aim [19]. For example, museum designers have recently leveraged immersive realities' capacity for spatial and temporal representation, narrative, interactivity, real-time personalized scaffolds, and collaboration to create meaningful learning experiences in medicine and human biology [149]. By enhancing a sense

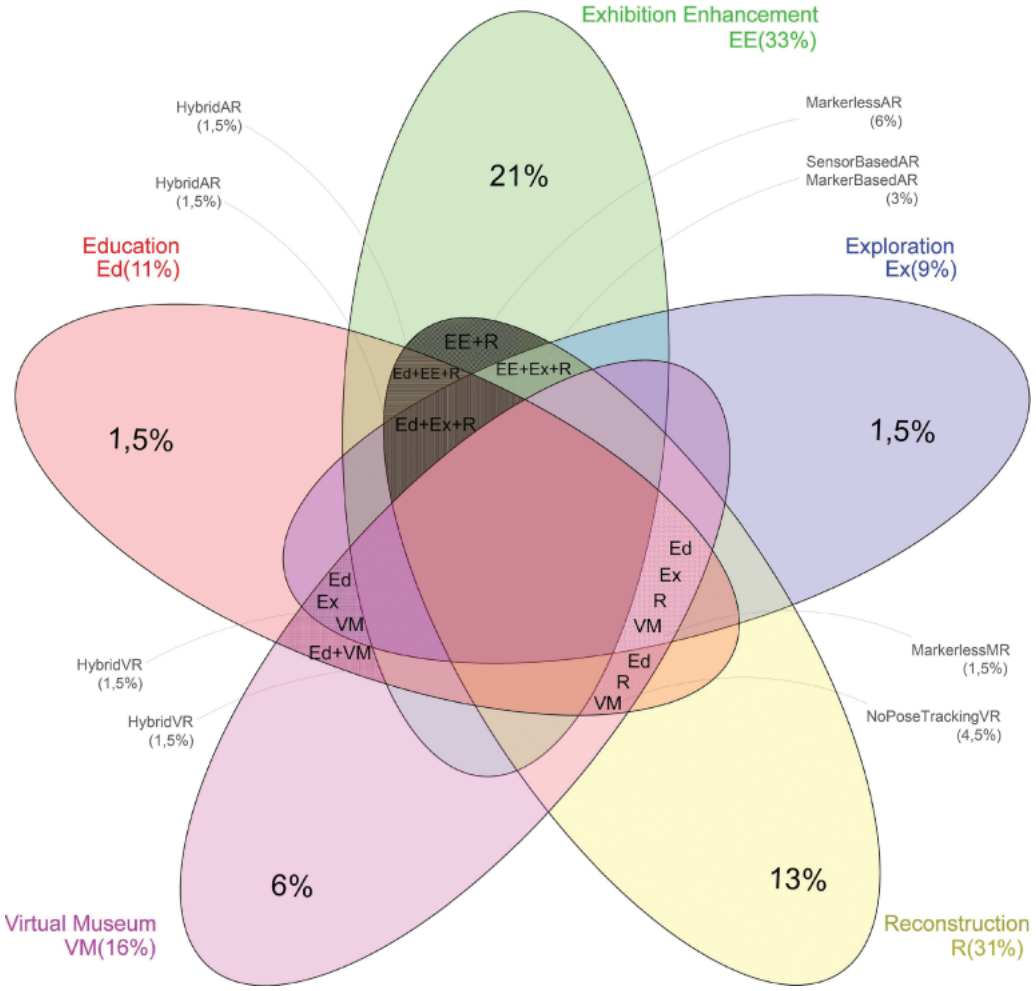


Fig. 3.1 A Venn diagram illustrating the relationships between application categories. The assignment of applications to overlapping categories is partially determined by the use of shared technologies (e.g., Markerless Augmented Reality for the overlap between Exhibition Enhancement and Reconstruction) [25].

of place, such as improving visits or way-finding in a virtual environment, learning activities can be significantly improved [44]. [84] discuss how playful learning can cross boundaries among schools, museums, and science centers by involving participants in extended episodes of digital interaction with exhibitions. This approach significantly improves learning outcomes, increasing students' curiosity, willingness to share their experiences, eagerness to use new technologies, and knowledge acquisition. Similarly, [113] considers various interventions and studies related to new technologies and new scientific languages based on the learning objective. The idea is to enhance the usability of the MAUS Museum through an AR application and Virtual Reality projections related to the natural sciences (Plankton 3D and Tarbosaurus 3D). Evidence

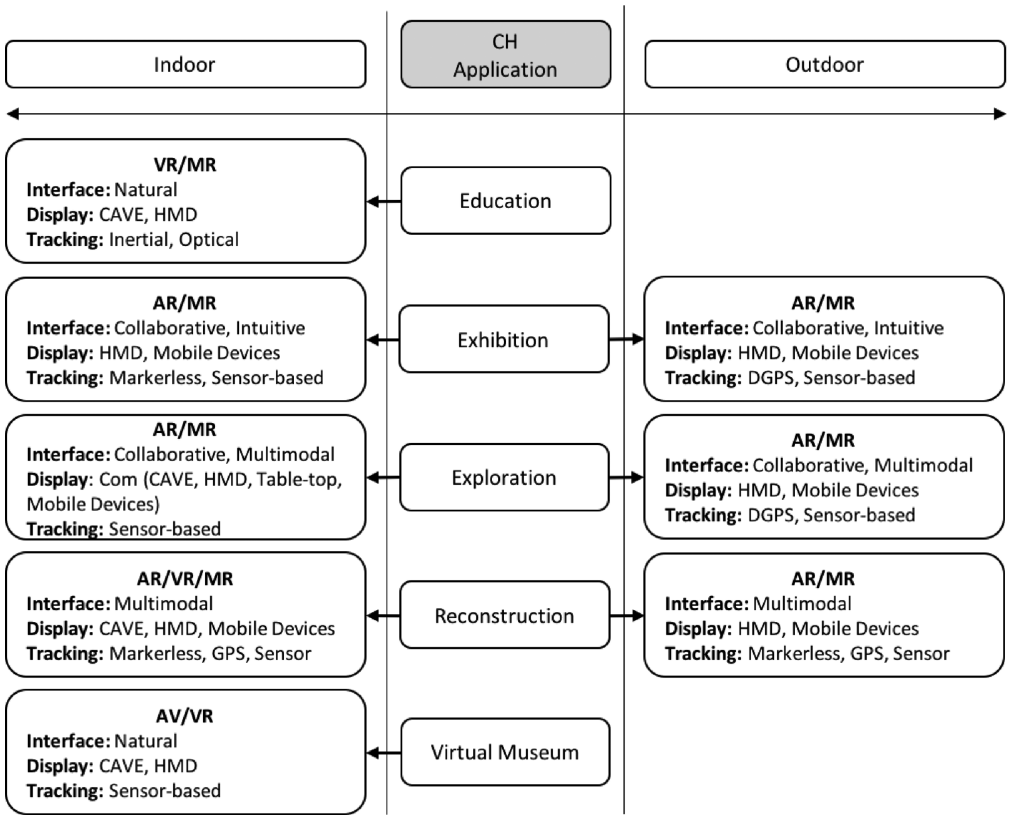


Fig. 3.2 Considering the identified application areas of cultural heritage, the diagram above illustrates the technical requirements for augmented, virtual, and mixed reality systems in both indoor and outdoor settings[25].

suggests that educating users on the historical aspects of both tangible and intangible CHs requires presenting the content in an entertaining environment. To this end, the system should be immersive and interactive. Tangible educational CH needs users' active interaction with the displayed content. Interfaces must be as natural and intuitive as possible. Users' inexperience with such applications should not constrain delivering the historical aspects as intended. A user's age, background, and domain knowledge may differ, and the system should adapt accordingly. Intangible CHs, however, do not need adaptive interfaces to the same extent, as ready-made audio-visual content is presented, and user interaction is often limited to playing and pausing the content. For tangible CH, HMDs and CAVEs with high-resolution displays and realistic rendering capabilities can achieve the required immersion. Tracking is mandatory to enhance immersion and interaction. A combination of inertial and IR sensors can provide the user's pose. Interface-related tracking is better achieved through natural interaction such as gesture, gaze, and speech sensors. Tracking is not required if intangible CHs are presented, but it can enhance the experience. Both VR and MR can be used to

achieve educational support in a fully immersive environment. AR may not be suitable as it overlays virtual and real-world views, while the focus of educational applications tends to be historical.

2. **Exhibition enhancement:** Enhancing a visitor's experience can take place indoors, outdoors, or both, depending on the location of CH assets. In all cases, a virtual element, such as a description, guide map, or virtual human character, is superimposed over the users' current view of the real world. The number and quality of applied research articles in this field are high, as AR can provide various solutions to help museums fulfill their role and goals [49]. Regardless of the type of installation, evidence shows that visitor interest grows when immersive solutions are adopted [43]. In the user study conducted during the ARCO project [217], this trend was evident. In many cases, an immersive reality approach enables new media and storytelling, representing the major highlight of a user's experience [170]. The work of Liestøl [133] shows that such practices can damage the historical value of CH assets. Regarding outdoor systems, a combination of location-based, sensor-based, and markerless methods can be used to achieve tracking. If users can approach the CH asset, a markerless method is suitable; otherwise, long-range optical sensors and GPS localization are more appropriate. Visitors often attend museums and CH sites in groups. Hence, the interaction interface should be collaborative and intuitive so users can experience the visit from their perspective while collaborating with co-located visitors. This can extend to remote collaboration between spatially distributed visitors.

3. **Exploration:** Exploration-based applications focus on the historical and current aspects of tangible archaeological CHs, allowing users to discover, explore, visualize, and manipulate the content, leading to knowledge creation and new insights. Users of such applications are assumed to have expertise in the domain; therefore, the system can assume prior knowledge of domain-specific visualization. Exploration-focused applications need hybrid tracking, a combination of complementary displays, collaborative and multimodal interfaces, and distributed and immersive environments. Single-technology tracking is insufficient due to its lack of accuracy. As users are assumed to be experts, the tracking method can focus on accuracy over user experience without compromising user comfort. A combination of sensor-based methods can fit indoor environments, while location and sensor-based methods can achieve outdoor tracking. If exploration tasks are at distributed locations, pose readings from all locations must be synchronized; otherwise, users cannot collaborate seamlessly. A suitable combination of displays includes HMDs and table-top projectors for indoor settings

and HMDs for outdoor settings. Table-top projectors and CAVEs can accommodate multiple users at a time, and HMDs can display user-centered perspectives. Interactions with the displayed content rely heavily on the accuracy of interface-related tracking methods. For this, sensor-based input devices and natural interaction mode sensors should be used in combination. MR and AR are best for exploration purposes, as users can see both real-world and virtual views. This feature is especially invaluable in archaeological settings. The possibility of enriching reality with computer-generated information, and providing innovative information access at CH sites, has been noted in recent research. Verykokou et al. [227] visualize a part of the Middle Stoa in the Ancient Agora of Athens (see Figure 3.3a), allowing users to see what this building looked like in ancient times as its three-dimensional model is displayed on their device's camera view, projected onto modern-day ruins. Related examples of AR for exploration can be found in the work of Etxeberria et al. [73], Pierdicca et al. [174], Stanco et al. [212], Deliyiannis and Papaioannou [63], and Empler [72]. This brief analysis highlights three main points: First, AR applications in this domain are restricted to sensor-based AR, and interaction with virtual archaeological content in museums using AR has not spread (VR provides more possibilities in this regard). Second, AR is mainly used to visualize lost or posited artifacts. Third, for archaeology, the use of geomatics applications is unavoidable [178]. Virtual reconstructions (see Figure 3.3b) must rely on accurate data sources [173, 182]. Interaction is key to making such experiences more attractive for users and involving them actively in the exploration process, as shown in the well-designed mobile interaction solutions of Wiley and Schulze [237] and Kang [118]. Since the mid-2000s, archaeology has benefited from the widespread availability of digital 3D models [54], allowing developers to represent difficult-to-reach environments, such as underwater conditions [101]. The challenge, discussed further in Section 5, is twofold: describing a known workflow that moves from data acquisition to visualization in an immersive environment and making these data portable and suitable for different devices or platforms. These technologies can better accommodate collaborative and multimodal interaction. VR systems cannot meet these requirements to the same extent. However, if collaboration is not needed, HMD-based VR can suffice.

4. **Reconstruction:** Reconstruction applications display reconstructed views of tangible and intangible CHs, allowing users to visualize CH assets that existed only in the past or that partially exist. Reconstructed assets can be presented as tangible, intangible, or a blend of both. AR and MR are best suited to tangible and blended forms because they can superimpose reconstructed views over their historical location. Additional



Fig. 3.3 Examples of AR applications for archaeological purposes include visualizing ancient architecture in its original location using location-based AR technology.

information beyond the virtual reconstruction itself can also be overlaid [196]. To ensure the preservation of artifacts, such as statues or paintings, they must be analyzed to diagnose physical frailties that could result in permanent damage. Advancements in digital imaging techniques and computer-aided analysis aid such diagnoses, though the ability to work directly with artifacts in the field remains limited. Several examples of different kinds of diagnosis are reported by Colizzi et al. [53]. Of particular interest is the work of Girbacia et al. [95] on a workflow for the restoration of religious heritage, starting with the reconstruction of statues and extending to their in-place geo-located visualization in AR. Vanoni et al. describe ARtifact [222], a tablet-based augmented reality system that enables on-site visual analysis (see Figure 3.4a). Their idea is to use overlaid layers to represent images acquired from different data sources. Another tool, the "revealing flashlight" [192], distinguishes details obscured by aging effects by projecting an expressive 3D visualization that highlights features based on an analysis of previously acquired geometry at multiple scales (see Figure 3.4b). The novelty mainly lies in the interaction, which is based on gestures. VR, on the other hand, is suitable for intangible reconstruction and visualizing tangible assets in indoor environments, as it does not rely on displaying reconstructed views over their historical location. A review of such applications with a focus on interaction and gamification is provided by Kateros et al. [119]. For AR and MR, positional tracking can be achieved using a combination of GPS, orientation sensors, and markerless tracking. VR requires tracking to correct perspective, achieved through orientation sensors attached to HMDs or stereo glasses. Users of reconstruction applications may range from domain experts and students to the general public, preferably with gamification [164**?**]. Therefore, the system should be inclusive of these groups' backgrounds, focusing on interaction and presentation aspects. As a result, such applications should have a multimodal interface



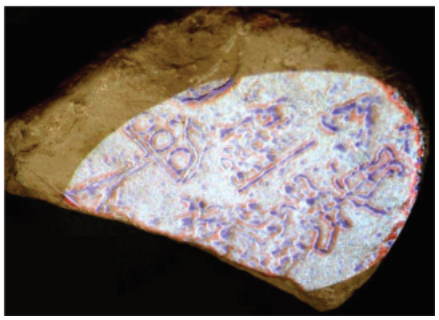


Fig. 3.4 Some examples of tools designed to facilitate restoration and enhance analysis. These applications reveal features that are not visible to the naked eye.

and immersive features. In general, see-through HMDs fit the requirements for AR and MR systems, and CAVEs suit VR.

5. **Virtual museums:** Virtual museums simulate physical museums and CH sites, including their tangible and intangible assets. Often, such assets are inaccessible and fragile, so the simulation must be very realistic and detailed to serve as a replica of artifacts, making it difficult for users to discern differences between the originals and their replicas. Such simulations enhance users' presence in virtual museums, creating the illusion of being physically present at an actual museum or CH site. This can be extended to represent users as virtual human characters inside the simulated environment, allowing co-located users to see each other. To achieve this, the chosen modeling method should blend pre-rendered scenes and virtual human characters in real time. These virtual characters should not be mere avatars but close approximations of actual users to simulate real-life interaction, considering their behavioral and physical properties. The system should be fully immersive, allowing interaction via gesture, gaze, speech, and movement. The enabling sensors for such interfaces should not remind users of their attachment to the real world, as this would interrupt their presence in the virtual museum. Sensors physically attached to users may require direct manipulation compared to remotely placed sensors, which may decrease presence. Interaction should create a perception of physical movement inside a digital environment. HMD-based AV and CAVE-based VR environments can achieve a fully immersive virtual museum. However, large-scale CAVEs are more appropriate, as they can accommodate multiple users, making virtual human characters unnecessary.

The following sections discuss the surveyed articles from the perspectives of enabling technology, systems, and purpose. Additionally, suggestions are made based on these observations as to which technologies are most suited to a given purpose. These suggestions

differ from those discussed above, which are based on the central objective of the identified application areas, whereas the suggestions below are based on the technologies adopted by the surveyed works.

3.4.1 AR Applications

Most AR applications aim at exhibition enhancement, followed by reconstruction and exploration. Table 3.1 shows the details of these applications. Hybrid tracking is a relatively common approach, with markerless, sensor-based, and marker-based methods used in that order. In terms of presentation devices, mobile displays make up the majority, followed by HMDs, a combination of diverse displays, desktop, custom-built, and SAR displays. Most applications use a tangible interface, followed by natural, collaborative, multimodal, and hybrid interfaces. Additionally, most of the surveyed applications target indoor conditions, with a few applications suitable for both indoor and outdoor environments. There are two notable differences in the environmental settings for AR applications.

Indoor Systems

Most indoor applications focus on exhibition enhancement. The tracking methods used are marker-based and markerless. A significant number of indoor applications use mobile devices for display, and tangible interaction interfaces are the most common.

Outdoor Systems

Most outdoor applications aim at reconstruction. Hybrid tracking is often adopted. Mobile devices and tangible interfaces are common choices for display and interface, respectively. Indoor applications are more suited to exhibition enhancement experiences since physical museums often use such applications for virtual tour guidance more than outdoor CHs. Outdoor applications are more suitable for reconstruction, as it allows overlaying reconstructed historical views over the real world. Moreover, a reconstruction theme is often applied to outdoor sites that have been demolished or worn away.

3.4.2 VR Applications

In our findings, most VR applications serve virtual museums, followed by education, reconstruction, and exploration purposes, in that order. Table 3.2 shows the details of these applications. Most do not use any tracking methods, as these applications are non-immersive or semi-immersive. The remaining applications use hybrid, electromagnetic, and optical

tracking methods. Presentation devices are mostly screens/projectors, followed by CAVEs and HMDs. Interfaces are mostly device-based, followed by sensor-based, multimodal, and hybrid interfaces. Applications tend to be semi-immersive. We discuss these applications from the perspective of their level of immersion, as their systems range across non-immersive, semi-immersive, and fully immersive environments.

Non-immersive

Non-immersive applications are primarily for education, virtual museums, and reconstruction. These applications do not use any tracking methods. They employ desktop screens for displaying virtual content and use device-based interfaces. Desktop screens and device-based interfaces seem sufficient for non-immersive experiences for education, virtual museum, and education themes. Pose tracking is not required for non-immersive systems.

Semi-immersive

Virtual museums and education are the primary areas for semi-immersive applications. However, a few applications serve reconstruction and exploration purposes. Electromagnetic, optical, and hybrid methods are used to track users' poses and interaction devices in a few applications, but most do not use any tracking. This is acceptable since users see pre-rendered virtual content and most often tracking is only required for interaction. Moreover, tracking may be unnecessary if a gamepad or mouse is used. Back-projected screens and 3D stereo displays are common choices for semi-immersive applications. Most applications use device-based interfaces, with a few using sensor-based and hybrid interfaces. Therefore, optical, electromagnetic, and hybrid tracking methods; back-projected screens and 3D stereo displays; and device-based, sensor-based, and hybrid interfaces are viable for semi-immersive systems intended for virtual museums, reconstruction, education, and exploration.

Fully Immersive

Virtual museums are the most frequent application area for fully immersive VR, with a few applications for education and exploration. Most applications use hybrid tracking, with a few systems employing electromagnetic and optical methods. Fully immersive experiences are achieved by CAVE and HMD displays. Device-based interfaces are widely adopted, though a few applications use multimodal and sensor-based interfaces. Fully immersive VR experiences supporting virtual museums, education, and exploration are best achieved by adopting electromagnetic, optical, and hybrid tracking methods, CAVE and HMD displays, and device-based, sensor-based, and multimodal interfaces.

3.4.3 VR in Intangible Culture Heritage

The utilization of VR technology for cultural heritage preservation, both tangible and intangible, has become a topic of growing interest within the research community in the past decade. VR's potential for restoration, analysis, and conservation of cultural landscapes and Intangible Cultural Heritage (ICH) is increasingly recognized.

For example, VR was used to assess the visual aesthetic effect of restored cultural landscapes in southwest China, finding that the restoration not only reduced visual fatigue but also enhanced the visual aesthetic experience [134]. In the context of ICH preservation, VR has been employed in the learning of traditional crafts, such as Dongyang bamboo weaving, effectively reducing the learning time required for subjects and showcasing its potential as an educational tool for ICH preservation [248].

Further research has investigated the use of VR in ICH High-Definition Digital Mobile Display Technology. One such study focused on Dai ceramics, demonstrating that VR can enhance the virtual visualization of ICH digital display technology by digitally and artistically processing ceramics while maintaining their authenticity [241]. VR platforms have also been developed based on Convolutional Neural Networks (CNNs) for the conservation of ICH, such as the boat-shaped houses of the Li nationality in Hainan, China, highlighting the advantages of digitizing ICH using artificial intelligence [64].

Other researchers have implemented cloud-based VR systems for preserving traditional art forms, such as Chinese glove puppetry, showcasing its effectiveness and ease of use [235]. An immersive VR learning experience was developed to teach prehistoric intangible cultural heritage to young history students, highlighting VR's potential as a medium for learning and disseminating traditional cultural practices [22]. Additionally, Serious Games have been employed to bring ancient stories to life, such as The Story of the Shipwrecked Sailor, while teaching players about hieroglyphs [177]. An ontology-based system encoded dance and music performances, providing a semantic navigation environment for Indian Classical Dance [144]. A conceptual framework was proposed for the design of virtual environments for cultural learning, emphasizing the importance of information design, information presentation, navigation mechanisms, and environment setting [111].

The preservation of ICH has also been addressed in diverse areas such as archaeology, history, folklore, and performing arts. An immersive VR application was developed for the Museum of Preclassical Civilizations of the Southern Murgia in Ostuni, allowing users to explore the Upper Paleolithic settlement and experience the burial ritual of Ostuni 1 [62]. Research into preserving the heritage of bridge diving at Stari Most in Mostar, Bosnia, used interactive digital storytelling with 360-degree videos in VR [200]. Another study focused on creating a multimedia platform to communicate the tradition of bronze "investment casting"

Table 3.1 The purposes and enabling technologies utilized by AR applications in cultural heritage.

Application	Purpose	Tracking	Display	Interface	Setting
Wojciechowski et al. [240]	Exhibition enhancement	Marker-based	Desktop screen	Hybrid (Tangible and Web-based)	Indoor
Miyashita et al. [155]	Exhibition enhancement	Hybrid (Markerless and IMU)	Mobile	Tangible	Indoor
Portalés et al. [178]	Reconstruction	Markerless	HMD	Natural (Movement-based)	Indoor
Zoellner et al. [252]	Reconstruction	Markerless	Combo (Mobile (UMPC) and MovableScreen)	Tangible	Indoor
Kim et al. [122]	Exhibition enhancement	Markerless	Laptop	Tangible	Indoor
Haydar et al. [101]	Reconstruction	Marker-based (optical)	HMD	Tangible	Indoor
Damala et al. [58]	Exhibition enhancement	No pose tracking, gaze, acoustic, and biosensor	Custom-built (See-through glass)	Natural (Sensor-based interface)	Indoor
Ridel et al. [192]	Exploration	Hybrid (electromagnetic and optical)	SAR	Tangible	Indoor
D'Agnano et al. [56]	Exhibition enhancement	No pose tracking (audio-based AR)	Mobile	Tangible (3D print with tactile surface)	Indoor
Damala et al. [57]	Exhibition enhancement	Hybrid (IMU and markerless)	Mobile	Tangible (iPhone in a loupe-like wooden case)	Indoor
Breuss-Schneeweis [35]	Exhibition enhancement	Marker-based	Mobile	Tangible	Indoor
Invitto et al. [113]	Education	Hybrid	Desktop	Natural	Indoor
Petri et al. [171]	Exhibition enhancement	Markerless	Mobile	Tangible	Indoor

Table 3.2 The purposes and enabling technologies utilized by VR applications in cultural heritage.

Application	Purpose	Tracking	Display	Interface	Setting
Wojciechowski et al. [240]	Virtual museum	No pose tracking (pre-rendered perspective)	Desktop screen	Web-based (Mouse and keyboard)	Non-immersive
Zara [246]	Virtual museum	No pose tracking	Desktop screen	Device-based	Non-immersive
Hsieh et al. [110]	Virtual museum	No pose tracking	Projector	Hybrid (Device-based and sensor-based)	Semi-immersive
Marton et al. (2014)	Exploration	Custom-designed (Image-based)	Back-projected screen	Device-based	Semi-immersive
Reunanen et al. [190]	Reconstruction	No pose tracking	Back-projected stereo screen with goggles	Sensor-based (through natural gesture)	Semi-immersive
Bustillo et al. [37]	Education	Hybrid (optical and markerless)	Projector	Device-based	Semi-immersive
Bruno et al. [36]	Virtual museum	Optical	Stereoscopic screen	Device-based	Semi-immersive
Haydar et al. [101]	Virtual museum	Optical	Large screen with LCD glasses	Device-based	Semi-immersive
Pietroni et al. [176]	Reconstruction	No pose tracking	Screen/wall	Sensor-based (through natural gesture)	Semi-immersive
Richards-Rissetto et al. [191]	Education	No pose tracking	3D stereo display	Sensor-based	Semi-immersive
Christou (2006)	Virtual museum	Electromagnetic	CAVE	Multimodal	Fully-immersive
Hernández et al. [13]	Virtual museum	Hybrid (Acoustic and Inertial)	HMD	Sensor-based	Fully-immersive
Barsanti et al. [89]	Virtual museum	Hybrid (IMU and optical)	HMD	Sensor-based	Fully-immersive
Katsouri et al. [120]	Virtual museum	Optical	CAVE	Device-based	Fully-immersive

Table 3.3 Comparative Analysis of Cultural Heritage Projects Utilizing VR

Project	**Mobility**	**Interact. in two ways**	**On-location Story-telling**	**Gamif.**	**Sense of virtual envir.**	**3D spatial multi. represent.**	**Guid.**	**UNESCO ICH inscr.**		**3D spatial Audio**	**Custom. exper.**
Meta museum [148]	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓
Empty museum [106]	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗
Wide FOV displays [243]	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗
Cypriot CH [139]	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗	✗
Santa Maria project [79]	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗
World war I [157]	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗
Mobile VR [167]	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
3DCG [210]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Tokoloshe [205]	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗
HieroQuest [177]	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗
old bridge in Mostar [200]	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	✗
Dunhuang Cave [96]	✓	✗	✓	✗	✓	✓	✗	✓	✓	✓	✗
V-Museum [117]	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗
Hua'er in VR [138]	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗
Re-Live History [22]	✗	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗
Ancient Contexts [62]	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗
Chinese Glove [235]	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✗
Bamboo Weaving [248]	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗
El-FnaVR	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓: The project implemented the technique.
✗: The project did not implement the technique.

in Pietrasanta, Italy [40]. Research has been conducted into the application of VR systems for the inheritance of intangible cultural heritage in gongs and drums [46]. A unique example of digitizing ICH involves the African oral tradition about the Tokoloshe, cautionary tales passed down through generations [205]. An immersive VR system for exploring the UNESCO-listed Dunhuang Caves was developed [96]. Their user study highlights the importance of spatial context and interactive elements in enhancing users' understanding of cultural heritage within virtual environments. Finally, an interactive VR system was developed to preserve "Huaer," a traditional oral performance that is part of China's national ICH [138].

In our comprehensive analysis of existing VR projects in cultural heritage preservation, as summarized in the provided Table 3.3, a notable trend emerges: while many projects have successfully incorporated VR technology, they predominantly focus on tangible cultural heritage, leaving a significant gap in the representation of Intangible Cultural Heritage (ICH), especially those recognized by UNESCO. This gap is partly due to the inherent complexity of ICH activities and the high level of interactivity required to faithfully replicate them. Most previous efforts, like the Meta Museum [148], Empty Museum [Hernández et al.], and Wide FOV displays [243], focus mainly on virtual world creation, offering limited user mobility and interaction. Projects like HieroQuest [177] show advancements in interactive elements but often fall short of providing comprehensive sensory experiences and deep engagement with ICH. Against this backdrop, El-FnaVR distinguishes itself with its rare use of 3D spatial audio, an essential feature for authentically recreating the atmosphere of Jemaa El-Fna, a UNESCO-inscribed ICH site. Jemaa El-Fna, known for its diverse and intricate activities, serves as an exemplary model of ICH. El-FnaVR captures this essence by offering an immersive experience that transcends visual fidelity, embracing auditory and interactive dimensions. By focusing on two-way interactions, gamification, and user preferences, El-FnaVR not only conserves the dynamic spirit of Jemaa El-Fna but also pioneers a new path in VR for cultural heritage, demonstrating the technology's potential to vividly encapsulate and convey the richness of the world's diverse cultural landscapes.

3.4.4 MR Applications

Most surveyed MR applications are designed for reconstruction purposes, followed by education, exploration, and virtual museums. Table 3 provides details on these applications. Most are designed for non-immersive experiences. Hybrid tracking methods are commonly used, often combining GPS and markerless tracking, GPS and IMU, as well as inertial, electromagnetic, and acoustic tracking methods. Mobile displays are frequently used to present visual and audio content, although some systems utilize custom-built HMDs and combinations of different types of presentation devices to display real-virtual content.

MR applications in the CH domain are not as widespread as AR and VR, which is understandable given that the technological aspects of MR are still developing. However, as robust real-time tracking, 3D registration, realistic virtual environments, natural interfaces, and advanced presentation devices become more prevalent, more MR applications will likely emerge in the CH domain. Given the current systems in the domain, hybrid tracking, HMDs, mobile displays, and tangible interfaces seem to satisfy the requirements for implementing MR in the CH domain, especially when focusing on reconstruction.

3.5 Virtual Reality Evaluation

The Technology Acceptance Model (TAM), conceptualized by Davis [59], is widely utilized by researchers to assess individuals' attitudes and intentions toward adopting new technological systems. Davis initially adapted reasonable action theory into what is now known as the TAM [60], which serves to forecast the adoption and workplace utilization of technological innovations. Fundamental to TAM are two primary factors: perceived usefulness (PU) and perceived ease of use (PEU), which collectively drive user motivation. In this context, PEU signifies the extent to which an individual anticipates the technology to be free of effort. It has been identified that not only external elements but also PEU can significantly impact PU. The resultant attitudes toward new technology, shaped by PU and PEU, can subsequently dictate the system's actual application. Subsequently, Davis and others [61] expanded upon the original framework to include an 'intention to use' component, further influencing PU and user attitudes, hence affecting technology's actual usage.

The latest iteration of the TAM was articulated by Venkatesh and Davis [224], who posited that PU and PEU directly sway the behavioral intention (BI), thus omitting the need for separate constructs for attitude and intention. This evolution led to the creation of a BI factor directly affecting usage behaviors. Building on these foundations, Venkatesh et al. [225] developed the Unified Theory of Acceptance and Use of Technology (UTAUT), which introduces four primary determinants of user behavior: PU, PEU, social influence, and facilitating conditions, as depicted in Figure 3.5 of their work.

In the realm of virtual reality (VR), the application of TAM has become a focal point due to VR's growing prevalence. VR's acceptance was analyzed in clinical environments [29], concluding that the intention to use VR hinges predominantly on its perceived usefulness. In their 2012 investigation [76], delved into the impact of personal innovation in IT alongside the fundamental TAM constructs, PU and PEU, within a medical emergency simulation context, underscoring the predictive capacity of these constructs on the intent to utilize VR. In a more recent study, TAM framework by incorporating 'perceived enjoyment' was extended to

Table 3.4 The purposes and enabling technologies utilized by MR applications in cultural heritage.

Application	Purpose	Tracking	Display	Interface	Setting
Schnädelbach et al. [199]	Reconstruction	Hybrid (rotation sensors and GPS)	Custom-built (Tripod-mounted display)	Tangible (Device-based)	Non-immersive
Magnenat-Thalmann et al. [143]	Reconstruction	Markerless	Mobile (laptop)	Tangible	Non-immersive
Dow et al. [69]	Education	Hybrid (GPS and IMU)	Mobile (Audio presentation device and tablet)	Hybrid (Tangible, device-based)	Non-immersive
Chrysanth et al. [50]	Reconstruction	Marker-based	Plasma screen	Tangible	Non-immersive
Durand et al. [70]	Reconstruction	Sensor-based (IMU)	Mobile	Tangible	Non-immersive
Okura et al. [166]	Reconstruction	Hybrid (GPS and Markerless)	Mobile	Hybrid (Tangible, device-based)	Non-immersive
Santos et al. [197]	Reconstruction	Hybrid (markerless, GPS, IMU)	Custom-built (HMD-like)	Collaborative	Semi-immersive
Benko et al. [26]	Exploration	Hybrid (inertial, electromagnetic and acoustic)	Combo (HMD, projected table, large screen)	Hybrid (Multimodal and collaborative)	Fully immersive
Magnenat-Thalmann and Papagiannakis [142]	Reconstruction	Markerless	HMD	Hybrid (Tangible, device-based)	Fully immersive

gauge acceptance of VR hardware, highlighting the evolving nature of technology acceptance models [145].

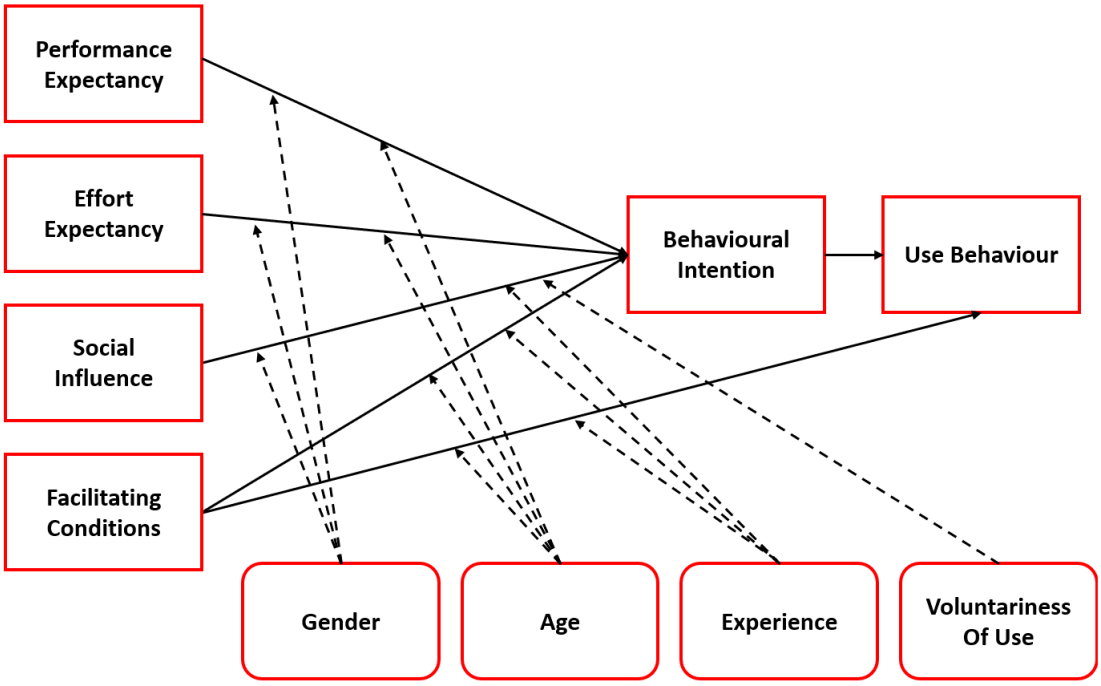


Fig. 3.5 UTAUT [225].

3.6 Conclusion

The literature review has explored key technological advancements in cultural heritage preservation, spanning terrestrial photogrammetry, 3D human pose estimation, and virtual reality applications. It has highlighted the transformative potential of these technologies in enhancing documentation, analysis, and interactive experiences related to cultural heritage. Despite the progress, the review also underscores existing limitations and the need for further research, particularly in improving accessibility, reducing costs, and enhancing user engagement. The insights gained from this review serve as a foundation for developing more effective and inclusive preservation strategies. As these technologies continue to evolve, they hold the promise of making cultural heritage more accessible and engaging, ensuring its preservation for future generations. The ongoing challenge for scholars and practitioners is to harness these technologies responsibly, ensuring they complement traditional preservation methods and contribute to a deeper understanding and appreciation of our shared cultural legacy.

Chapter 4

Low-cost Terrestrial Photogrammetry for 3D Modeling of Historic Sites

4.1 Introduction

In an era where technology's rapid advancements continually redefine the preservation of cultural heritage, it becomes imperative to harness tools and techniques that are not only effective but also economically viable. This chapter delves into the fusion of ubiquity and innovation by exploring the potential of terrestrial photogrammetry through tools readily available to a vast majority, including archaeologists, researchers, and even lay users.

As of the close of 2019, the global landscape witnessed a staggering 5.2 billion mobile service subscribers, accounting for roughly 67% of the world's population. Even more intriguing is the fact that of these, nearly 3 billion are smartphone users [65]. Such widespread adoption of smartphones underscores their potential to be more than just communication devices. They represent an opportunity, especially in the realm of cultural heritage conservation.

To contextualize and assess this potential, this chapter embarks on a comparative study, juxtaposing two distinct digital cameras – a high-end DSLR and an entry-level smartphone camera. In the quest to strike a balance between cost and accuracy, control points further play a crucial role. Our choice of subject for this endeavor is deliberate and poignant. The Marinid tombs of the Marinid Royal Necropolis, constructed in 1582, stand as a testament to Morocco's rich history. Nestled within the historic precincts of Fez's old medina – a UNESCO World Heritage site – these tombs bear the scars of time, having endured water erosion and neglect, resulting in a deterioration that underscores the urgency of preservation [42].

Central to our investigation are two critical research questions: In what ways can terrestrial photogrammetry be optimized for the 3D reconstruction of cultural heritage sites? How do these methods impact the sustainability and accuracy of heritage preservation? By addressing these questions, we aim to elucidate the potential enhancements in terrestrial photogrammetry techniques and evaluate their implications on the enduring preservation and accurate documentation of heritage sites.

This chapter provides a comprehensive examination of how accessible technologies, particularly smartphones, can be leveraged in photogrammetry to achieve sustainable and precise cultural heritage conservation. The study's findings promise to offer valuable insights into optimizing photogrammetry practices and advancing the field of heritage preservation.

4.2 Study area and data acquisition

4.2.1 Marinid Tombs

Between the 13th and 15th centuries, the Marinids were the kings of the Maghreb and the Iberian Peninsula. They built a large part of the old city of Fez, "Fez el-Bali." They were the ones who made Fez the cultural capital of the country. The medina, which was listed by UNESCO in 1981, has become an icon of architecture and history. The sultans were given beautiful tombs after their hard work. The first was buried in Rabat, in the necropolis of Chellah, and the last in Fez, in Marinid Royal Necropolis (Figures 4.1 and 4.3). After the fall of the Marinids, the following dynasties did not wish to maintain the tombs. The sites fell into ruin, worn out by sun, wind, and weather Figure 4.2 Today, the Marinid tombs are dilapidated. There are still a few structures on which beautiful motifs and friezes can be seen, but it is evident that the site has lost some of its splendor. That is why it is necessary to work on the conservation of this dilapidated heritage in order to avoid losing it.

One tomb was chosen for the reconstruction. The tomb also is in a bad situation, and it has lost some of its parts, especially the roof and the corners. The tomb dimension is approximately 9.5 m in height and 7 m in width. The photographs were taken during the journey using three different sensor cameras. Table 4.1 below shows the list of the cameras and their sensor parameters.

4.2.2 Taking pictures

Photographs were captured using an entry-level smartphone camera (Xiaomi Redmi 7) and DSLR camera sensors (Canon EOS 800D). Thus, the sizes of the images are dependent on the camera used. The images were taken during the day and without a tripod or any



Fig. 4.1 Study area from Google Maps.



Fig. 4.2 Marinid tombs ruins (‘Les tombeaux des mérinides à Fès’, 2016)



Fig. 4.3 Location of the site at the top of the hill overlooking the old medina of Fez

Table 4.1 Characteristics of utilized DSLR and mobile phone cameras.

Sensor	Image size	Focal length (mm)	Number of used images	ISO	Price points
Canon EOS 800D	6000 x 4000	18	7	100	DSLR (expensive)
Xiaomi Redmi 7	3000 x 4000	3.84	7	100	Entry-level Budget

fixing tools with a concern to get as close as possible to limited use conditions. The camera settings were fixed according to sensor types, including maximum f-numbers, fixed focus, and minimal ISO to improve the quality of the images and to reduce noise.

Two sets of several images with different resolutions were taken on different days, close to the building, but just seven from each set were chosen to make the tests after the elimination of redundancy

4.2.3 Reference data

Photographs were captured using an entry-level smartphone camera (Xiaomi Redmi 7) and DSLR camera sensors (Canon EOS 800D). Thus, the sizes of the images are dependent on the camera used. The images were taken during the day and without a tripod or any fixing tools with a concern to get as close as possible to limited use conditions. The camera settings were fixed according to sensor types, including maximum f-numbers, fixed focus, and minimal ISO to improve the quality of the images and to reduce noise. Two sets of several images

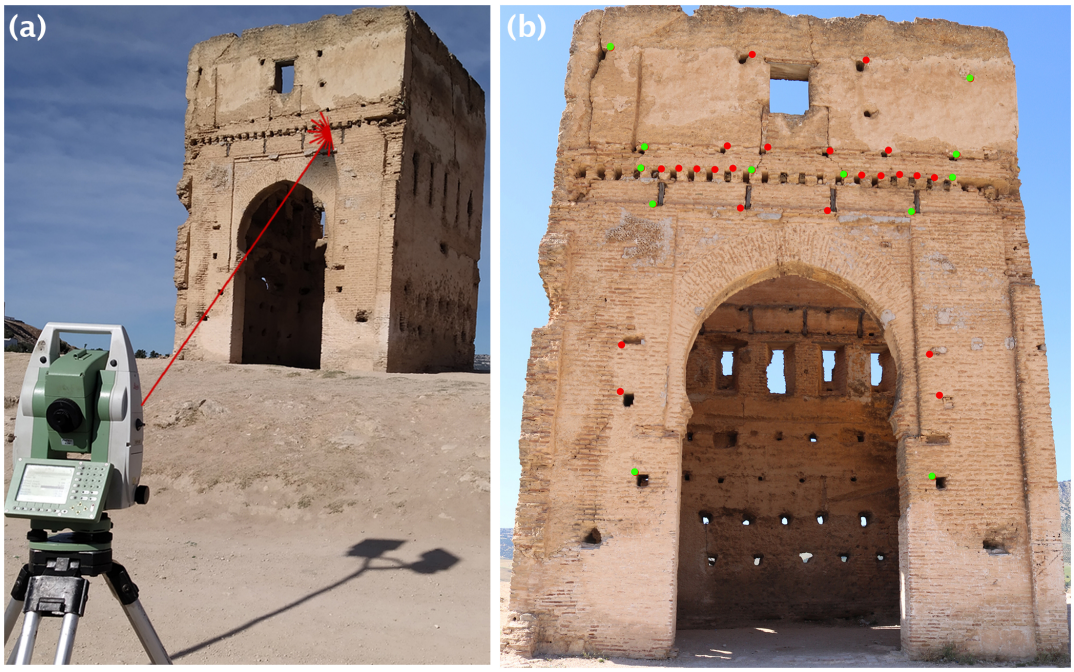


Fig. 4.4 The position of the used total station. (b): The building facade and the chosen CPs.

with different resolutions were taken on different days, close to the building, but just seven from each set were chosen to make the tests after the elimination of redundancy

4.3 Methodology

4.3.1 From images to point clouds

A point clouds is a set of points that represent the 3D objects using the geometric coordinates system, each of the point clouds that forms the clouds representation are defined by X, Y, and Z coordinates and by its color characteristics. The point clouds are obtained by the identification of several photographs. Therefore, the density is influenced by the quality and quantity of the image sets. The point clouds can extract valuable information, such as the accuracy of point positions and other metric information, by using several techniques and algorithms, including the used software. The processing involves camera calibration, image orientation, and finally generating the point clouds

Bundle adjustment

Bundle block adjustments (BA) have been conceived by the photogrammetric community in the ’50s, and used to compute 3D coordinates through a bundle adjustment method that

was developed for the first time by Brown in 1958. It is an optimization problem on the sought 3D structure and camera parameters, such as position, orientations, and eventually internals, in order to estimate 3D reconstruction. BA has become the standard solution for 3D reconstruction problems in the '80s.

In our experience, two different BA solutions were proposed to make the comparison between different cameras from different ranges. The first one is the use of 12 CPs that were computed employing the total station, as shown in Figure 4.4. The second one is to reconstruct the object without using any CPs; the model was controlled by using scale bars. Agisoft Metashape Professional software (commercial software developed by Agisoft Corporation) was used to build models and to compute the residuals

Ground control points

Agisoft Metashape professional version allows users to add additional measurements that could improve the results such as CPs and the Scale Bars. In this case, 12 CPs were chosen and measured using a local coordinates system (Merchich/North of Morocco). Those points are already known on the facade of the building. It remains to match each point in the photographs to its corresponding coordinates obtained by the total station. The accuracy obtained using the software for both of the cameras was computed by comparing the CPs coordinates from the total station, and the CPs coordinates obtained using the software. A second test was conducted by reducing the number of CPs to six and evaluating the impact on the accuracy, as depicted in Figure 4.5. The third test was about the creation of the models without any CPs, and only scale bars were employed to control the models at that last test.

4.3.2 Cloud-to-cloud distance and data comparison

CloudCompare is a user-friendly open-source processing software [5], oriented mainly to calculate the distances between two dense 3D point clouds [88]. The software has several tools to calculate the distances; in our case, we used the cloud-to-cloud (C2C) distance tool. It measures the difference between each photogrammetric point clouds and a reference cloud. Several performing algorithms were used to calculate the distance, such as the iterative closest point (ICP) [7]. As shown in Figure 4.6, CloudCompare takes the nearest point in the reference point clouds and computes their C2C distance using the nearest neighbor distance algorithm.



Fig. 4.5 Configuration of the 4, 6, 8, and 10 control points (green points) from 34 collected points (red points). Clockwise from top-left

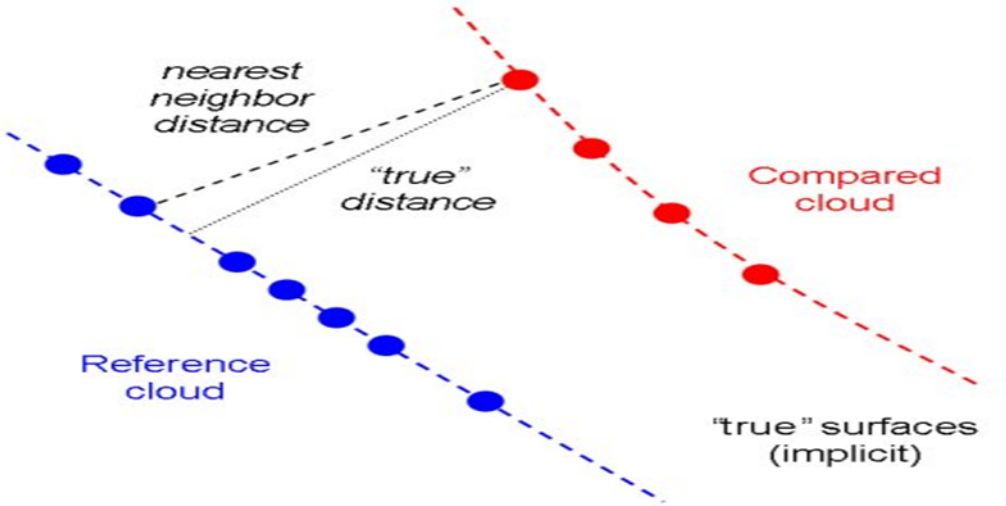


Fig. 4.6 The nearest neighbor distance that cloud-to-cloud (C2C) distancing is based on

4.4 Results and Analysis

The software used to obtain points and process and create image-based 3D models of point clouds is the Agisoft Metashape professional version. The software extracts the tie points before the external orientation of the camera is automatically calculated from these points. For the camera calibration, Metashape uses Brown's distortion model to simulate the distortion of the lens (Agisoft and R., 2018). As shown in Figure 4.7, the parameters of accuracy and the number of tie points were set at the best configurations possible.

4.4.1 Comparison of the bundle block adjustment for the cameras

In order to compare the two cameras, two different bundle block adjustments were used. The first one was based on the CPs, and the second one used only the software without using any CPs.

BA accuracy based on using CPs

The statistical parameters of the Canon and the Redmi cameras measured after georeferencing using 12 CPs are summarised in Table 4.2. For the Redmi camera, one CP was considered an outlier as it affected the results negatively; consequently, it was removed. The RMS (Root Mean Square) values of the two cameras were calculated spatially and along the three spatial axes. For the Canon camera: the observed RMS value along the x-axis was 24.99 mm, the RMS value along the y-axis was 22.08 mm, the RMS value along the z-axis was 15.27 mm,

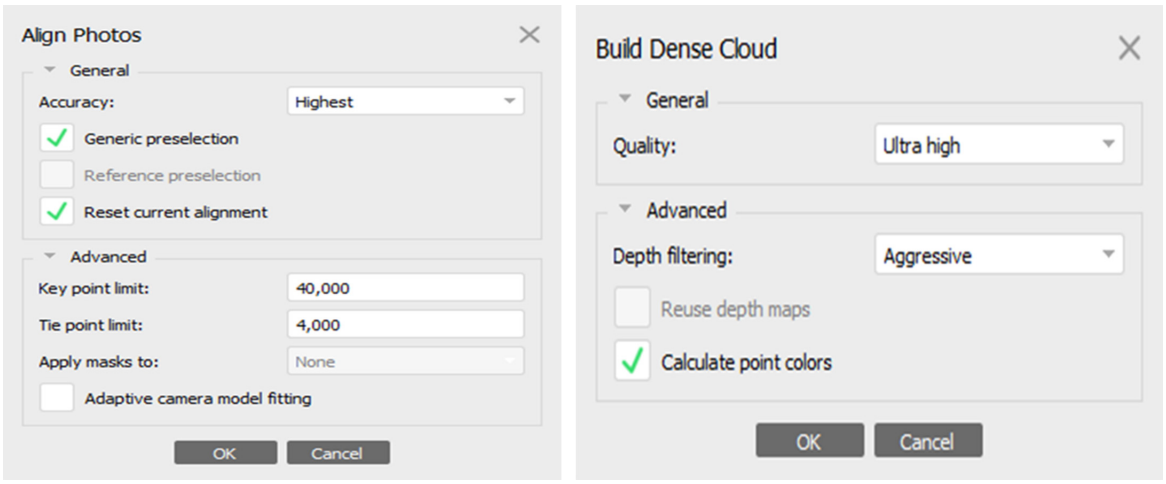


Fig. 4.7 Initial using parameters in Agisoft Metashape

and the spatial RMS was 35.12 mm. For the Redmi camera, the RMS value along the x-axis was 9.75 mm, the RMS value along the y-axis was 15.51 mm, the RMS value along the z-axis was 11.93 mm, and the spatial RMS value was 21.86 mm, which is more accurate than canon camera. As shown in the histogram and Q-Q plot for the Canon camera Figure 4.8, the residual data are typically distributed. As can be seen, the central part of the curve follows the straight line well. The skewness value is between -0.5 and 0.5, which means an asymmetrical distribution of the data. The kurtosis value is negative -0.8, which means that the data follows a platykurtic distribution with flat tails that indicate the small outliers in the distribution.

Figures 4.9 and 4.10 show the histogram and Q-Q plots of discrepancies of the Redmi camera before and after the removal of the outliers. The histogram before filtering shows that the distribution is not similar to a normal distribution while it is highly positively skewed. The positive excess kurtosis on its side indicates a leptokurtic distribution with a heavy tail, which indicates significant outliers. As can be seen, after removing the outlier, the tails of the residual distribution are flat and close to the normal distribution while the standard deviation was improved to 6.86 mm from 10.99 mm, and the data follow a straight line as shown in the Q-Q plot (Figure 4.10).

On the other hand, the mean value of the Redmi camera was around 2 cm and 3.5 cm for the Canon camera, and their standard deviations were between 6.87 mm and 15.67 mm.

Cloud-to-cloud comparisons

In order to compare the two proposed low-cost cameras, we used 34 control points that were taken using the total station as a reference instead of the laser scanner point clouds. On the

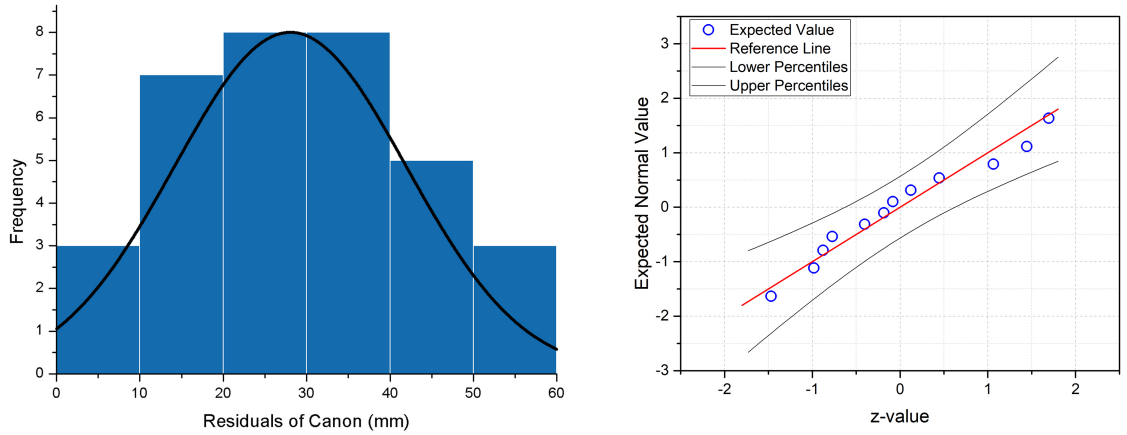


Fig. 4.8 Histogram and Q-Q plot of photogrammetry using Canon and Total station residuals with normal distribution

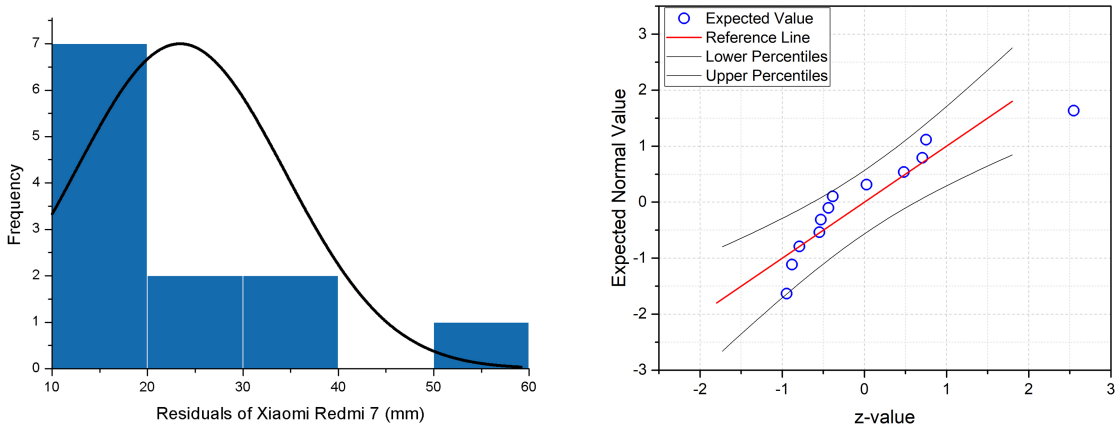


Fig. 4.9 Histogram and Q-Q plot of photogrammetry using Redmi phone and Total station residuals with normal distribution (before outliers removing)

Table 4.2 Obtained statistical parameters of the Canon and Redmi 7 cameras results

		Canon	Redmi	Redmi (After outliers)
X-direction	Mean (mm)	0.00	0.00	-3.01
	Standard deviation (mm)	24.99	13.95	9.72
	RMS (mm)	24.99	13.95	9.75
Y-direction	Mean (mm)	0.00	0.00	3.55
	Standard deviation (mm)	22.07	19.48	15.83
	RMS (mm)	22.08	19.48	15.51
Z-direction	Mean (mm)	0.00	-0.42	0.38
	Standard deviation (mm)	15.27	12.25	12.50
	RMS (mm)	15.27	12.26	11.93
Spatial position	Mean (mm)	31.75	23.4	20.86
	Standard deviation (mm)	15.67	10.99	6.87
	RMS (mm)	35.12	25.66	21.86
Skewness	-	0.42	1.63	0.63
Kurtosis	-	-0.8	3.03	-1.14

other side, two models of points clouds of the two cameras were extracted from Agisoft Metashape software. Each of these models was compared with the reference model using CloudCompare as a comparison tool that calculates the cloud-to-cloud distances (C2C) by taking each point from the compared model, finding its nearest neighbor on the reference model, and then computing their distance. Table 4.3 summarizes the results of several statistical variables of comparison between the CPs and point clouds models. The first statistical variable is the mean value of all distances calculated by CloudCompare software, which was used to compute the accuracy. The second one is the standard deviation, which was used to take an idea about the error margin, and the third one is the RMS which gives high weight to large residuals. As shown in Table 4.3, the mean C2C distance value of the Redmi has a lower value (6.84 mm) than that of Canon (11.66 mm). The residuals of the Canon are more significant when compared to that of Redmi. That is confirmed by standard deviation and RMS values. In Figure 4.11, the histogram on the right indicates that 91% of points were under 7.4 cm using the Redmi Model, while the histogram on the left shows that 91% of points were smaller than 11.4 cm in the case of the Canon camera.

4.4.2 Relative comparison between Canon and Redmi

CloudCompare software allows making a relative comparison between two models of point clouds. Therefore, the achieved models using Agisoft Metashape were compared by comput-

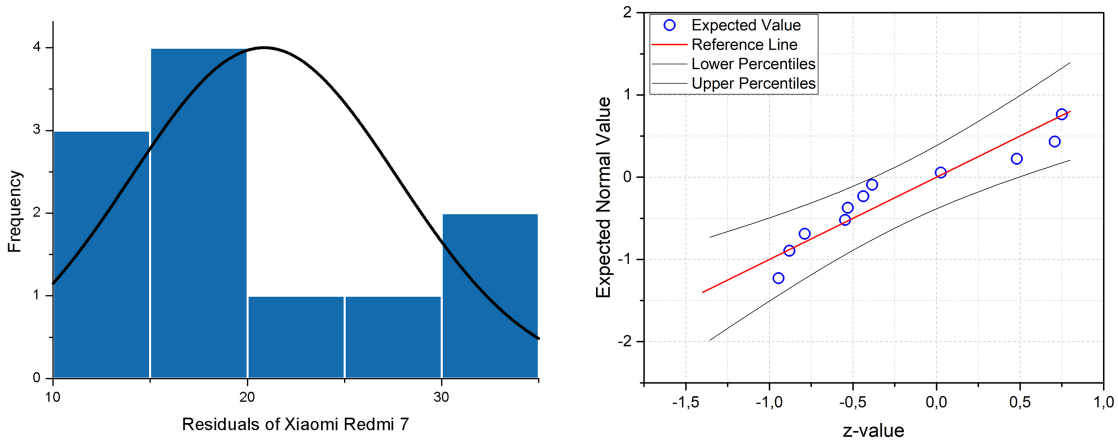


Fig. 4.10 Histogram and Q-Q plot of photogrammetry using Redmi phone and Total station residuals with the normal distribution (after outliers removing)

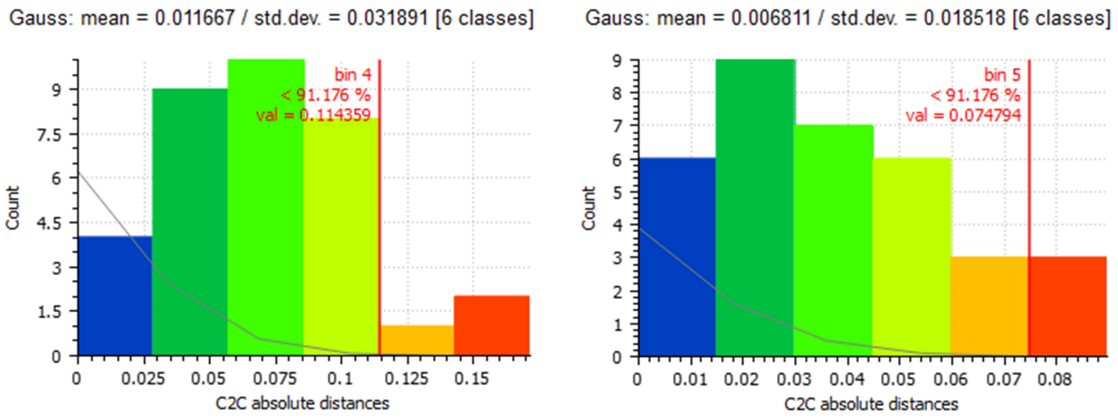


Fig. 4.11 Histograms of C2C distances for Canon model (left) and Redmi one (right)

Table 4.3 Achieved statistical results of C2C distances between compared models and point clouds of total station as a reference

Reference model	Compared model	Mean (mm)	Standard deviation (mm)	RMSE (mm)
34 points	Canon	11	31	93
	Redmi	6	18	54

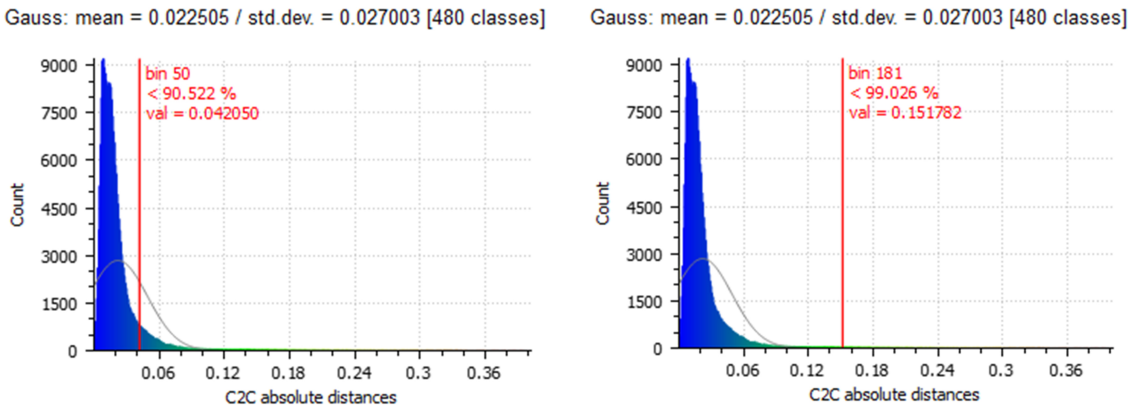


Fig. 4.12 Histograms of C2C distances of Canon camera using Redmi phone camera as a reference model

Table 4.4 Achieved statistical results of C2C distances for relative comparison between Redmi as a reference and Canon as a compared model

Reference	Compared	Mean (mm)	Standard deviation (mm)	RMSE (mm)
Redmi	Canon	22	27	35

ing cloud-tocloud distances (C2C). Since the Redmi model was found more accurate than the Canon one, it was selected as a reference, while the Canon model was the compared model. As shown in Table 4.4, the mean and the standard deviation values of cloud-to-cloud distances were 2.25 cm and 2.7 cm, respectively, and the RMS value was 3.15 cm.

The histograms shown in Figure 4.12 indicate that for 90% of points, the mean of C2C distances is under 4.2 cm, while it is less than 15.17 cm for 99% of points are less than 15.17 cm. Figure 4.13 is a graphical representation of the C2C distances between the reference and compared point clouds, where the blue area indicates a lower distance (less than 5 cm), and the green area presents distances around 12 cm.

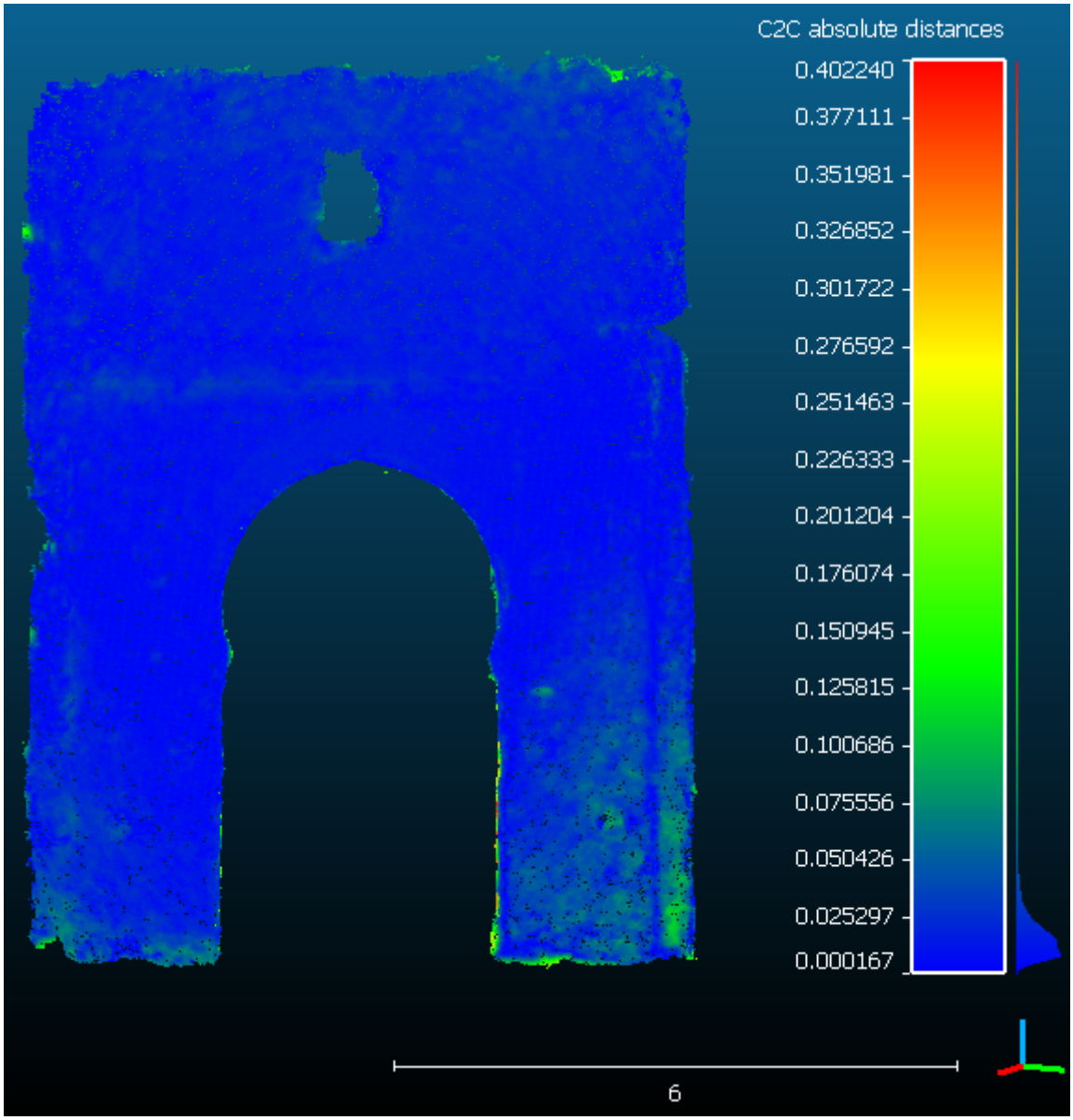


Fig. 4.13 C2C distances of Canon and Redmi relative comparison. Blue indicates low distances

Table 4.5 obtained statistical results of C2C distances between compared models (without using CPs) and point clouds of the total station as a reference

Reference model	Compared model	Mean (mm)	Standard deviation (mm)	RMSE (mm)
34 points	Canon	21	60	66
	Redmi	11	32	70

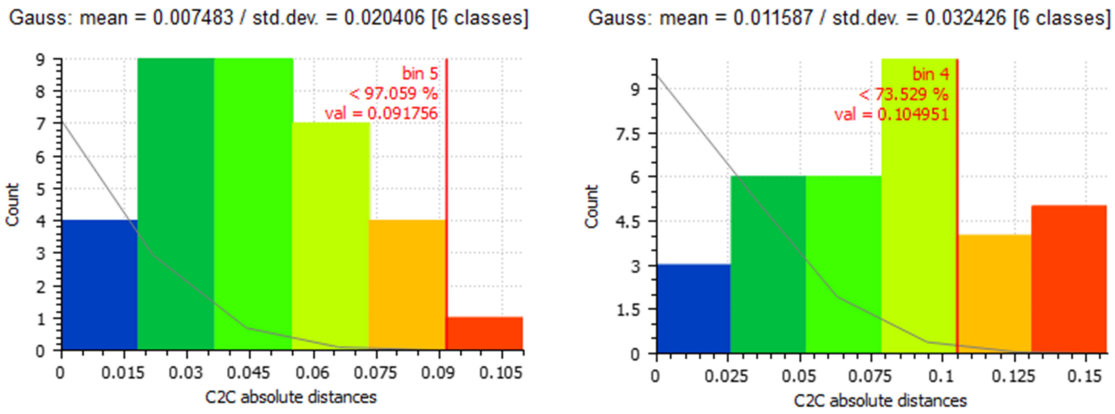


Fig. 4.14 Histograms of C2C distances for Canon model (left) and Redmi one (right)

4.4.3 Bundle block without using CPs

The self-calibration proposed by Agisoft Metashape allows automatic extraction of the tie points without any CPs before the external and internal orientation is computed from them (tie points). The obtained point clouds models were controlled by scale bars. The models were extracted with the local coordinates system (Merchich/North of Morocco), which were compared to the aligned reference points using iterative closest points (ICP) before computing C2C distances. The 34 CPs taken by the total station were considered as a reference model. The results of the statistical parameters are summarised in Table 4.5. The mean value of the Redmi was always better (11 mm) than that of the Canon (21 mm). The standard deviation and RMS of the Redmi (32 mm and 94 mm, respectively) were less than those of the Canon (60 mm and 176 mm respectively), and that means that the residuals of Canon were more significant than that of the Redmi. As shown in Figure 4.14, over 73% of the points have a mean of C2C distances within 10 cm for the Redmi camera, while for the Canon camera, 85% of points have a mean of C2C distances that is under 22 cm for the Canon camera.

Table 4.6 Statistical results of the models as a function of CPs number

	Canon			Redmi		
CPs number	**Mean (mm)**	**SD (mm)**	**RMSE (mm)**	**Mean (mm)**	**SD (mm)**	**RMSE (mm)**
0	21	60	66	11	32	70
4	11	34	47	5	16	42
6	10	32	44	5	16	35
8	10	31	39	6	17	31
10	10	31	39	6	17	32

4.4.4 Effect of the number of CPs

In order to study the effect of the number of CPs and their distribution on the accuracy of the results, several tests have been conducted. Each test has given a reduced number of CPs. Combinations of 10, 8, 6, and 4 CPs have respectively been tested. Agisoft Metashape Professional software has been used to obtain the five models corresponding to the different combinations above. The CPs were distributed, as shown in Figure 4.5. The obtained 3D models of each set of points were compared into CloudCompare software to the laser total station model, which is used as a reference model. The achieved results involving accuracies, mean values, and standard deviation of C2C distances are presented in Table 4.6.

The results suggested that both cameras exhibited stable means distances and standard deviation, as shown in Figures 4.15 and 4.16 for the models obtained using CPs. Redmi models yielded accurate results (under 6 mm), better than Canon for all scenarios.

4.5 Discussion & conclusion

This chapter presented an evaluation of the accuracy of two solutions for low-cost image-based 3D modeling of the Marinid Royal Necropolis in Fez. Smartphone camera and Canon DSLR camera have been used through different configurations, with or without CPs collected by a total station. The aim was to investigate the impact of many parameters, such as camera type and the proposed bundle block adjustment solutions. The 3D models were obtained using Agisoft Metashape professional software and were compared to a set of CPs.

The results indicate that the models obtained using CPs were generally more accurate, with mean values of the C2C distances ranging between 5 mm and 21 mm. When using 12 CPs, the highest accurate model corresponded to the smartphone camera (5 mm), which results in a mean C2C distance equal to 6 mm mean. Otherwise, when four or more CPs were used, the outcomes indicated that the accuracy was stable and showed almost no changes.

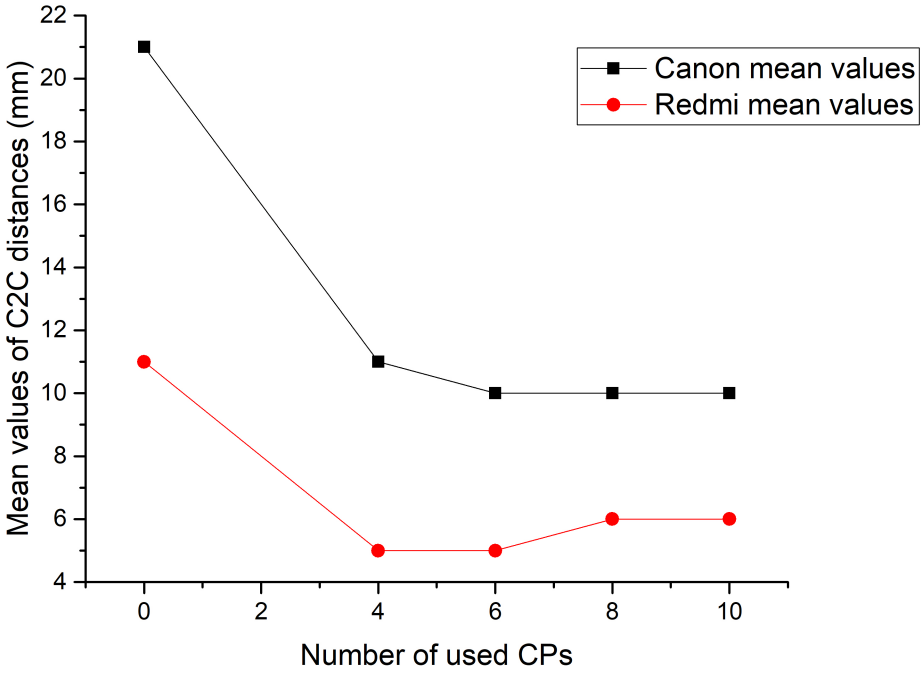


Fig. 4.15 Graph of mean variation as a function of CPs number

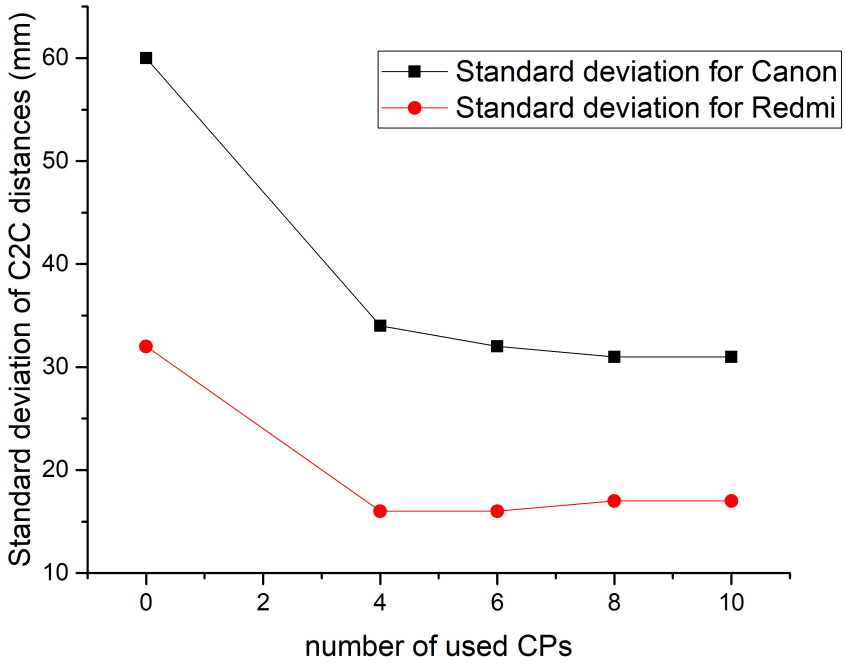


Fig. 4.16 Graph of mean variation as a function of CPs number

The smaller mean value of the C2C distances corresponded to 5 mm, and it was obtained when four or six CPs and the Redmi camera were used. Although the challenging task of CPs surveying and marker location on such deteriorated historical heritage, explains the increase in standard deviation values, the results were satisfactory.

The non-use of CPs has, as expected, negatively affected the accuracy of the obtained 3D models for both cameras. This is because of the known coordinates provided by CPs, which ensure that any points in the model correspond accurately with the real 3D coordinates. The only thing that helped the software control the 3D model was the scale bars. Despite that, the accuracies were acceptably close to the obtained ones by using CPs. They ranged between 11 mm and 21 mm. In this case, again, the smartphone camera, with an 11 mm C2C mean distance, was more accurate than the DSLR camera.

After analyzing the results obtained from this case study, it can be stated that 3D modeling of historical heritage using smartphone cameras instead of DSLR cameras is reliable, stable, and effective for many reasons such as accuracy, cost, and availability of instruments. Therefore, this solution can be an optimal close-range low-cost alternative. For obtaining more accurate results, the use of at least 4 CPs is still recommended for a facade of approximately 60 m². In the case of the non-availability of a total station for CPs surveying, the use of a scale bar is widely recommended. The working distance between the smartphone camera and the captured monument is relative to the size of the monument. The angle of view should be adapted so that the monument occupies the maximum possible of the captured image.

Chapter 5

Evaluating the Accuracy of Cloud-Based 3D Human Pose Estimation Tools

5.1 Introduction

This chapter presents an evaluation of cloud-based 3D Human Pose Estimation (HPE) tools, specifically MOTiON by RADiCAL, within the context of developing VR applications aimed at cultural heritage conservation and promotion. The primary goal is to assess the accuracy and suitability of these tools as automated solutions for animating 3D characters in VR environments. This evaluation is crucial for understanding how such technologies can be effectively adopted in the VR development process, particularly for scenarios involving intangible cultural heritage, such as the dynamic and vibrant activities of Jamaa El-Fna in Marrakech.

The research questions guiding this evaluation are:

1. How accurate are cloud-based 3D HPE tools in estimating human poses compared to ground truth data?

2. Can cloud-based 3D HPE tools be feasible alternatives to traditional motion capture systems in various scenarios?

3. What improvements are necessary for cloud-based 3D HPE tools to extend their applicability to high-precision tasks, such as advanced biomechanical analysis, sports performance analysis, or precise VR interaction?

To address these questions, we cleaned and filtered the CMU dataset for use in this study, focusing on scenarios that include a range of actions visible in videos and their corresponding

3D human poses. The 3D coordinates for each human joint were obtained from these videos. The quantitative evaluation involved comparing the results to the ground truth after Procrustes alignment [90], using metrics such as Mean Per Joint Position Error (MPJPE) and Percentage of Correct Keypoints (PCK). Qualitative evaluation involves a visual assessment of selected frames from each scenario.

Quantitative results revealed that MOTiON by RADiCAL is adequate for most scenarios, though it has limitations, particularly with occlusion and dynamic motions. Cloud-based tools like this can potentially replace expensive traditional tools for simple scenarios with slow movements and minimal occlusion. Qualitatively, nine scenarios were accurately estimated, while others had misaligned skeletons or components.

5.2 Methodology

After extracting videos from CMU and their associated BVH pose files, a preprocessing stage comprising: cleaning (i.e., avoid corrupted sequences or those that do not verify the necessary conditions), reorganization (i.e., reclassifying all sequences into 12 scenarios), and synchronization (because the BVH poses files are not synchronized with the associated videos) was performed. The sequences in video format were then processed in the cloud with RADiCAL. Both BVH poses of CMU and RADiCAL were rendered using a virtual camera to get the 3D coordinates of each joint. Then two evaluation types were performed; the first one was quantitative, which compared both poses of RADiCAL (as predicted results) and those of CMU (as a ground-truth one). The other evaluation type is a qualitative one based on visual analysis of the predicted 3D human pose scenario according to the ground truth. The workflow was summarized in Figure 5.1.

5.2.1 Data preprocessing

The CMU Graphics Lab Motion Capture Database (CMU) was chosen for this study due to its extensive and diverse range of scenarios that closely simulate real-world movements found in intangible cultural heritage. This dataset includes a variety of activities such as dancing, daily activities, sports, and more, making it an ideal reference for evaluating the effectiveness of cloud-based 3D HPE tools like MOTiON by RADiCAL in capturing complex human motions.

The CMU dataset offers several advantages:

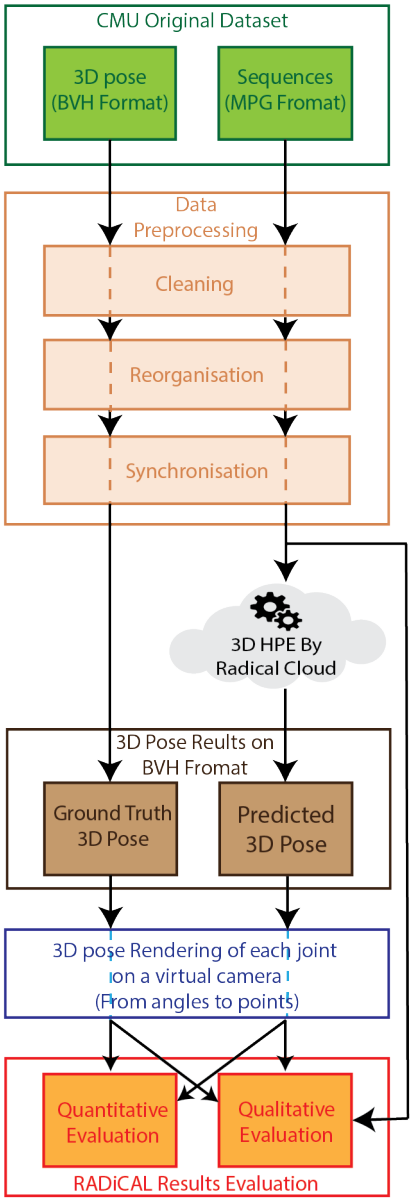


Fig. 5.1 Overview of the suggested approach.

1. **Diversity of Scenarios:** The dataset covers a wide array of actions that mirror the dynamic and varied movements found in cultural heritage activities, such as traditional dances, performances, and daily routines.

2. **Diversity of Scenarios:** The dataset covers a wide array of actions that mirror the dynamic and varied movements found in cultural heritage activities, such as traditional dances, performances, and daily routines.

3. **High-Quality Data:** Captured using 12 Vicon MX-40 infrared cameras, the CMU dataset provides precise and reliable 3D motion capture data, essential for accurate evaluation.

4. **Comprehensive Joint Data:** The dataset includes detailed 3D coordinates for each human joint, which is crucial for comparing the performance of HPE tools in replicating human poses accurately.

5. **Established Benchmark:** As a widely recognized and utilized dataset in the field of motion capture, the CMU database serves as a robust benchmark for evaluating new technologies and methods.

The preprocessing stage involved several critical steps to ensure the dataset's suitability for this study:

- **Cleaning:** some corrupted videos were eliminated, and the rest were repaired by hiding the second actor if that is possible.

- **Reorganisation:** the sequences were classified into 12 essential scenarios, as shown in Table 5.1. Additionally, the dataset was reorganized by gender, separating the sequences into male and female categories to allow for a more nuanced analysis. By comparing the results by gender, we can identify any biases or discrepancies in the tool's accuracy and ensure that it performs reliably for both male and female subjects, which is essential for creating inclusive and accurate VR applications.

- **Synchronisation:** since all BVH frames are not synced with videos, a manual process was manually done using Blender. Also, the sequences captured with 120 or 60 FPS were decreased to 30 FPS since the RADiCAL supports only motion capturing with 30 FPS.

5.2.2 MOTiON by RADiCAL

MOTiON by RADiCAL is a model-based 3D HPE AI-driven and cloud-based software that converts 2D movies into complete 3D animation with 6 degrees of freedom. The animation data is stored with 30 FPS into FBX (Filmbox), a format that allows the exchange of geometric and animation data between 3D animation software, such as Blender. For the study, the sequences in MPG format were imported to the RADiCAL cloud then the HPE was processed using the RADiCAL model. After a few moments, the FBX files were done. In order to compare those results to CMU's ground truth, the FBX output files were converted

Table 5.1 CMU dataset component after cleaning, filtering, and classification

Number of scenarios	Number of sequences	Number of views	Frequency	Scenarios	Number of frames
12	279	1	30 FPS*	Animal behaviors	62 454
				Climbing	981
				Daily activities	6 306
				Dancing	1 122
				Home activities	56 359
				Jumping	1 553
				Reactions	12 852
				Running	332
				Sitting	4 548
				Sport	9714
				Walking	16 493
				Working	11 952
				Female	71 759
				Male	112 907
				All	184 667

*FPS–Frames Per Second

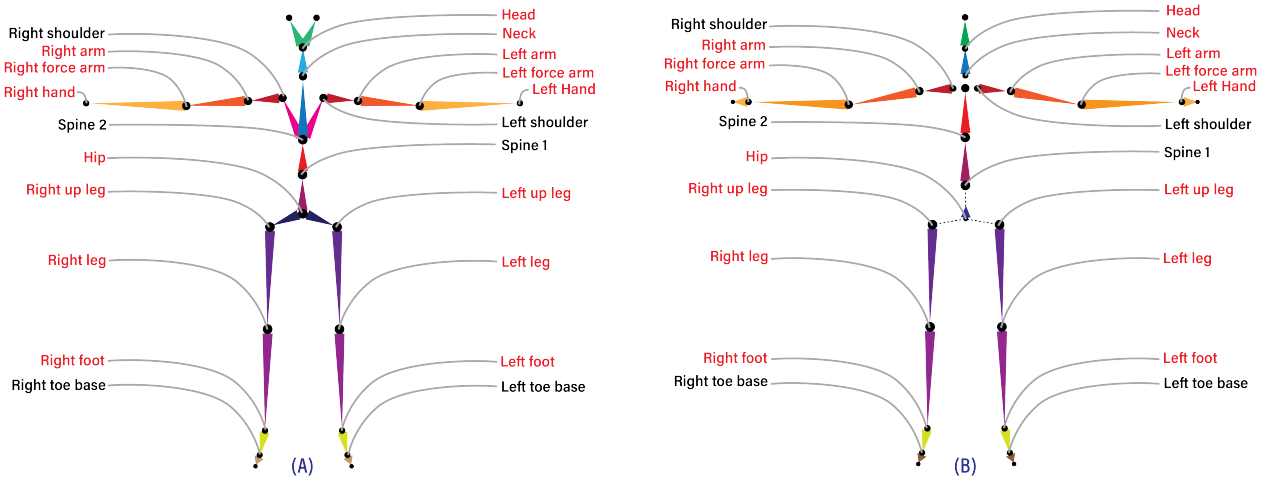


Fig. 5.2 (A): Skeleton model hierarchy of CMU. (B): Skeleton model hierarchy of RADiCAL. The red ones are the chosen joints to perform the quantitative evaluation.

to BVH format using Blender. Since the skeletons of CMU and Radical are not similar; the Procrustes analysis was used to determine the scale [90], rotation, and translation. The output skeleton and the joints are shown in Figure 5.2.

5.2.3 BVH projecting to 3D coordinates

The motion capture of videos from the RADiCAL and CMU datasets is stored in BVH format, including the root transaction coordinates and Euler angles for each joint. As illustrated in Figure 5.3, all the coordinates, including those in the BVH files, were projected to a 3D virtual camera to obtain the 3D coordinates of each joint. Algorithm 1 computes the joint coordinates in camera space from a BVH file containing joint hierarchy and motion data. The algorithm starts by defining the camera intrinsic matrix K, which represents the camera's internal parameters such as focal length and principal point. The camera extrinsic matrix C is also defined, which represents the camera's external parameters such as position and orientation in global space.

$$K = \begin{pmatrix} f_x & 2 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5.1}$$

Where f_x and f_y are the focal lengths of the camera in x and y directions, and c_x and c_y are the coordinates of the principal point of the camera.

$$C = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \tag{5.2}$$

Where r_{ij} is the rotation matrix that describes the camera's orientation in global space, and t_i is the translation offset. The joint positions and orientations for each frame in the motion data are then computed using forward kinematics, with the root joint's global position and orientation serving as the initial values. The global positions and orientations of the child's joints are then computed by traversing the joint hierarchy, and the resulting global joint positions are transformed into camera coordinates using the intrinsic and extrinsic matrices. This transformation can be represented mathematically as:

$$\begin{pmatrix} X_{position_global} \\ Y_{position_global} \\ Z_{position_global} \end{pmatrix} = K \times C \times \begin{pmatrix} X_{position_camera} \\ Y_{position_camera} \\ Z_{position_camera} \end{pmatrix} \tag{5.3}$$

Where $(X_{position_global}, Y_{position_global}, Z_{position_global})$ is the global joint position in 3D space, and the resulting $(X_{position_camera}, Y_{position_camera}, Z_{position_camera})$ is the joint position in camera coordinates. This algorithm provides a way to extract joint positions in camera space from a BVH file.

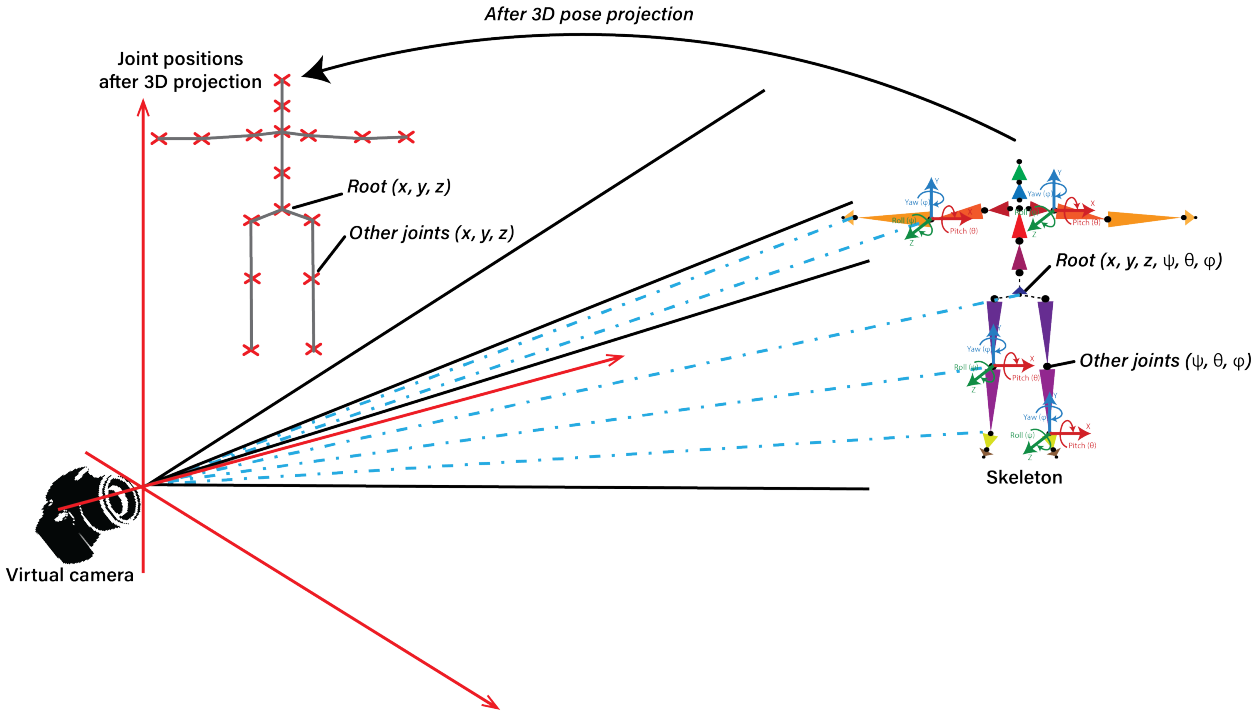


Fig. 5.3 Rendering process of CMU and RADiCAL skeletons. The purpose is to obtain the 3D pose coordinate of joints from angles.

5.3 Evaluation

As stated previously, qualitative and quantitative evaluations were performed. With the restructured CMU dataset, the initial step was to obtain the MPJPE and the PCK by scenario and gender. The second sort of evaluation consisted of picking 3D postures of various scenarios and visually analyzing the results' accuracy.

5.3.1 Quantitative evaluation

The results obtained using MOTiON by RADiCAL cloud-based were compared with the ground-truth 3D poses from the reconstructed CMU dataset. 15 keypoints were examined, which are indicated in red in Figure 5.3. using two metrics measurements (MPJPE and PCK). The 3D poses were classified by gender and scenario to assess the accuracy of each one. Then the accuracy of each joint was discussed.

Comparing by Joints

In this evaluation, 15 crucial joints were analyzed, as depicted in Fig. 5 where the red joints are highlighted. The results are presented in Table 5.2 and Figure 5.4, displaying the highest

Algorithm 1 Compute Joint Coordinates in Camera Space from BVH File

Input: BVH file with joint hierarchy and motion data
Output: 3D joint coordinates in camera space for each frame of motion data

1. Load BVH file and extract joint hierarchy and motion data.
2. Define camera intrinsic matrix K with focal lengths f_x and f_y and principal point coordinates c_x and c_y.
3. Define camera extrinsic matrix C with rotation matrix R and position vector P.
4. For each frame of motion data, traverse the joint hierarchy in forward kinematics to compute global joint positions.
5. Transform global joint positions to camera coordinates using K and C.
6. Output the 3D joint coordinates in camera space for each frame of motion data.

End algorithm.

Table 5.2 Mean error by joints

Joints	**Mean (mm)**	**Standard deviation (mm)**
Nose	85.28	36.11
Neck	87.15	27.30
RShoulder	77.96	33.85
RElbow	99	41.29
RWrist	116.96	77.72
LShoulder	70.52	30.86
LElbow	103.61	49.12
LWrist	122.73	85.58
MidHip	170.49	20.33
RHip	94.78	24.28
RKnee	80.91	37.07
RAnkle	92.17	59.06
LHip	91.13	25.73
LKnee	89.61	42.77
LAnkle	99.07	50.24

mean error values of the Middle Hip, Left Wrist, and Right Wrist joints. While, the lowest mean error values were obtained for the Shoulders, Knees, Nose, and Nick.

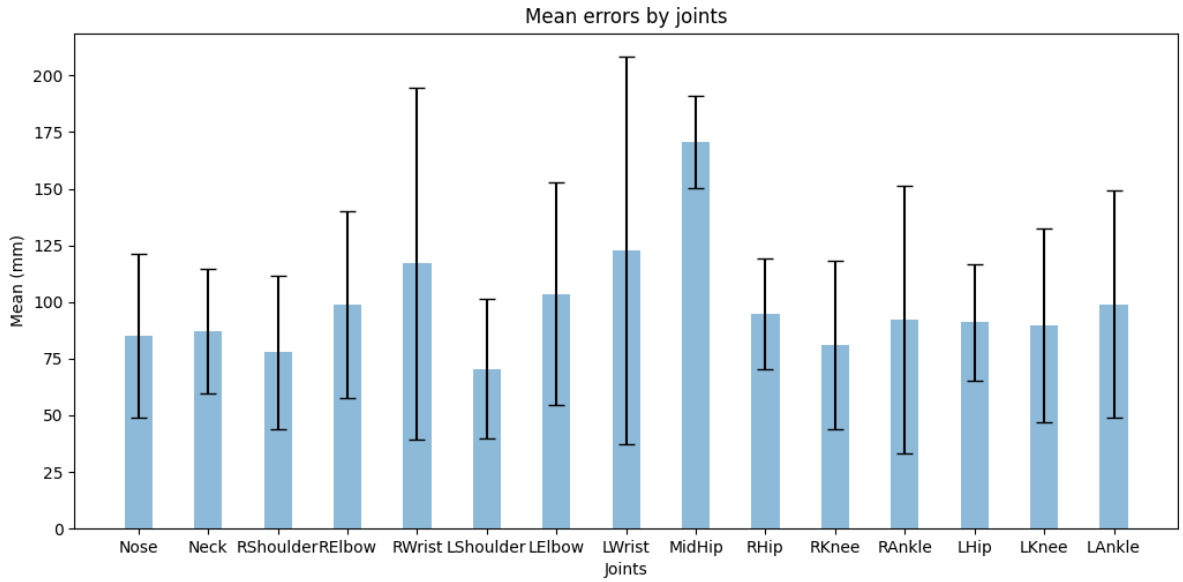


Fig. 5.4 The mean error results by joints.

Comparing by scenarios

Figure 5.5 and Table 5.3 show that the MPJPE varied from 90,7 mm to 119,1 mm, depending on the scenario. The walking scenario was the most accurate, with an MPJPE of 90.7 mm, whereas the running scenario was the least accurate. Each scenario's MPJPE (walking, jumping, dancing, reaction, and animal behavior) was under 100 mm while they were near one another, except for home activities, whose MPJPE was just under 100 mm. The MPJPE is more than 100 mm for the remaining scenarios (daily activities, working, climbing, sitting, sports, and running). Expect "Running," "Sitting," and "Sport"; all the scenarios were accurate with higher than 90% of correct joints according to the PCK values of each one. The scenarios: "Home activities," "Jumping," "Reactions," and "Walking" had a PCK near 100%. Expect running and sports scenarios with a standard deviation of around 70 mm. Every other scenario was within 50 mm. Figure 5.6 shows a clear progression of the Percentage of Correct Keypoints (PCK) in response to the changing Mean Per Joint Error (MPJPE) threshold. As the MPJPE threshold increases, the PCK also follows suit. Notably, at a relatively stringent threshold of 75 mm, the PCK already reaches about 44%, almost half of the keypoints estimated correctly within this error range. This trend continues and the PCK grows to about 96% when the MPJPE threshold is relaxed to 150 mm, indicating a substantial portion of the estimated keypoints are accurately detected within this margin of error. As we further expand the MPJPE threshold beyond 150 mm, the PCK continues to increase, albeit at a slower rate. The curve eventually approaches a saturation point near 100%, indicating that practically all keypoints are accurately estimated within these larger margins of error.

Table 5.3 MPJPE by scenarios

scenarios	**MPJPE (mm)**	**Standard deviation (mm)**	**PCK (%)**
Animal behaviors	95.62	52.33	96.5
Climbing	112.65	54.48	92.14
Daily activities	101.92	53.56	95.8
Dancing	92.14	63.09	94.55
Home activities	99.44	47.66	97.3
Jumping	91.30	40.58	99.03
Reactions	92.97	47.67	98.63
Running	119.10	72.22	83.68
Sitting	116.75	59.32	87.28
Sport	118.31	68.82	83.75
Walking	90.71	45.39	98.13
Working	104.79	56.47	91.57
All	98.76	52.06	95.8

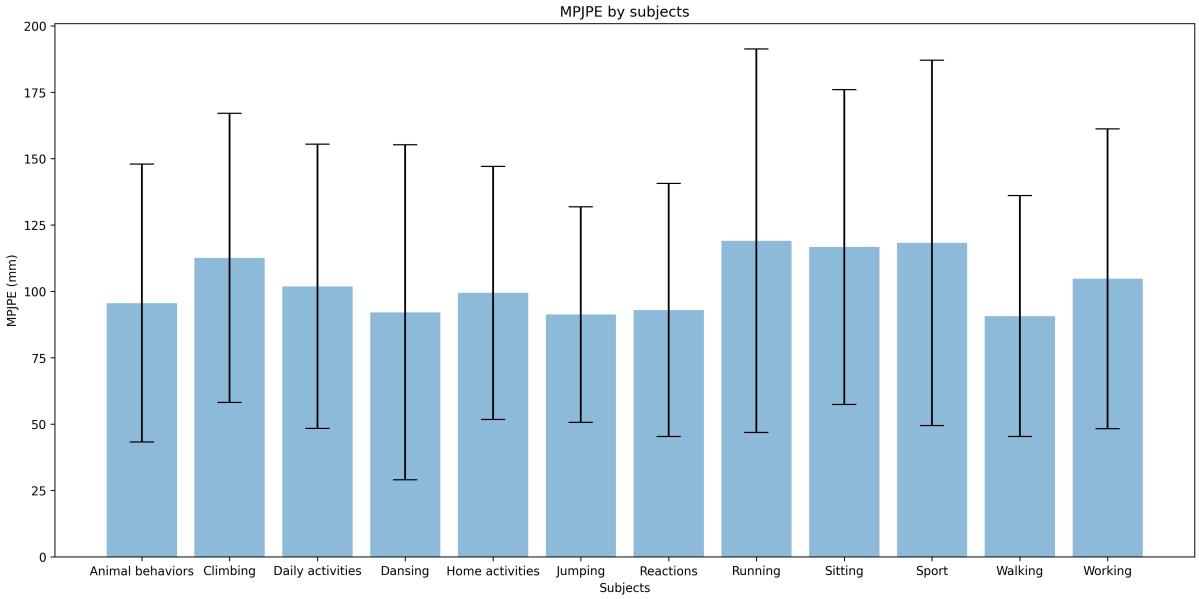


Fig. 5.5 The results of MPJPE by scenarios.

Comparing by gender

The findings of gender-based evaluations are depicted in Table 5.4 and Figure 5.7. Male and female MPJPEs were comparable, with the female MPJPE (95 mm) being 5 mm better than the male MPJPE (100 mm). The same holds for the standard deviation, which was almost

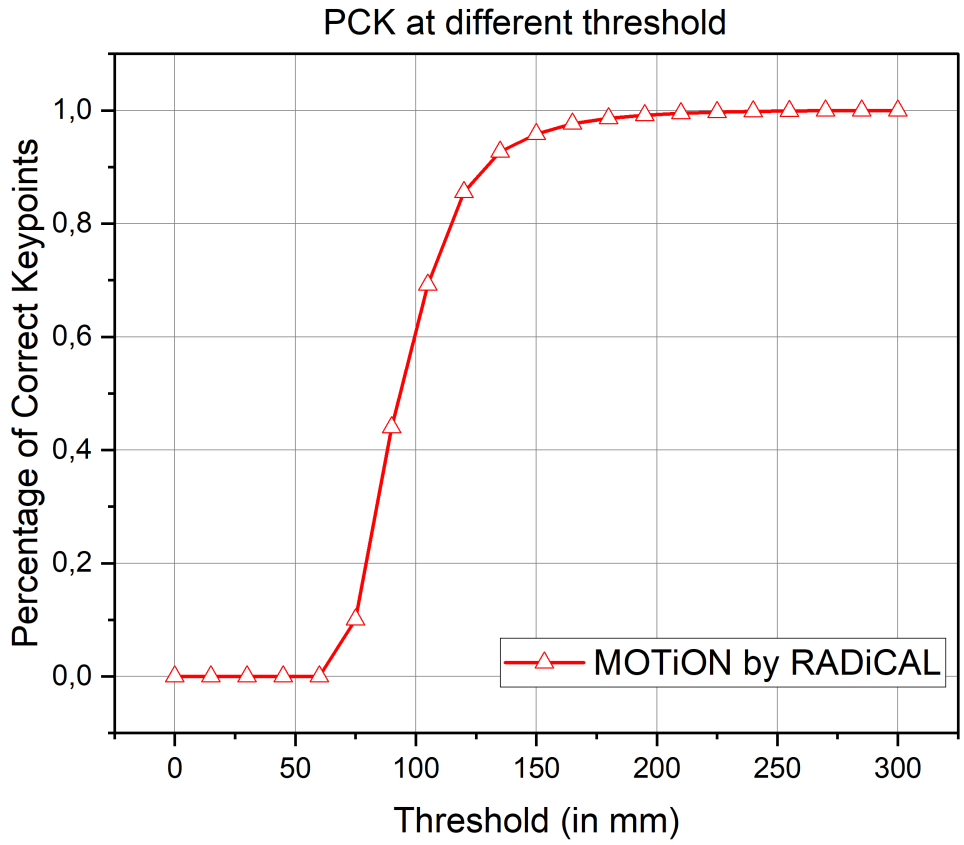


Fig. 5.6 The thresholds and the corresponding PCK

Table 5.4 MPJPE by gender

	Female	**Male**
MPJPE (mm)	95.7	100.71
Standard deviation (mm)	53.06	51.31
PCK (%)	95.9	95.7

identical. The PCK of both genders was almost the same, with 95.7%. As shown in Figure 5.8, comparing all joints by gender reveals a lower mean error for eight male joints.

5.3.2 Qualitative Evaluation

One challenging frame from each scenario was selected, and RADiCAL output was visually compared to the ground-truth frame. As demonstrated in Fig. 12, the scenarios such as "Animal Behaviours," "Daily Activities," "Reactions," "Dancing," "Jumping," "Home

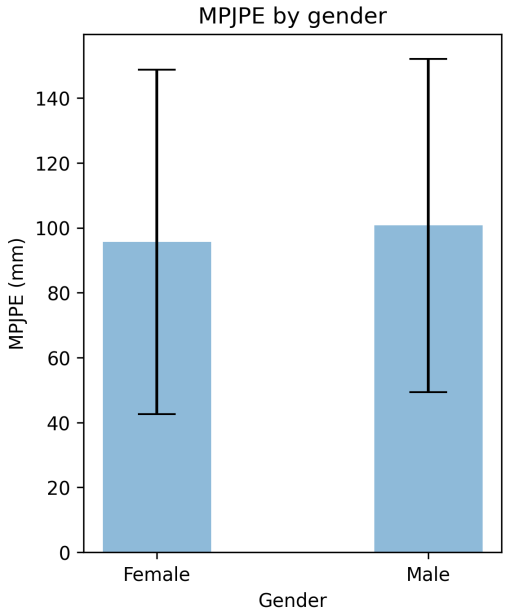


Fig. 5.7 The results of MPJPE by gender.

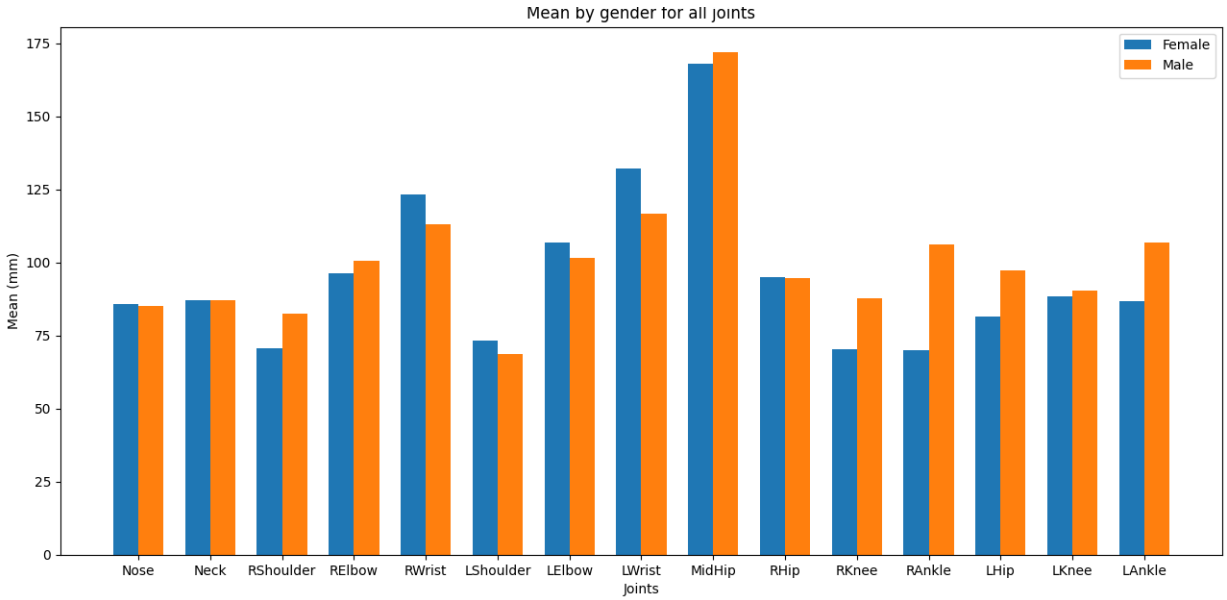


Fig. 5.8 The MPJPE results of comparison of joints by gender.

Activities," and "Walking" imitate the ground-truth quite accurately. In addition, all skeletal parts are in their proper locations. In the remaining instances, RADiCAL correctly estimated all skeleton parts. However, its orientation was incorrect. The "Working" scenario was estimated correctly, except for the head in several frames. Some parts of the "Sports" scenario, such as the hand, were not precisely estimated.

5.4 Discussion

This study provides the first comprehensive evaluation of MOTiON by RADiCAL, a cloud-based 3D human motion estimation tool, revealing both its potential advantages and inherent limitations. In scenarios involving less complex actions or slower movements, such as walking or light exercise routines, our evaluation demonstrated a relatively low Mean Per Joint Position Error (MPJPE) and a high Percentage of Correct Keypoints (PCK), signaling promising performance. These results indicate that MOTiON by RADiCAL can serve as an effective tool for animating 3D characters in VR applications, especially for intangible cultural heritage animations. For instance, traditional dances, performances, and daily routines can be effectively captured and animated using this tool, thus contributing to the preservation and promotion of cultural heritage in immersive VR experiences.

The distinction in performance directly influences the range of applications suitable for MOTiON by RADiCAL. For digital content creation fields such as simple animation for games, films, or casual fitness tracking where millimeter-level precision may not be paramount, the tool's cost-effectiveness and accessibility offer substantial benefits. This can help automate the animation process and democratize access to advanced animation techniques, allowing independent creators, small studios, and hobbyists to produce high-quality content without the need for expensive traditional motion capture systems.

However, for applications demanding high-precision motion capture, such as advanced biomechanical analysis, sports performance analysis, or precise VR interaction, the current version of MOTiON by RADiCAL may not provide the necessary accuracy. The MPJPE of 98mm found in our study, while acceptable in some contexts, could lead to significant errors in these precision-demanding applications. This limitation needs to be addressed to extend the tool's applicability to more complex and dynamic scenarios, such as sports movements or actions involving occlusion.

To further enhance the platform and democratize its use, several recommendations are proposed:

1. **Improving Occlusion Handling:** Enhancing the tool's capability to accurately capture poses in scenarios with significant occlusions will make it more reliable for a wider range of applications.

2. **Dynamic Motion Capture:** Developing algorithms that better handle rapid and complex movements can improve the tool's performance in sports and other high-energy activities.

3. **User-Friendly Calibration:** Implementing more intuitive calibration processes can help users achieve better accuracy without extensive technical knowledge.

4. **Enhanced Data Integration:** Facilitating seamless integration with other animation and VR development tools can streamline the workflow for creators and developers.

5.5 Threats to Validity

Addressing threats to validity is crucial for ensuring the robustness and reliability of research findings. According to Wohlin et al. (2012), a structured approach to identifying and mitigating these threats enhances the credibility and applicability of the study's conclusions. In this section, we discuss the potential threats to validity [239]:

5.5.1 Internal validity

In this research, potential threats to internal validity include the accuracy of the CMU dataset as the ground truth and the manual synchronization process. Any errors or inconsistencies in these areas could impact the results. To mitigate these threats, rigorous data cleaning and precise synchronization methods were employed to ensure the dataset's reliability.

5.5.2 External Validity

While the CMU dataset provides a diverse range of scenarios, it may not encompass all possible real-world movements, particularly those specific to certain cultural heritage activities. Furthermore, the findings might not be fully applicable to other HPE tools or different motion capture systems. To address this, future studies should consider expanding the range of scenarios and validating the results across multiple datasets and tools to enhance external validity.

5.5.3 Construct Validity

Potential threats include the selection of evaluation metrics (MPJPE and PCK) and whether they comprehensively capture the tool's performance. Employing additional metrics and qualitative assessments can provide a more holistic evaluation of the tool's capabilities, thereby strengthening construct validity.

5.5.4 Conclusion Validity

Potential threats include statistical errors and biases in data interpretation. To mitigate these threats, robust statistical analysis techniques were used, and efforts were made to avoid overgeneralization of the results. Clearly defining the scope and limitations of the conclusions enhances the reliability and applicability of the study's findings.

5.6 Conclusion

MOTiON by RADiCAL, as evaluated in this study, shows promise as a cost-effective, user-friendly alternative to traditional sensor-based motion capture systems. However, the tool's current performance suggests its best fit for applications where absolute precision is not a critical requirement. In realms like basic animation for gaming, motion-guided user interface design, or casual fitness tracking, the tool's slight inaccuracies are unlikely to substantially impact the result, making it a beneficial tool. Its cost and usability advantages are particularly beneficial for independent creators, small studios, or hobbyists in these fields.

However, in precision-critical applications, such as advanced biomechanical research, sports performance analysis, or high-end VR systems that require nuanced interaction, the existing error levels in MOTiON by RADiCAL may be prohibitive. For these applications, traditional sensor-based systems, despite their higher cost and complexity, may remain the gold standard.

In summary, MOTiON by RADiCAL represents a significant step forward in democratizing access to 3D human motion estimation. The tool can play a vital role in preserving and promoting intangible cultural heritage through VR applications by automating the animation of traditional dances, performances, and daily activities, thus making cultural experiences more immersive and accessible. However, its current performance limitations suggest that it is not a one-size-fits-all solution. Future research and development should focus on improving the precision of such tools to extend their applicability to a broader range of scenarios and meet the needs of more demanding applications.

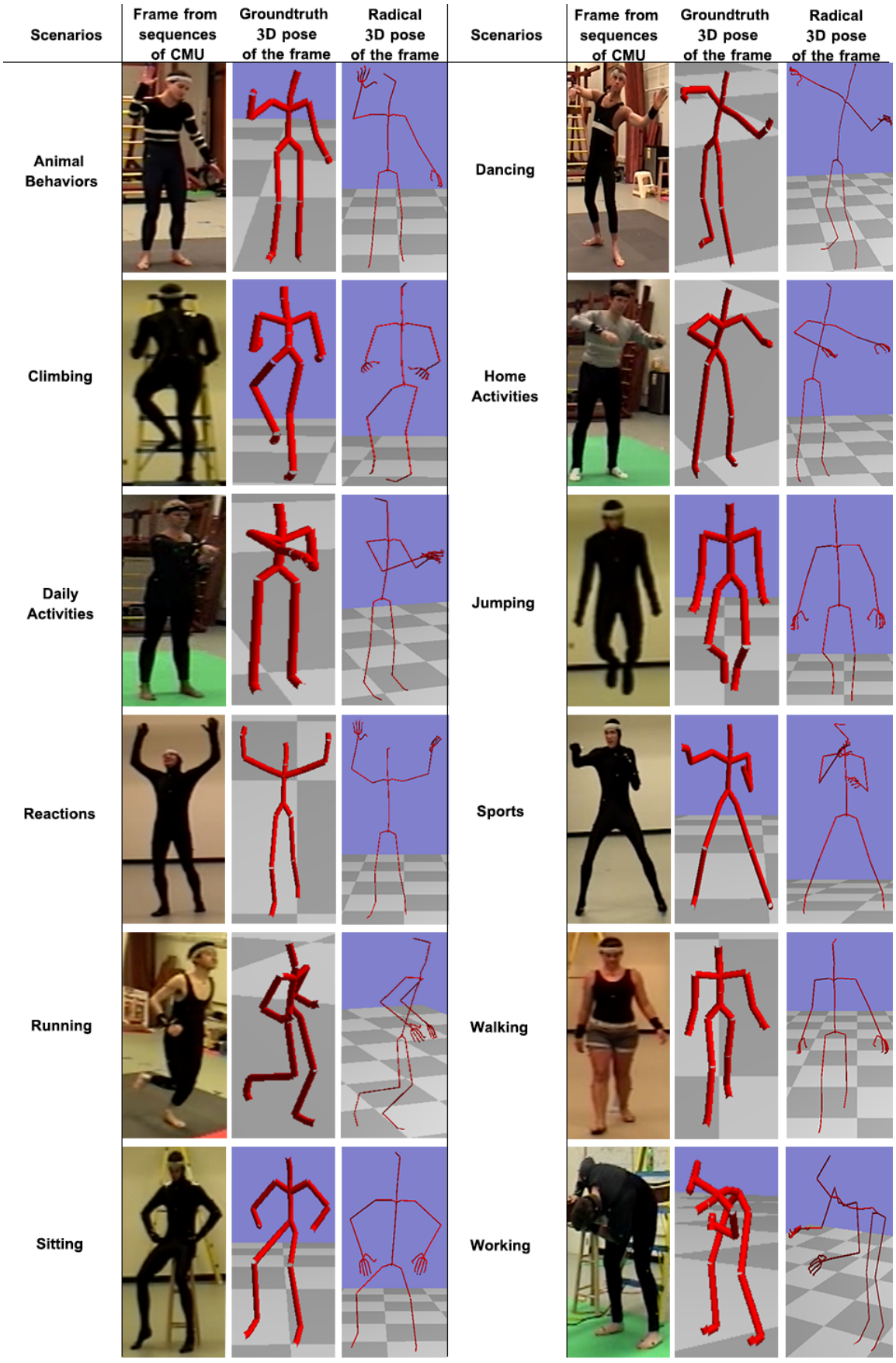


Fig. 5.9 The qualitative results of each scenario.

Chapter 6

Design, Development, and User-Centric Evaluation of El-FnaVR: An Immersive Virtual Reality Representation of Jemaa El-Fna in Marrakech for Intangible Cultural Heritage Experiences

6.1 Introduction

This chapter aims to address the potential of VR technology in enhancing the cultural experience of Jemaa El-Fna and promoting intangible cultural heritage tourism through the development and evaluation of El-FnaVR, an immersive VR application.

The primary motivation for this case study stems from the need to preserve and promote the rich cultural heritage of Jemaa El-Fna, a UNESCO-listed site recognized for its historical and cultural significance. The dynamic and sensory-rich environment of Jemaa El-Fna makes it an ideal candidate for exploring the capabilities of VR in providing an immersive and engaging cultural experience.

The goals of this case study are threefold:

Develop a VR Application: Create a VR application that accurately represents the cultural and historical essence of Jemaa El-Fna, including its unique activities, performances, and ambiance. Evaluate User Engagement and Understanding: Assess how the VR application influences user engagement and understanding of the cultural heritage and history of Jemaa El-Fna through a comprehensive user study. Analyze Practical Implications: Examine the

practical implications of using VR for cultural heritage tourism, focusing on its potential benefits and challenges for both research and industry. The research questions guiding this study are:

1. **Tool Evaluation:** How effective is the El-FnaVR application in providing an immersive and interactive cultural experience of Jemaa El-Fna?
2. **Impact on User Engagement and Understanding:** What is the impact of the VR experience on user engagement and understanding of Jemaa El-Fna's cultural heritage compared to traditional media?
3. **Influence on Subsequent Perceptions:** How does exposure to a VR experience influence participants' reactions and perceptions when subsequently watching a related video, compared to those who only watch the video?

The chapter is organized as follows: The next section will provide a literature review on the use of VR in intangible cultural heritage preservation and promotion. Then, the methodology section will describe the design and development of the VR application and the user study. The results and discussion section will present the findings of the user study and the potential implications of the research. Finally, the conclusion will summarize the main findings and provide recommendations for future research.

6.2 CASE STUDY: Jemaa El-Fna

6.2.1 Historical Background

Jemaa El-Fna, a vibrant square situated in the heart of Marrakech, Morocco as depicted in Figure 6.1, has an intricate historical background that embodies the living pulse of Moroccan heritage. The square, whose name roughly translates to "Assembly of the Dead," is nothing short of an emblematic symbol of Morocco's rich history and vibrant cultural life.

Originating in the 11th century under the rule of the Almoravids, Jemaa El-Fna was initially conceived as a meeting point for traders, nomads, and the like. Throughout its existence, it has witnessed significant transformations reflective of Morocco's sociopolitical changes, enhancing its historical significance.

As the city expanded, the square grew into an essential hub of cultural, commercial, and spiritual activities. By day, the Jemaa El-Fna square is populated by snake charmers, fruit vendors, traditional healers, and storytellers; come twilight, it transforms into an enormous open-air restaurant filled with food stalls, accompanied by music and dance performances, embodying the diverse sensory experience that characterizes Moroccan culture.

Moreover, the square has been a crucial site for the oral transmission of stories, historical events, and traditional customs, effectively preserving the intangible cultural heritage of the region. As a testament to its significance, Jemaa El-Fna was recognized as a Masterpiece of the Oral and Intangible Heritage of Humanity by UNESCO in 2009, emphasizing its importance as a living museum of Moroccan tradition and heritage.

The inclusion of digital and VR technologies in the depiction and preservation of the Jemaa El-Fna square presents a unique opportunity to capture its dynamic atmosphere and cultural vibrancy. Such technological advances can offer a novel perspective, enhancing the understanding of the square's historical background while fostering a deeper appreciation of Morocco's rich cultural tapestry.



Fig. 6.1 Jemaa El-Fna, Marrakech, Morocco.

6.2.2 Activities

Halqa

Traditional storytelling circles featuring master storytellers, snake charmers as shown in Figure 6.2a, and monkey performers, as depicted in Figure 6.2b.

These performances and activities form an intrinsic part of the Halqa experience, contributing to the vibrant and dynamic character of Jemaa El-Fna. They serve not only as sources of entertainment but also as living records of Moroccan history, culture, and societal norms.

Gnawa

Gnawa, UNESCO-listed[1] in 2019, another integral part of Jemaa El-Fna, is a traditional music and dance performances rooted in the sub-Saharan African heritage of Morocco.

[1] https://ich.unesco.org/en/RL/gnawa-01170

The Gnawa performances in Jemaa El-Fna are a sight to behold, drawing both locals and tourists into their rhythmic swirl as shown in Figure 6.2d. The melodic harmony of the guembri, the sharp clatter of the Krakebs, the vibrant colors, and the swirling dancers make it an unforgettable experience.

Traditional Foods

From dusk until late into the night, Jemaa El-Fna becomes a bustling open-air restaurant with stalls offering a variety of traditional Moroccan dishes. The air is filled with the aroma of grilling meats, simmering tagines, and exotic spices, creating a vibrant sensory experience.

Essential dishes such as couscous, Harira, and various tagines are abundant. Street vendors serve grilled meats like Kefta, Merguez, and Brochettes, as shown in Figure 6.2f. Adventurous food enthusiasts can try snails in spicy broth or sheep's head, while a variety of freshly caught seafood is also available.

For those with a sweet tooth, options include Pastilla and freshly squeezed juices from local fruits like oranges (Figure 6.2e), pomegranates, and dates. The communal atmosphere of Jemaa El-Fna's food stalls fosters interactions between locals and visitors, encouraging cultural exchange and conversations over shared meals.

Horse Carriage

As shown in Figure 6.2c, the horse carriages are an integral part of the bustling atmosphere at Jemaa El-Fna, offering both locals and tourists an alternative way to explore the square and its surroundings. Traditionally adorned and guided by experienced carriage drivers, these horse-drawn vehicles add a touch of historic charm to the modern vibrancy of the square.

6.3 El-FnaVR: Design and Implementation

The development of El-FnaVR unfolds across five distinct stages summarized in Figure 6.3, with the first being bifurcated. Stage 1.A focuses on identifying the salient features of Jamaa El-Fna for accurate representation, while Stage 1.B scours relevant literature to pinpoint suitable tools for design and execution. In Stage 2, the earmarked features and characteristics are methodically designed to facilitate an engaging virtual tour. Stage 3 involves creating essential assets such as 3D models and animated characters. This phase undergoes iterative self-reviews to ensure authenticity and accuracy. Upon validation, Stage 4 ensues, centering on prototype development. This involves integrating the validated assets and scripting interactions to enable smooth navigation throughout the virtual environment. The



Fig. 6.2 Some Jemaa El-Fna activities and traditional street foods. (a): Halqa of snakes from Jemaa El Fna square. (b): Halqa of monkeys from Jemaa El Fna square. (c): Horse carriage. (d): Gnawa dancers group performing traditional dance and songs. (e): Orange Jus at the left. (f): The grilled meats and Harira at right.

prototype undergoes rigorous self-review until it meets the desired benchmarks. Concluding the approach, Stage 5 employs a comprehensive evaluation strategy, blending both subjective and objective assessment methods, to gauge the prototype's efficacy.

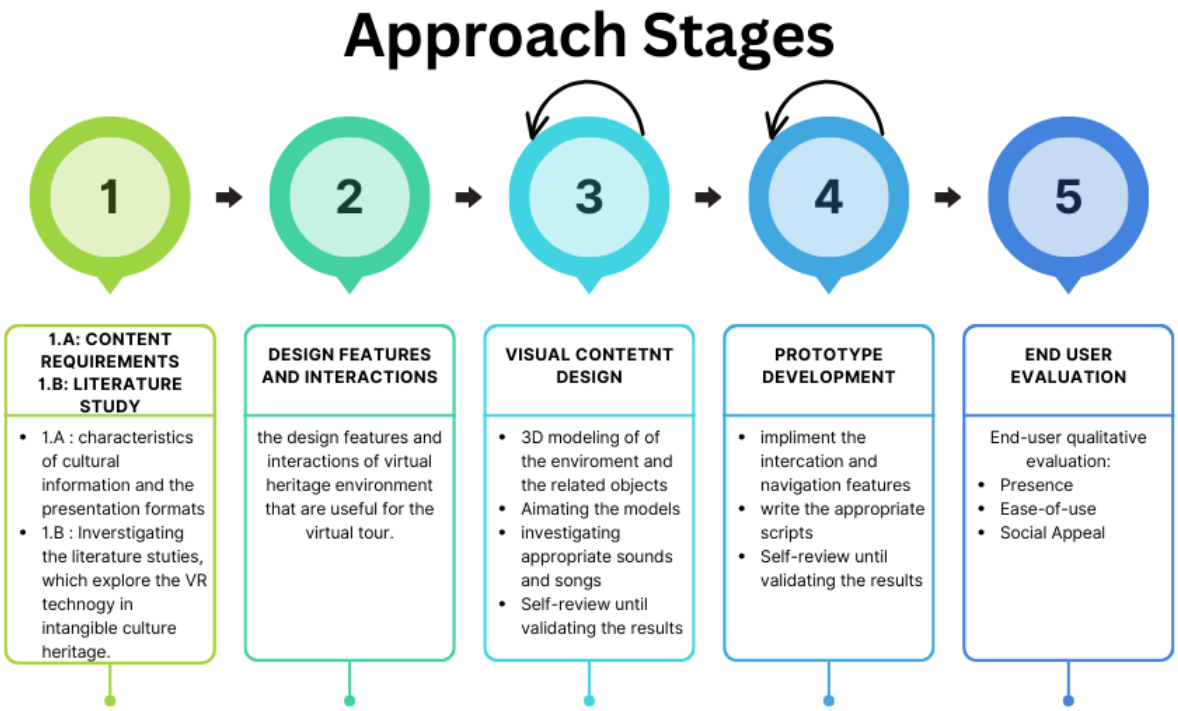


Fig. 6.3 Sequential Stages of El-FnaVR Design and Implementation approach.

6.3.1 Stage 1: Content Requirements & Literature Study

The initial stage of our approach focuses on the requirements of the design features and characteristics of the VR application, subdivided into two key areas. The second stage is used to design the features and characteristics.

Stage 1.A:

It revolves around the characteristics of cultural information and the presentation formats. Here, the essence lies in capturing the richness and diversity of Jamaa El-Fna's cultural elements and transforming them into a digital format that is accurate, immersive, and interactive. By doing so, we can provide an authentic and engaging user experience. From the traditional Gnawa music to the vibrant food stalls and captivating Halqas, we aim to encapsulate all facets of this vibrant square. The presentation formats are carefully chosen to facilitate this immersive experience. The environment is designed using realistic 3D models and animations, complemented with spatial audio to render a 360-degree sensory experience.

Stage 1.B:

We delve into the literature to investigate studies that have explored VR technology's application in the preservation and dissemination of intangible cultural heritage. This research provides critical insights into best practices, challenges, and potential solutions when developing VR applications for cultural heritage. Additionally, it aids in understanding how VR can enhance accessibility, immersion, and interactive learning experiences for users, enabling them to connect with the cultural content on a deeper level. These insights guide us in designing a VR application that not only replicates Jamaa El-Fna in a digital format but also provides an enriching and meaningful cultural experience for users.

6.3.2 Stage 2: Design Features and Characteristics

In the second stage of our development approach, we focus on the detailed design of features and characteristics that make our Jamaa El-Fna VR application an authentic representation of the cultural heritage site. This stage is all about bringing the richness and diversity of Jamaa El-Fna into a VR environment and ensuring the accuracy and immersion of the cultural information presented. This involves careful consideration of the aesthetic aspects, interactive elements, and the overall sensory experience for the user.

El-FenaVR functions

The following functions describe the features that have been designed to achieve these objectives.

- **Function 1** - Aesthetic Design: Environment
  Implements the visual aesthetics of the environment, reflecting the color palette and Islamic design principles of Marrakech. Also, the red color of the buildings, which indicates the city.
- **Function 2** - Activity Design: Snakes Halqa
  Introduces the Halqa of snakes, where users can interactively pick up animated snakes that slither around the Halqa.
- **Function 3** - Activity Design: Monkies Halqa
  Adds the Halqa of monkeys, where users can observe animated monkeys walking around the Halqa.
- **Function 4** - Activity Design: Gnawa Groups Incorporates Gnawa music, enabling users to experience virtual Gnawa groups performing traditional songs.

- **Function 5** - Activity Design: Traditional Foods
  Introduces the aspect of gastronomy with virtual food stalls serving barbeque, Tajine, and couscous.
- **Function 6** - Activity Design: Horse Carriage
  Integrates horse carriages into the environment, which users can interact with and drive.
- **Function 7** - Activity Design: Everyday Activities
  Creates an immersive atmosphere by adding virtual people performing everyday activities like walking, talking, watching Halqa, Gnawa, and food stalls.
- **Function 8** - User Interaction
  Provides users with a range of interactive capabilities, such as navigating the square, observing activities, and picking up objects.
- **Function 9** - Points of Interest
  Implements informative points of interest related to each activity and their locations.
- **Function 10** – Spatial 3D Audio Design
  Establishes an immersive 3D soundscape that corresponds with every activity in the environment, from the hissing of snakes to the rhythm of Gnawa music, and even the ambient sound of conversation among virtual people.
- **Function 11** – Preferences
  The system empowers users to personalize their experience by selecting from a variety of hand models. To mitigate the effects of cybersickness, users have the option to switch to a teleportation-based navigation mechanism as an alternative to continuous movement.

Tour and spatial design

Crafting an optimal tour scenario is paramount to guiding visitors through all activities in an immersive sequence. The vast expanse of Jamaa El-Fna square allows tourists unrestricted movement, with no predetermined path. Most, however, often experience walking to Jamaa El-Fna by horse carriage. As depicted in Figure 6.4 — which serves as a guideline rather than a strict route — our virtual application simulates this experience, beginning with a horse carriage journey towards the square. Visitors are first introduced to the Halqa of monkeys, offering a glimpse of Lamaalem's playful interactions with them. Adjacently, the enthralling rhythms of Gnawa music beckon from the Halqa of the Gnawa group. From there,

an adventurous stride leads to the snake Halqa, where audacious visitors can handle snakes and observe Lamaalem's deft maneuvers with them. A quintessential part of the Jamaa El-Fna experience is indulging in traditional street foods and refreshing orange juice, so the tour naturally progresses to stalls offering these delights. To conclude, visitors can immerse themselves in the vibrant atmosphere of three additional Gnawa Halqas, each presenting a unique medley of songs and music.

Application structure

The design of the El-FnaVR application is centered on user interaction and navigation. This structure is layered, integrating physical and virtual components to enhance user control and engagement within the virtual environment, as depicted in Figure 6.5. The first layer, the physical, involves the controllers and sensors of the Oculus Quest headset. These devices enable users to navigate, control, and interact within the virtual setting. The second layer is represented by a virtual guide. This guide helps users navigate the virtual space by providing points of interest, coins, and storytelling, thus enhancing the localization of each activity within the virtual environment. The third layer, the most comprehensive, constitutes the entire 3D virtual environment. It encompasses all 3D characters, objects, spatial 3D audio, and effects, immersing users in a rich, detailed representation of the cultural setting.

6.3.3 Stage 3: Visual content design

In the third stage of development, our focus shifts to the Visual Content Design (Figure 6.6), which forms the core of the VR experience. This stage ensures that the immersive digital environment accurately represents the authentic spirit and visual richness of Jemaa El-Fna. The stage involves the following steps:

1. **3D Modeling of the Environment and Related Objects:** This step involves creating a 3D model of Jemaa El-Fna and all associated objects like the Halqa, Gnawa musicians, and local food stalls. This involves detailed architectural modeling of the square and careful sculpting of related elements, using high-quality 3D graphics to accurately represent the physical appearance of the cultural site and its elements. To achieve this goal, AutoDesk MAYA[2] was chosen for 3D modeling and Pixologic ZBrush[3] for the sculpting process.

2. **Animating the Models:** Static models do not capture the vibrant life of Jemaa El-Fna. Therefore, we bring our 3D models to life through animation. Snakes slither, monkeys

[2]https://www.autodesk.com/products/maya
[3]https://www.maxon.net/en/zbrush

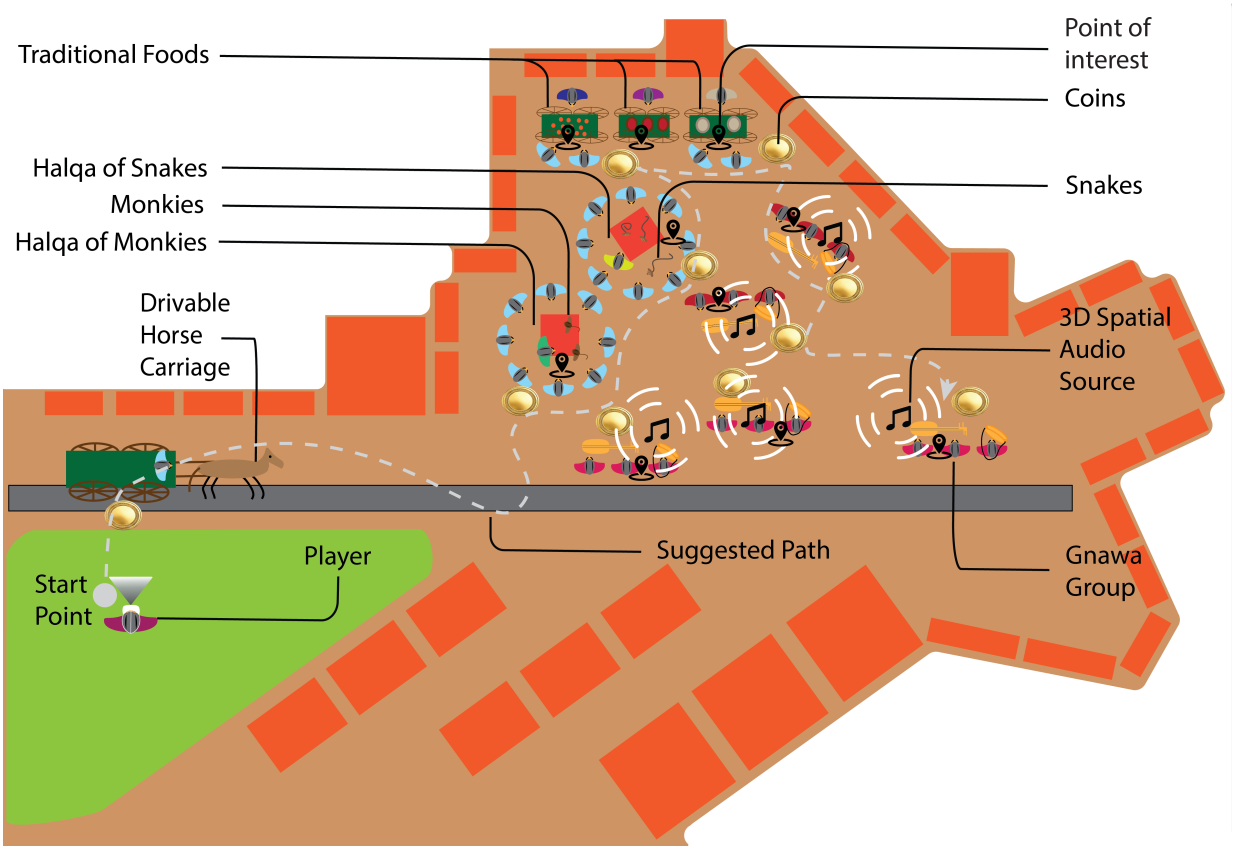


Fig. 6.4 Jamaa El-Fna mapping and tour design.

wander, Gnawa musicians perform, and virtual inhabitants carry out their activities, lending dynamism and realism to the experience. Autodesk Maya was used to animate virtual objects.

3. **Investigating Appropriate Sounds and Songs:** The audio design is a critical element of the immersive experience. We carefully research and curate a selection of appropriate ambient audios, musical performances, and sound effects, ensuring they are culturally accurate and contribute to the overall authenticity of the environment.

4. **Self-Review and Validation of Results:** Lastly, the team conducts iterative reviews of the designed content, verifying the accuracy and quality of the visual and auditory elements. This step is crucial to ensure that the VR representation remains faithful to the cultural richness and diversity of the real Jemaa El-Fna, providing users with an immersive and authentic experience.

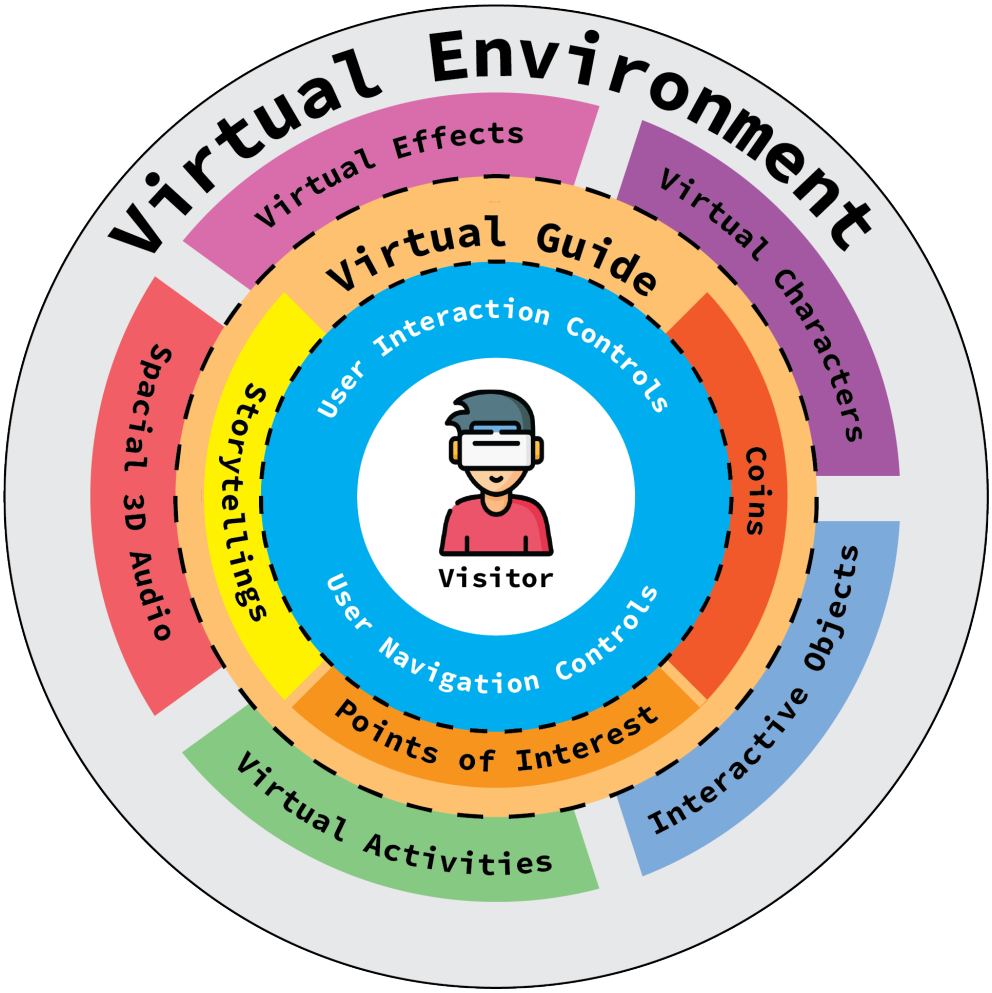


Fig. 6.5 The layers-based application structure.

6.3.4 Stage 4: Prototype implementation

The fourth stage, Prototype Implementation, delves into the actual construction of our VR application, bringing the design features and visual content to life. This stage is split into two main components.

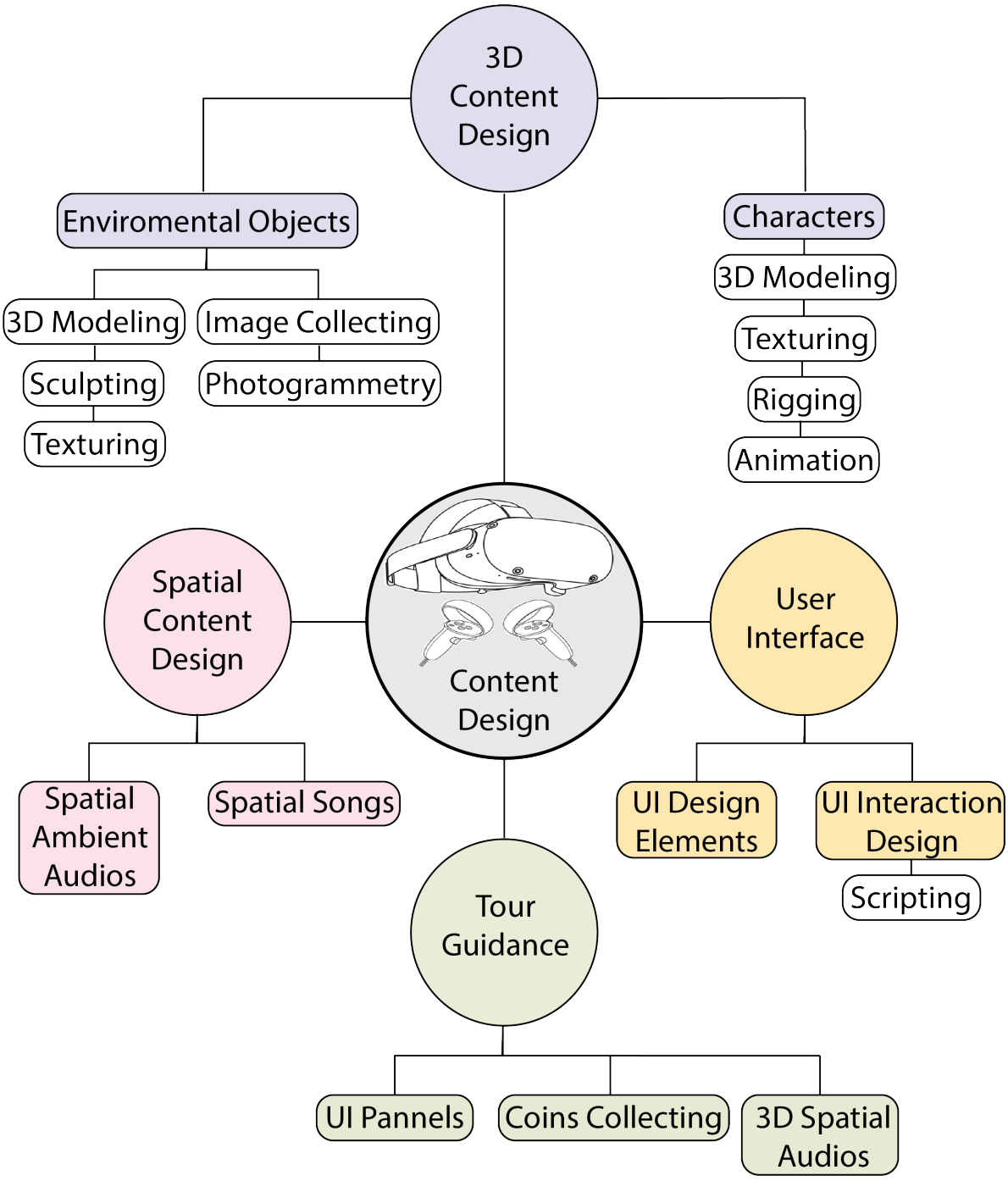


Fig. 6.6 Visual Content Design of El-FnaVR.

Development pipeline

The VR application is developed using the Unity[4] 3D game engine and the VR Interaction Framework. It is specifically designed to be compatible with the Oculus Quest 2 and later

[4]https://unity.com/

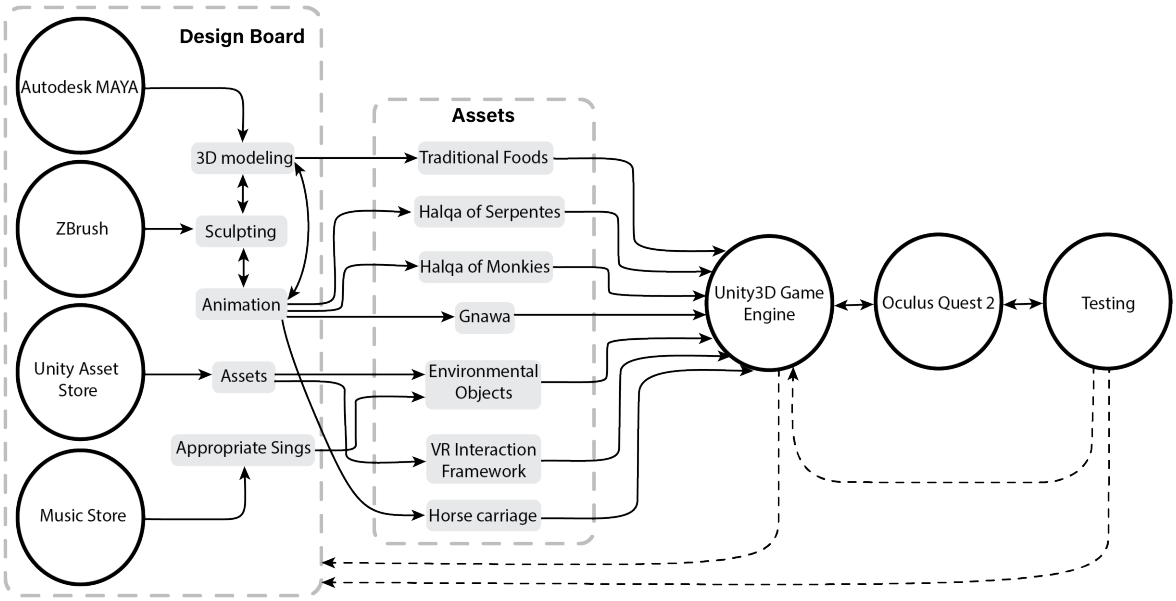


Fig. 6.7 Development pipeline of El-FnaVR.

models. A critical aspect of this phase is the implementation of interaction and navigation features, which are essential for user engagement and the overall user experience within the VR environment. This includes coding the necessary scripts that drive these features and constantly reviewing and adjusting until satisfactory results are achieved. Such a systematic approach ensures that the application is not only immersive but also interactive and responsive to user input, providing an enriching virtual experience. The development pipeline is summarized in Figure 6.7.

Virtual scenario

The system scenario was designed to be adaptable, as depicted in Figure 6.10. The system provides users with the freedom to navigate through the virtual space and jump from one scene to another at their own pace. Such flexibility not only increases the level of user engagement but also enhances their ability to learn and enjoy the tour. To aid navigation, "points of interest" markers have been placed to localize each activity, as demonstrated in Figure 6.8a. Furthermore, to encourage exploration, 10 coins have been distributed next to each activity or point of interest for users to collect, as shown in Figure 6.8b. To help reduce any potential anxiety from continuous movement, an alternative teleportation movement mechanism has been implemented [34], as depicted in Figure 6.9. This allows users to move to different areas quickly, providing them with a seamless and comfortable VR experience.



Fig. 6.8 Aid-navigation tools. (a): a point of interest placed beside the Halqa of snakes. (b): collectible coin.

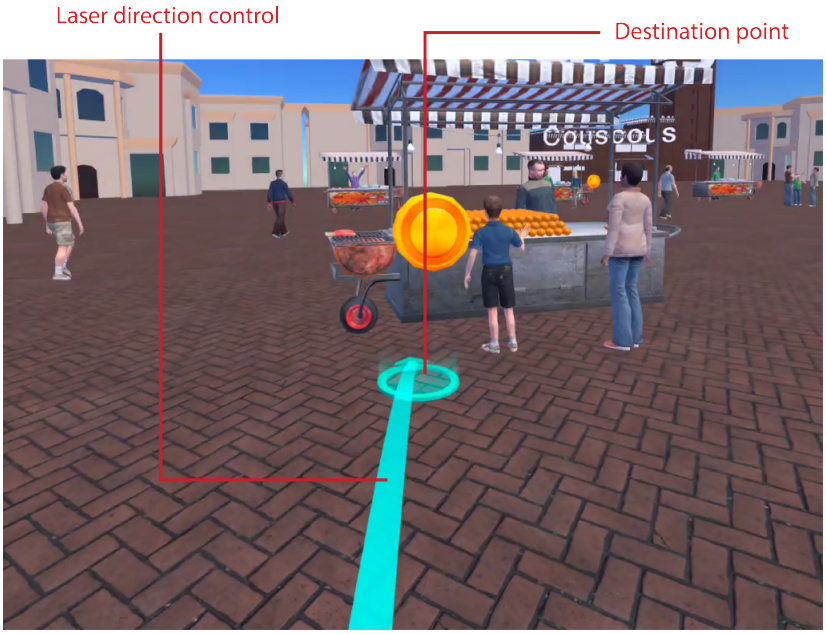


Fig. 6.9 Movement using Point & Teleport Locomotion Technique principle [34].

As outlined in Function 11, El-FnaVR offers users the ability to customize their experience by selecting their preferred hand model within the virtual environment. This feature allows for a more personalized and immersive interaction, tailoring the virtual world to individual tastes. Figure 6.11 demonstrates this functionality.

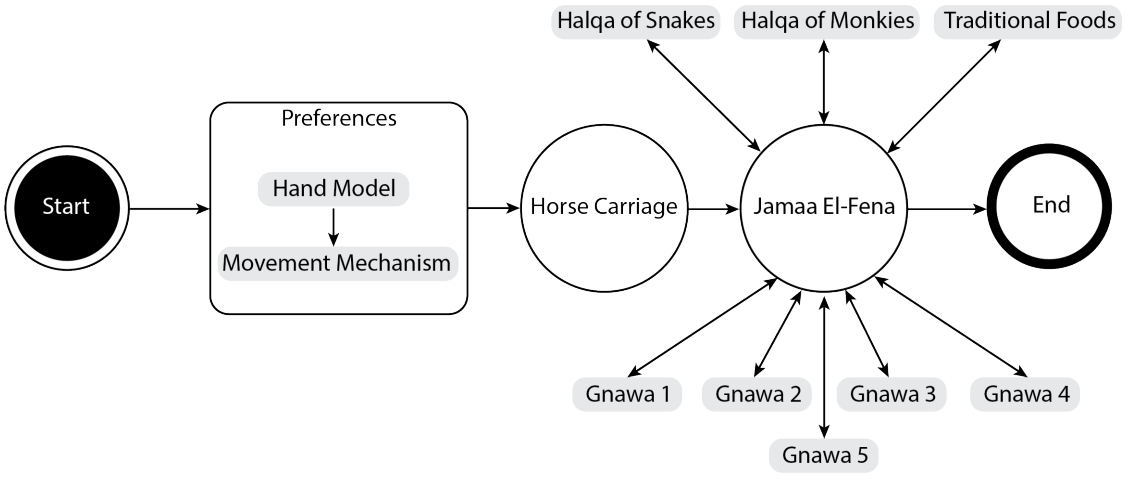


Fig. 6.10 El-FnaVR system scenario.

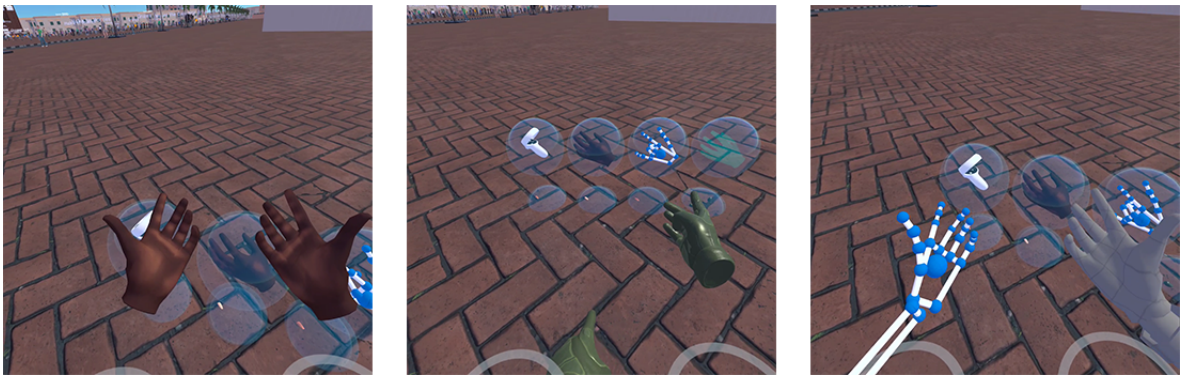


Fig. 6.11 Customizing hand model from suggested models.

Refined Implementation and Deployment

Following rigorous self-testing, the ultimate version of El-FnaVR has been fully developed and is now primed for end-user evaluation. The application meticulously replicates the ambiance of Jemaa El-Fna with a high degree of detail, allowing users to interact with a multitude of objects and select their desired tour scenario through points of interest. Figure 6.13 showcases the most prominent activities implemented in the application.

6.3.5 Stage 5: End user evaluation

To comprehensively evaluate our VR rendition of Jamaa El-Fna, we utilized a dual-pronged methodological approach that combines the analytical rigor of the Unified Theory of Acceptance and Use of Technology (UTAUT) model with a more dynamic, task-based objective evaluation technique.



Fig. 6.12 Participants during the El-FnaVR experience.

Subjective Approach

At the heart of the subjective assessment was the UTAUT model, a potent tool employed to gauge user receptivity, behavior, and engagement levels with technology. Our participants, clustered into two distinct groups, interfaced with the digital reality of Jamaa El-Fna through different lenses.

Group One, composed of 42 participants as shown in Figure 6.12, was introduced to the bustling milieu of Jamaa El-Fna through the Oculus Quest 2 VR headset. Once they had navigated the VR environment, they were asked to reflect on their experience by completing the UTAUT questionnaire. This introspective exercise nudged them to contemplate the VR application's performance expectancy, the anticipated ease of use (effort expectancy), the social dynamics influencing their interaction (social influence), and the facilitating conditions, which encapsulate the belief in the necessary support to use the system. Once their VR journey concluded, they embarked on a parallel exploration using traditional mediums (YouTube videos, and other applications), followed by the completion of a second UTAUT questionnaire.

Group Two, a pool of 32 participants, exclusively interacted with Jamaa El-Fna through these traditional mediums and offered their insights via the UTAUT questionnaire. This bifurcated approach provided a comparative perspective on user experiences, unveiling a nuanced understanding of the VR application's utility and effectiveness.

Our study revolves around three core hypotheses, each investigating different aspects of VR's impact on participants' subsequent video-viewing experiences:

- **Usage Difference Hypothesis:** There is a significant difference between participants' responses when they watch a video after VR exposure compared to when they watch a video alone. This hypothesis is examined using Cramer's V and a Chi-Square test, which allow us to quantify and compare the distinct experiences of VR and standard video users.
- **Experience Effect Hypothesis:** Exposure to VR prior to viewing a video significantly alters participants' subsequent video viewing experience, as measured through their responses. This hypothesis probes the potential of VR to enrich users' engagements with similar content in a non-VR format.
- **Association Hypothesis:** There is a notable association between the type of video viewing condition (VR followed by video or video only) and participants' responses. This hypothesis explores whether the viewing condition can influence user responses.

Objective Approach

Augmenting our subjective inquiry was an innovative, task-based evaluation method. Users journeyed through the vibrant alleys of VR Jamaa El-Fna, undertaking a treasure hunt of sorts, where the treasure was virtual coins peppered around various locations. Nestled next to engaging activities like Halqa or a Gnaoua performance, each coin became a beacon, urging users to explore and interact more deeply with the virtual environment. The coin collection served as an objective indicator of user engagement, their navigation skills, and the overall usability and immersion of the VR application.

6.4 Results

6.4.1 Subjective Evaluation Results

The study engaged 43 participants for the VR First and Video After VR experiences and 32 participants for the Video Only group, with an approximately equal gender distribution in the former groups and a slight male predominance in the latter (as shown in Table 1). Predominantly aged 18-24 and skewed towards students, the participant pool also included a few individuals from the <18 and 25-34 age brackets across all groups. Furthermore, while participants' educational backgrounds were diverse, most had at least graduated high school

Table 6.1 The statistics of the demographic data for the participants in the study.

Questions	Responses	VR First	Video after VR	Video Only
Gender	Female	20	20	14
	Male	23	23	18
Age	<18	3	20	14
	18-24	36	36	22
	25-34	3	3	4
	35-44	0	0	0
	45-54	0	0	0
	55-64	0	0	0
	>65	0	0	0
Occupation	Student	40	40	31
	Employed	3	3	1
Age	High school graduate	16	16	12
	Some college or associate degree	15	15	10
	Bachelor's degree	8	8	1
	Less than high school	3	3	9
	Master's degree	1	1	0

or undertaken some college or associate degree, with only a minimal representation from those with a master's degree, especially in the Video Only group (as depicted in Table 1).

The Cronbach's alpha was calculated to assess the internal consistency of our survey measures across the three groups: VR First, Video after VR, and Video Only. Notably, all the measures indicated satisfactory to excellent reliability. For the Performance Expectancy (PE) measure, Cronbach's alpha was 0.94, 0.96, and 0.97 for the VR First, Video after VR, and Video Only groups respectively, demonstrating a high level of consistency. The Effort Expectancy (EE) measure yielded reliability coefficients of 0.74, 0.76, and 0.85, reflecting acceptable consistency across all groups. Similarly, the Social Influence (SI) measure showed strong reliability with alpha values of 0.84, 0.87, and 0.87. Lastly, the Facilitating Conditions (FC) measure also indicated satisfactory consistency with alpha coefficients of 0.77, 0.83, and 0.92 across the groups. These results affirm the robustness of our survey measures across different groups and are presented in detail in Table 2.

A comparison of the Diverging Stacked Bar Charts for Likert Scales for both the VR Only and Video after VR groups, as shown in Figures 6.14 and 6.15, reveals distinctive differences in the distribution of participant responses for each construct item.

For the VR Only group, a strong positive sentiment is evident as the majority of the responses skewed towards 'Agree' and 'Strongly Agree' across all the construct items. This is

Table 6.2 Cronbach's Alpha Measures for Survey Consistency Across Different Groups.

	Cronbach alpha		
Construct	**VR First**	**Video after VR**	**Video Only**
PE	0.94	0.96	0.97
EE	0.74	0.76	0.85
SI	0.84	0.87	0.87
FC	0.77	0.83	0.92

particularly noticeable in Perceived Enjoyment (PE) and Ease of Use (EE) items. Conversely, in the Facilitating Conditions (FC) items, particularly FC1 and FC2, responses were more evenly distributed across all categories, hinting at a more neutral stance from the participants.

On the other hand, the Video after VR group displayed a wider dispersion of responses across the Likert scale. Interestingly, a considerable number of participants either 'Strongly Agreed' or 'Agreed' with the statements in PE and EE categories, despite a larger presence of 'Neutral', 'Disagree', and 'Strongly Disagree' responses compared to the VR Only group.

This observation points towards varied user perceptions and experiences within the two groups. While the VR Only group generally demonstrated a more favorable response to the VR experience, the Video after-VR group revealed mixed responses, potentially reflecting the influence of previous video experience on subsequent VR perceptions. These visual patterns, as displayed in Figure 6.10, set the stage for a deeper statistical investigation into the differential effects of VR-only and video-then-VR exposure on user perceptions.

The descriptive statistics of each construct across the three groups; VR First, Video after VR, and Video Only, including the mean and standard deviation, were explored (as shown in Table 3). The means for the VR First group ranged from 3.23 to 4.51, indicating generally high values. The Video after the VR group's means varied between 3.18 and 3.83, which is slightly lower than those of the VR First group. The means for the Video Only group ranged from 3.21 to 3.81, slightly higher than the Video after VR group but still less than the VR First group. Standard deviations for all groups across constructs, which indicate the dispersion of responses, ranged from 0.82 to 1.54 for VR First, 1.02 to 1.38 for Video after VR, and 1.09 to 1.38 for Video Only. The dispersion was more prominent in the VR First group than the other two, implying responses that are more varied. Overall, this data suggests that there might be differences between the three groups in terms of perceived enjoyment, ease of use, social influence, and facilitating conditions. However, these potential differences, suggested by the variability in means and standard deviations across constructs and groups, will be further examined in the forthcoming inferential statistical analyses.

Table 6.3 Statistical results of the models as a function of CPs number

	Mean			**SD**		
Item	**VR First**	**Video after VR**	**Video Only**	**VR First**	**Video after VR**	**Video Only**
PE.1	4.51	3.39	3.68	0.82	1.38	1.22
PE.2	4.23	3.32	3.81	1.04	1.35	1.20
PE.3	4.02	3.48	3.53	1.26	1.50	1.26
PE.4	4.02	3.34	3.59	1.23	1.15	1.13
PE.5	3.90	3.18	3.43	1.04	1.40	1.26
PE.6	3.90	3.32	3.53	0.94	1.22	1.19
PE.7	4.16	3.48	3.68	0.84	1.24	1.11
PE.8	4.16	3.44	3.78	1.06	1.25	1.18
EE.1	4.30	3.53	3.78	1.05	1.14	1.09
EE.2	4.37	3.37	3.31	0.84	1.15	1.22
EE.3	4.46	3.83	3.37	0.82	1.02	1.38
EE.4	3.95	3.34	3.21	1.32	1.27	1.36
EE.5	4.18	3.65	3.40	1.02	1.15	1.18
SI.1	4.27	3.65	3.53	1.07	1.28	1.29
SI.2	4.32	3.62	3.5	1.01	1.19	1.34
SI.3	4.32	3.69	3.37	0.83	1.12	1.15
FC.1	3.51	3.41	3.43	1.50	1.38	1.38
FC.2	3.23	3.25	3.5	1.54	1.31	1.36
FC.3	3.58	3.23	3.43	1.13	1.08	1.21
FC.4	3.79	3.27	3.40	1.24	1.24	1.24

With respect to the chi-square test results shown in Table 4, we see notable differences across each construct (Perceived Enjoyment (PE), Ease of Use (EE), Social Influence (SI), and Facilitating Conditions (FC)). This affirms our Usage Difference Hypothesis, as significant statistical associations were found between the constructs and the three study conditions (VR First, Video after VR, Video Only). Our p-values, being less than the accepted significance level of 0.05, support these associations, particularly for EE, SI, and FC ($p < 0.00001$).

Cramer's V results, as shown in Table 4, further substantiate our Experience Effect Hypothesis. While PE demonstrated a weaker association with the study conditions (Cramer's V=0.14), EE, SI, and FC exhibited moderate to strong associations (Cramer's V values: 0.38, 0.63, and 0.53 respectively). Notably, SI presented a strong correlation, suggesting the VR experience significantly shaped participants' social influence perception.

Our Association Hypothesis is validated by these findings, as the significant associations and their respective strengths reinforce the impact of the viewing conditions on participant responses. Together, these results underscore the effectiveness of VR in delivering immersive

Table 6.4 Chi-Square Test Results and Cramer's V Values for UTAUT Constructs Across Study Conditions.

Construct	χ^2	Degree of Freedom	P-value	Cramer's V
PE	11.12	4	0.025	0.14
EE	89.29	4	<0.00001	0.38
SI	242.05	4	<0.00001	0.63
FC	157.11	4	<0.00001	0.53

Table 6.5 Summary of Hypotheses and Outcomes.

Core hypotheses	Hypotheses	Outcome
Usage Difference Hypothesis	There is a significant difference between participants' responses when they watch a video after VR exposure compared to when they watch a video alone.	Supported
Experience Effect Hypothesis	Exposure to VR prior to viewing a video significantly alters participants' subsequent video viewing experience, as measured through their responses	Supported
Association Hypothesis.	There is a notable association between the type of video viewing condition (VR followed by video or video only) and participants' responses.	Supported

experiences that shape user perceptions and interactions with intangible cultural heritage content.

6.4.2 Objective Evaluation Results

As a crucial part of this study, objective measures were incorporated to complement and confirm the subjective findings. These measures involved participants collecting virtual coins across various activities within the Jamaa El-Fna VR environment, serving as quantifiable evidence of user engagement and interaction. As shown in Table 6, the average coin collection count, a solid 8.84 out of 10 with a relatively minor standard deviation of 2.21, not only underscores the participants' effective navigation and interaction within the VR environment but also lends credence to the subjective assessments.

Such a high average suggests that users were indeed engaged and could successfully explore a significant majority of the activities provided in the VR experience, which resonates with the high Perceived Enjoyment (PE) scores from the subjective evaluations.

Table 6.6 Objective Evaluation Metrics: Mean Coin Collection in the VR Environment.

Mean	Standard Deviation
8.84	2.21

This tangible, objective evidence of interaction and engagement, combined with the subjective data, form a robust corroboration of the findings. It aligns with our expectations and bolsters the validity of the study by illustrating that users were not just stating their enjoyment and ease of use, but their actions within the virtual environment were in agreement with these sentiments. In essence, this objective evaluation acts as a form of action-based validation of the subjective self-reported data, thereby providing a more comprehensive and balanced insight into the user experience with El-FnaVR.

6.5 Discussion

The current study embarked on a path-breaking exploration into how different modes of initial exposure—Video or VR—shape the perceptions of users towards a cultural heritage tourism experience, as exemplified by the innovative El-fenaVR. This VR application, designed to replicate the vibrant and culturally rich environment of the Jamaa El-Fna square, not only signifies an innovative approach to cultural preservation but also stands as a testament to the potential of VR in creating immersive, interactive, and educational experiences for users. Despite the geographical and temporal constraints often associated with cultural heritage tourism, VR applications like El-fenaVR can offer users the opportunity to experience and learn about different cultures from the comfort of their homes.

Subjective assessments offered a profound insight into users' experiences and perceptions, shaped by constructs of perceived enjoyment, ease of use, social influence, and facilitating conditions. Perceived enjoyment and ease of use were particularly pronounced in the VR Only group, indicating that the VR environment was both enjoyable and user-friendly. Such findings reaffirm the immersive and interactive potential of VR, making it a viable medium for promoting and preserving cultural heritage.

However, the construct of Facilitating Conditions (FC) emerged as a challenge across all groups, particularly visible in the VR Only group. Given the cost and accessibility concerns related to VR technology, it's not surprising that FC had lower agreement levels. Many participants were first-time VR users and faced difficulties in using the VR headsets, which could be attributed to the lack of prior experience, the complexity of setup, or the high costs associated with owning such technology. This calls for an increased focus on user education

about VR technology and a need for more affordable and user-friendly VR systems to ensure a wider adoption and a more inclusive user experience.

Interestingly, the Video after the VR group demonstrated a wider dispersion of responses. This suggests that an initial video experience might temper the novelty effect of the VR experience, leading to a more diverse range of user perceptions. This extends our understanding of how different modes of initial exposure could influence subsequent experiences, emphasizing the role of familiarity and novelty in shaping user experiences.

Both descriptive statistics and chi-square tests revealed noteworthy differences among the three groups. The VR First group demonstrated generally higher means across all constructs, emphasizing the positive effect of initial VR exposure. Moreover, the strong correlation between social influence and the study conditions shows the significant impact of the method of exposure on perceived social acceptance of VR technology. Objective evaluations further corroborated these insights, offering not just reported enjoyment, but also tangible evidence of this enjoyment through substantial interaction within the VR environment. This action-based validation of the VR experience underlines the potential of VR as an engaging platform for cultural heritage tourism, even in the face of challenges like accessibility and ease of use.

This study is not devoid of limitations, such as the sample size and diversity, or potential individual factors like motion sickness or other physical discomforts. Future research should aim to explore these factors, replicate these findings in other cultural settings, and investigate further the impact of the sequential order of exposure on user perceptions.

In conclusion, this study sheds light on how different modes and orders of exposure to content can influence user experiences in a VR environment like El-fenaVR. In the face of the rising prominence of VR technology, especially in domains like cultural heritage tourism, these insights are invaluable for the design and delivery of more engaging and effective VR experiences. As we navigate the challenges and possibilities of this digital era, the potential for VR to revolutionize cultural heritage tourism is not just exciting, but also truly transformative.

6.6 Conclusion

In conclusion, this study reveals the compelling potential of VR as an immersive medium for cultural heritage tourism, using El-fenaVR as a case study. It demonstrated how modes and sequences of content exposure can significantly shape users' experiences while underscoring challenges related to VR accessibility and usability.

However, this research opens the door for future studies. These could include broadening participant demographics to ensure representation and exploring individual factors like

motion sickness that could affect the VR experience. The findings concerning Facilitating Conditions invite further exploration into making VR technology more affordable and accessible. It also highlights the importance of user education for better adoption and inclusivity.

A valuable next step would be to validate these findings in other cultural heritage settings, contributing to the wider application of VR in this domain. As the world digitalizes, the transformative potential of VR for cultural heritage tourism becomes increasingly exciting, promising a future where cultures can be explored without time or space constraints.



Fig. 6.13 Depiction of Activities in the El-FnaVR Experience. (a) shows the horse carriage, which visitors can virtually ride and steer. (b) features the Gnawa Group engaged in song and dance. (c) showcases animated monkeys in various poses. (d) displays the option to handle animated snakes. (e) highlights street-side orange juice vendors. (f) illustrates some traditional street food offerings.

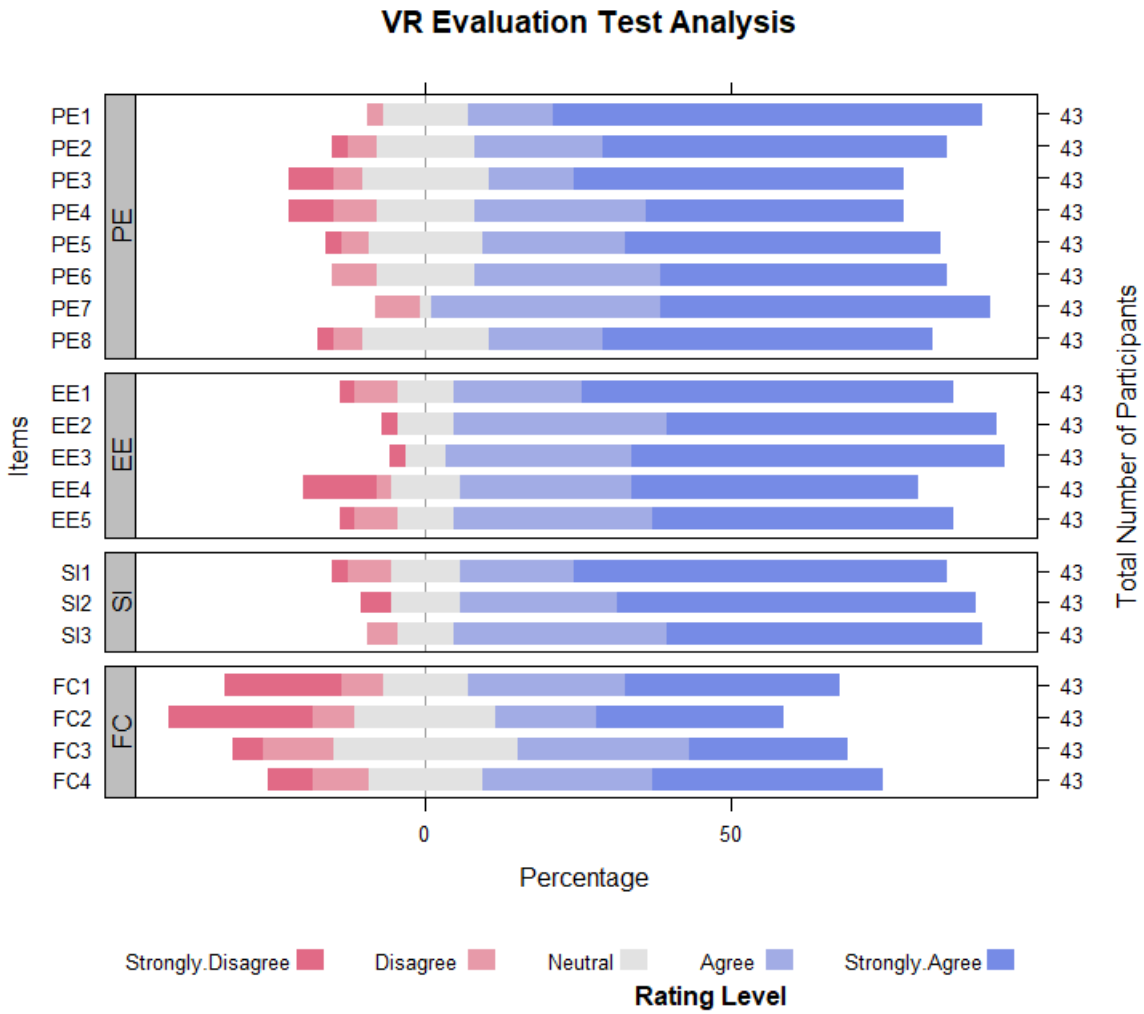


Fig. 6.14 VR evaluation test analysis.

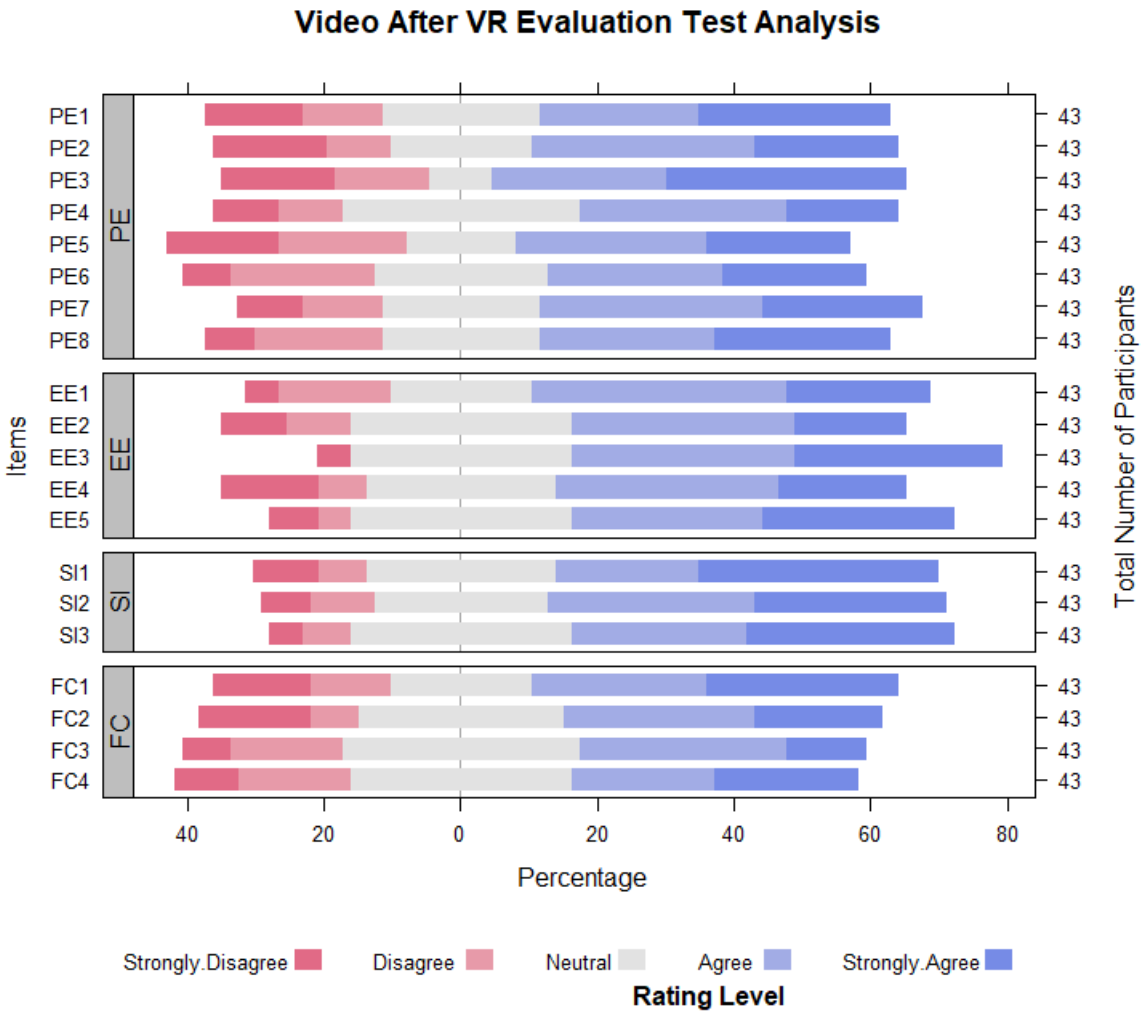


Fig. 6.15 Video after VR evaluation test analysis.

Chapter 7

Conclusion

7.1 Summary and Main Results

In this thesis, we delved into the realms of virtual reality, cloud-based 3D human pose estimation, and low-cost terrestrial photogrammetry, making notable contributions across these fields. Our work has demonstrated the power of emerging technologies to democratize access to cultural heritage preservation, motion analysis, and historical site documentation.

The research revealed that low-cost terrestrial photogrammetry, especially applied to the Marinid Royal Necropolis in Fez, can rival traditional methods in accuracy and effectiveness. This approach democratizes archaeological documentation, making it accessible and feasible for a broader audience. The findings advocate for the widespread adoption of photogrammetry, especially in resource-constrained environments, enhancing global heritage conservation efforts.

Additionally, our evaluation of MOTiON by RADiCAL has provided insights into the current landscape of cloud-based 3D human pose estimation. While showcasing the benefits of accessibility and cost-effectiveness, our analysis also highlights the limitations of these systems in high-precision-required scenarios, underscoring the need for a balanced approach tailored to specific end-use cases.

The El-FnaVR application illustrates the significant impact of VR on cultural heritage preservation, enabling immersive experiences that bridge geographical gaps, particularly in times when physical travel is restricted. This underlines VR's potential as a vital tool for educational and preservation efforts worldwide.

7.2 Implications

The implications drawn from the contributions in this thesis intersect at several critical junctures, notably in the realms of accessibility, educational outreach, and methodological innovation, reshaping our approach to cultural heritage, motion analysis, and archaeological documentation.

- **Technological Accessibility and Democratization:** The findings across the three contributions highlight a significant shift towards more accessible and cost-effective technological solutions. The El-FnaVR application demonstrates how VR can be employed beyond entertainment, serving as a vital tool for cultural preservation accessible from anywhere in the world. Similarly, the evaluation of MOTiON by RADiCAL underscores the growing accessibility of motion capture technologies, once confined to well-funded laboratories, now available to a broader audience including small studios and educational settings. The investigation into low-cost photogrammetry further emphasizes this trend, showcasing that consumer-grade cameras can perform comparably to professional equipment in documenting historical sites. This democratization of technology has profound implications for professionals and enthusiasts alike, enabling wider participation and fostering innovation in fields previously limited by technological and financial barriers.

- **Enhancement of Educational Outreach:** The immersive experience provided by the El-FnaVR application opens new avenues for educational outreach, offering a dynamic and engaging way to learn about cultural heritage remotely. This has important implications for educational curricula and cultural studies, suggesting a shift towards more interactive and immersive learning environments. Similarly, the accessibility of motion capture technology, as demonstrated by MOTiON by RADiCAL, introduces new possibilities for educational applications, from physical education to biomechanics and beyond. These technologies enable a hands-on approach to learning, enhancing student engagement and understanding.

- **Methodological Innovation in Professional Fields:** Each contribution introduces methodological innovations that challenge conventional practices in their respective fields. In cultural heritage preservation, the use of VR as demonstrated by the El-FnaVR project represents a novel approach to documenting and experiencing cultural sites, encouraging professionals to consider virtual solutions as part of their preservation strategies. In the realm of human motion analysis, the findings from the evaluation of MOTiON by RADiCAL prompt a reevaluation of how motion capture technologies

can be integrated into research, sports training, and healthcare diagnostics. Lastly, the success of using low-cost photogrammetry for archaeological documentation invites heritage professionals and archaeologists to reassess their methodologies, potentially incorporating more accessible technologies into their workflows.

7.3 Future Directions

The future directions delineated from this thesis underscore pivotal advancements and research opportunities poised to redefine the landscapes of virtual reality, motion capture, and photogrammetry. These areas, rich with potential, set the stage for transformative developments that promise to extend the reach, impact, and efficacy of these cutting-edge technologies:

- **Enhancements in Virtual Reality Experiences:** Future enhancements aim at making virtual reality experiences more personalized and interactive. This involves integrating artificial intelligence to tailor the VR content to individual users, enhancing their learning and exploration within virtual environments like El-FnaVR. Additionally, introducing features that allow for shared experiences can transform solitary VR journeys into collective explorations, enabling groups to simultaneously engage with and learn from cultural heritage sites in a communal virtual space.
- **Advancements in Cloud-based 3D Human Pose Estimation:** Improvements in this domain focus on refining the accuracy and functionality of motion capture technologies. By employing more sophisticated machine learning models and deep learning frameworks, future developments aim to enhance the precision of cloud-based 3D human pose estimation, making it suitable for a wider range of applications, from professional sports analysis to detailed biomechanical studies. This progression seeks to bridge the gap between the accessibility of cloud-based systems and the high precision demands of various professional fields.
- **Innovations in Terrestrial Photogrammetry for Archaeological Documentation:** The trajectory for terrestrial photogrammetry involves creating more robust and user-friendly methodologies for documenting historical and cultural sites. Enhancing algorithms for image processing and 3D model reconstruction will make low-cost cameras viable tools for professional-grade documentation, democratizing access to advanced archaeological and heritage conservation techniques. Establishing standardized procedures will further ensure the reliability and consistency of photogrammetric documentation across different sites and users.

- **Cross-disciplinary Applications and Research:** The integration of VR, 3D human pose estimation, and photogrammetry into broader research and application contexts represents a significant future direction. This encompasses not only the exploration of these technologies in various educational and entertainment settings but also a thorough examination of their ethical, social, and cultural impacts. Ensuring responsible and inclusive use is paramount as these technologies become more embedded in our daily lives.

- **Technology Integration and Hybrid Approaches:** Exploring the synergy between different technologies, such as combining VR with augmented reality (AR) for more immersive cultural experiences, represents an exciting frontier. Additionally, developing systems that merge the advantages of both cloud-based and sensor-based motion capture can offer flexible, high-quality solutions tailored to diverse user needs and application areas. This hybrid approach aims to leverage the strengths of each system to provide comprehensive and adaptable motion analysis tools.

References

[1] (1990). ASPRS ACCURACY STANDARDS FOR LARGE-SCALE MAPS. *The American Society for Photogrammetry and Remote Sensing*.

[2] (2009). Carnegie Mellon University - CMU Graphics Lab - motion capture library.

[3] Abdel-Bary Ebrahim, M. (2012). Using digital close-range photogrammetry as a QA/QC tool in construction: practical cases. *Survey Review*, 44(326):175–180.

[4] Achille, C., Adami, A., Chiarini, S., Cremonesi, S., Fassi, F., Fregonese, L., and Taffurelli, L. (2015). UAV-Based Photogrammetry and Integrated Technologies for Architectural Applications—Methodological Strategies for the After-Quake Survey of Vertical Structures in Mantua (Italy). *Sensors*, 15(7):15520–15539. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[5] Agisoft, L. L. C. and R., S. P. (2018). Agisoft PhotoScan User Manual - Professional Edition, Version 1.4. page 127.

[6] Ahmad, Y. (2006). The scope and definitions of heritage: From tangible to intangible. *International Journal of Heritage Studies*, 12(3):292–300.

[7] Ahmad Fuad, N., Yusoff, A. R., Ismail, Z., and Majid, Z. (2018). COMPARING THE PERFORMANCE OF POINT CLOUD REGISTRATION METHODS FOR LANDSLIDE MONITORING USING MOBILE LASER SCANNING DATA. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W9:11–21.

[8] Al-Ruzouq, R. and Abu Dabous, S. (2017). Archaeological Site Information Modelling and Management Based on Close-Range Photogrammetry and GIS. *Conservation and Management of Archaeological Sites*, 19(3):156–172.

[9] Al Zayer, M., MacNeilage, P., and Folmer, E. (2020). Virtual Locomotion: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 26(6):2315–2334. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[10] Albert, J. and Sung, K. (2018). User-centric classification of virtual reality locomotion. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18, pages 1–2, New York, NY, USA. Association for Computing Machinery.

[11] Altman, S., Xiao, W., and Grayson, B. (2017). EVALUATION OF LOW-COST TERRESTRIAL PHOTOGRAMMETRY FOR 3D RECONSTRUCTION OF COMPLEX BUILDINGS. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:199–206.

[12] Altuntas, C., Hezer, S., and Kırlı, S. (2017). Image Based Methods For Surveying Heritage Of Masonry Arch Bridge With The Example Of Dokuzunhan In Konya, Turkey. *SCIENTIFIC CULTURE*, 3:13–20.

[13] Anthes, C., García-Hernández, R., Wiedemann, M., and Kranzlmüller, D. (2016). State of the art of virtual reality technology. volume 2016-June.

[14] Arayici, Y. (2007). An approach for real world data modelling with the 3D terrestrial laser scanner for built environment. *Automation in Construction*, 16(6):816–829.

[15] Argelaguet, F. and Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136.

[16] Arnab, A., Doersch, C., and Zisserman, A. (2019). Exploiting temporal context for 3D human pose estimation in the wild. arXiv:1905.04266 [cs].

[17] Arns, L. L. (2002). A new taxonomy for locomotion in virtual environments. page 6069852, Ames. University: Iowa State University, Digital Repository.

[18] Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1968–1979, New York, NY, USA. Association for Computing Machinery.

[19] Bacca, J., Baldiris, S., Fabregat, R., Graf, S., and Kinshuk (2014). Augmented reality trends in education: A systematic review of research and applications. *Educational Technology and Society*, 17(4):133–149.

[20] Barazzetti, L., Previtali, M., and Roncoroni, F. (2018). CAN WE USE LOW-COST 360 DEGREE CAMERAS TO CREATE ACCURATE 3D MODELS? *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2:69–75.

[21] Barazzetti, L., Previtali, M., Roncoroni, F., and Valente, R. (2019). CONNECTING INSIDE AND OUTSIDE THROUGH 360° IMAGERY FOR CLOSE-RANGE PHOTOGRAMMETRY. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2-W9:87–92.

[22] Barbara, J. (2022). Re-Live History: An immersive virtual reality learning experience of prehistoric intangible cultural heritage. *Frontiers in Education*, 7:1032108.

[23] Barsanti, S. G., Remondino, F., Fenández-Palacios, B. J., and Visintini, D. (2014). Critical Factors and Guidelines for 3D Surveying and Modelling in Cultural Heritage. *International Journal of Heritage in the Digital Era*, 3(1):141–158.

[24] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded Up Robust Features. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision – ECCV 2006*, pages 404–417, Berlin, Heidelberg. Springer.

[25] Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., and Gain, J. (2018). A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage*, 11(2):1–36.

[26] Benko, H., Ishak, E., and Feiner, S. (2004). Collaborative mixed reality visualization of an archaeological excavation. pages 132–140.

[27] Benzine, A., Chabot, F., Luvison, B., Pham, Q. C., and Achard, C. (2020). PandaNet: Anchor-Based Single-Shot Multi-Person 3D Pose Estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6855–6864, Seattle, WA, USA. IEEE.

[28] Berkman, M. I. (2024). History of Virtual Reality. In Lee, N., editor, *Encyclopedia of Computer Graphics and Games*, pages 873–881. Springer International Publishing, Cham.

[29] Bertrand, M. and Bouchard, S. (2008). Applying the technology acceptance model to VR with people who are favorable to its use. *Journal of Cyber Therapy & Rehabilitation*, 1.

[30] Boletsis, C. (2017). The New Era of Virtual Reality Locomotion: A Systematic Literature Review of Techniques and a Proposed Typology. *Multimodal Technologies and Interaction*, 1(4):24. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[31] Boochs, F., Heinz, G., Huxhagen, U., and Müller, H. (2007). LOW-COST IMAGE BASED SYSTEM FOR NON-TECHNICAL EXPERTS IN CULTURAL HERITAGE DOCUMENTATION AND ANALYSIS. page 6.

[32] Bowman, D. (1999). Interaction Techniques For Common Tasks In Immersive Virtual Environments - Design, Evaluation, And.

[33] Bowman, D., Kruijff, E., and LaViola, I. (2004). 3D User Interfaces: Theory and Practice. page 512.

[34] Bozgeyikli, E., Raij, A., Katkoori, S., and Dubey, R. (2016). Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '16, pages 205–216, New York, NY, USA. Association for Computing Machinery.

[35] Breuss-Schneeweis, P. (2016). "The speaking celt" - Augmented reality avatars guide through a museum - Case study. pages 1484–1491.

[36] Bruno, F., Bruno, S., De Sensi, G., Luchi, M.-L., Mancuso, S., and Muzzupappa, M. (2010). From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage*, 11(1):42–49.

[37] Bustillo, A., Alaguero, M., Miguel, I., Saiz, J., and Iglesias, L. (2015). A flexible platform for the creation of 3D semi-immersive environments to teach Cultural Heritage. *Digital Applications in Archaeology and Cultural Heritage*, 2(4):248–259.

[38] Börlin, N., Murtiyoso, A., Grussenmeyer, P., Menna, F., and Nocerino, E. (2019). Flexible Photogrammetric Computations Using Modular Bundle Adjustment: The Chain Rule and the Collinearity Equations. *Photogrammetric engineering and remote sensing*, 85(5):361–368. Publisher: Asprs American Society for Photogrammetry and.

[39] Cameron, F. and Kenderdine, S. (2007). *Theorizing Digital Cultural Heritage: A Critical Discourse*.

[40] Carrozzino, M., Scucces, A., Leonardi, R., Evangelista, C., and Bergamasco, M. (2011). Virtually preserving the intangible heritage of artistic handicraft. *Journal of Cultural Heritage*, 12(1):82–87.

[41] Casiez, G., Vogel, D., Balakrishnan, R., and Cockburn, A. (2008). The Impact of Control-Display Gain on User Performance in Pointing Tasks. *Human-Computer Interaction*, 23(3):215–250.

[42] Centre, U. W. H. (2020). Medina of Fez.

[43] Chang, K.-E., Chang, C.-T., Hou, H.-T., Sung, Y.-T., Chao, H.-L., and Lee, C.-M. (2014). Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers and Education*, 71:185–197.

[44] Chang, Y.-L., Hon, H.-T., Pan, C.-Y., Sung, Y.-T., and Chang, K.-E. (2015). Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. *Educational Technology and Society*, 18(2):166–178.

[45] Chen, C.-H. and Ramanan, D. (2017). 3D Human Pose Estimation = 2D Pose Estimation + Matching.

[46] Chen, N. (2023). VR system application in the inheritance of intangible cultural heritage gongs and drums based on video simulation. *Soft Computing*, 27(14):10389–10399.

[47] Chiabrando, F., Donadio, E., and Rinaudo, F. (2015). SfM for Orthophoto to Generation: A Winning Approach for Cultural Heritage Knowledge. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5-W7:91–98.

[48] Chiabrando, F., Lo Turco, M., and Santagati, C. (2017). DIGITAL INVASIONS: FROM POINT CLOUDS TO HISTORICAL BUILDING OBJECT MODELING (H-BOM) OF A UNESCO WHL SITE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2-W3:171–178. Conference Name: TC II & CIPA
3D Virtual Reconstruction and Visualization of Complex Architectures (Volume XLII-2/W3) - 1–3 March 2017, Nafplio, Greece Publisher: Copernicus GmbH.

[49] Choi, H.-S. (2014). The conjugation method of augmented reality in museum exhibition. *International Journal of Smart Home*, 8(1):217–228.

[50] Chrysanthi, A., Papadopoulos, C., Frankland, T., and Earl, G. (2012). 'Tangible pasts': User-centred design of a mixed reality application for cultural heritage. *Archaeology in the Digital Era*, pages 31–39.

[51] Cignoni, P. and Scopigno, R. (2008). Sampled 3D models for CH applications: A viable and enabling new medium or just a technological exercise? *Journal on Computing and Cultural Heritage*, 1(1).

[52] Clini, P., Quattrini, R., Frontoni, E., Pierdicca, R., and Nespeca, R. (2016). Real/not real: Pseudo-holography and augmented reality applications for cultural heritage. In *Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling*, pages 201–227.

[53] Colizzi, L., Martini, A., and Chionna, F. (2010). Augmented reality applied to the diagnostics and fruition of cultural heritage. *Conserv. Sci. Cult. Herit.*, 10(1):195–238.

[54] Comes, R., Neamţu, C., Buna, Z., Badiu, I., and Pupeză, P. (2014). Methodology to create 3D models for augmented reality applications using scanned point clouds. *Mediterranean Archaeology and Archaeometry*, 14(4):35–44.

[55] Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., and Hart, J. C. (1992). The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM*, 35(6):64–72.

[56] D'Agnano, F., Balletti, C., Guerra, F., and Vernier, P. (2015). Tooteko: A case study of augmented reality for an accessible cultural heritage. Digitization, 3D printing and sensors for an audio-tactile experience. volume 40, pages 207–213. Issue: 5W4.

[57] Damala, A., Hornecker, E., van der Vaart, M., van Dijk, D., and Ruthven, I. (2016). The loupe: Tangible augmented reality for learning to look at ancient Greek art. *Mediterranean Archaeology and Archaeometry*, 16(5 Special Issue):73–85.

[58] Damala, A., Stojanovic, N., Schuchert, T., Moragues, J., Cabrera, A., and Gilleade, K. (2012). *Adaptive augmented reality for cultural heritage: ARtSENSE project*, volume 7616 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Pages: 755.

[59] Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems : theory and results*. Thesis, Massachusetts Institute of Technology. Accepted: 2005-08-08T16:11:32Z.

[60] Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340. Publisher: Management Information Systems Research Center, University of Minnesota.

[61] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8):982–1003. Publisher: INFORMS.

[62] De Paolis, L. T., Faggiano, F., Gatto, C., Barba, M. C., and De Luca, V. (2022). Immersive virtual reality for the fruition of ancient contexts: The case of the archaeological and Naturalistic Park of Santa Maria d'Agnano in Ostuni. *Digital Applications in Archaeology and Cultural Heritage*, 27:e00243.

[63] Deliyiannis, I. and Papaioannou, G. (2014). Augmented reality for archaeological environments on mobile devices: A novel open framework. *Mediterranean Archaeology and Archaeometry*, 14(4):1–10.

[64] Deng, X., Kim, I. T., and Shen, C. (2021). Research on Convolutional Neural Network-Based Virtual Reality Platform Framework for the Intangible Cultural Heritage Conservation of China Hainan Li Nationality: Boat-Shaped House as an Example. *Mathematical Problems in Engineering*, 2021:1–16.

[65] Denis, M. (2019). 39+ Smartphone Statistics You Should Know in 2019.

[66] Dewez, T. J. B., Girardeau-Montaut, D., Allanic, C., and Rohmer, J. (2016). FACETS : A CLOUDCOMPARE PLUGIN TO EXTRACT GEOLOGICAL PLANES FROM UNSTRUCTURED 3D POINT CLOUDS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5:799–804. Conference Name: XXIII ISPRS Congress, Commission V (Volume XLI-B5) - 12–19 July 2016, Prague, Czech Republic Publisher: Copernicus GmbH.

[67] Dhonju, H. K., Xiao, W., Sarhosis, V., Mills, J. P., Wilkinson, S., Wang, Z., Thapa, L., and Panday, U. S. (2017). FEASIBILITY STUDY OF LOW-COST IMAGE-BASED HERITAGE DOCUMENTATIONIN NEPAL. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3:237–242.

[68] Dominici, D., Alicandro, M., and Massimi, V. (2017). UAV photogrammetry in the post-earthquake scenario: case studies in L'Aquila. *Geomatics, Natural Hazards and Risk*, 8(1):87–103.

[69] Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J., and Gandy, M. (2005). Exploring spatial narratives and mixed reality experiences in Oakland Cemetery. volume 265, pages 51–60.

[70] Durand, E., Merienne, F., Pere, C., and Callet, P. (2014). Ray-on, an on-site photometric augmented reality device. *Journal on Computing and Cultural Heritage*, 7(2).

[71] Elkhrachy, I. (2019). Modeling and Visualization of Three Dimensional Objects Using Low-Cost Terrestrial Photogrammetry. *International Journal of Architectural Heritage*, pages 1–12.

[72] Empler, T. (2015). Cultural heritage: Displaying the Forum of Nerva with new technologies. pages 581–586.

[73] Etxeberria, A., Asensio, M., Vicent, N., and Cuenca, J. (2012). Mobile devices: A tool for tourism and learning at archaeological sites. *International Journal of Web Based Communities*, 8(1):57–62.

[74] Fabado, S., Seguí, A. E., Cabrelles, M., Navarro, S., García-De-San-Miguel, D., and Lerma, J. L. (2013). 3DVEM SOFTWARE MODULES FOR EFFICIENT MANAGEMENT OF POINT CLOUDS AND PHOTOREALISTIC 3D MODELS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5-W2:255–260. Conference Name: TC V
 XXIV International CIPA Symposium (Volume XL-5/W2) - 2–6 September 2013, Strasbourg, France Publisher: Copernicus GmbH.

[75] Fabbri, M., Lanzi, F., Calderara, S., Alletto, S., and Cucchiara, R. (2020). Compressed Volumetric Heatmaps for Multi-Person 3D Pose Estimation.

[76] Fagan, M., Kilmon, C., and Pandey, V. (2012). Exploring the adoption of a virtual reality simulation: The role of perceived ease of use, perceived usefulness and personal innovativeness. *Campus-Wide Information Systems*, 29(2):117–127. Publisher: Emerald Group Publishing Limited.

[77] Fangi, G. (2019). ALEPPO BEFORE AND AFTER THE WAR 2010–2018. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W15:449–456.

[78] Fassi, F., Troisi, S., Baiocchi, V., Del Pizzo, S., Giannone, F., Barazzetti, L., Previtali, M., and Roncoroni, F. (2018). Fisheye Photogrammetry to Survey Narrow Spaces in Architecture and a Hypogea Environment. MDPI. Book Title: Latest Developments in Reality-Based 3D Surveying and Modelling.

[79] Fineschi, A. and Pozzebon, A. (2015). A 3D virtual tour of the Santa Maria della Scala Museum Complex in Siena, Italy, based on the use of Oculus Rift HMD. In *2015 International Conference on 3D Imaging (IC3D)*, pages 1–5, Liège, Belgium. IEEE.

[80] Fioretti, G., Acquafredda, P., Calò, S., Cinelli, M., Germanò, G., Laera, A., and Moccia, A. (2020). Study and Conservation of the St. Nicola's Basilica Mosaics (Bari, Italy) by Photogrammetric Survey: Mapping of Polychrome Marbles, Decorative Patterns and Past Restorations. *Studies in Conservation*, 65(3):160–171.

[81] Frees, S. and Kessler, G. (2005). Precise and Rapid Interaction through Scaled Manipulation in Immersive Virtual Environments. *IEEE Virtual Reality Conference 2005 (VR'05)*, pages 99–106. Conference Name: IEEE Virtual Reality Conference 2005 (VR'05) ISBN: 9780780389298 Place: Bonn, Germany Publisher: IEEE.

[82] Fuchs, P., Moreau, G., and Berthoz, A. (2006). *Le traité de la réalité virtuelle volume 1 : L'Homme et l'environnement virtuel*.

[83] Furukawa, Y. and Ponce, J. (2010). Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[84] Gargalakos, M., Giallouri, E., Lazoudis, A., Sotiriou, S., and Bogner, F. (2011). Assessing the impact of technology-enhanced field trips in science centers and museums. *Advanced Science Letters*, 4(11-12):3332–3341.

[85] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., and Crawford, K. (2021). Datasheets for Datasets. arXiv:1803.09010 [cs].

[86] Gerval, J.-P. and Le Ru, Y. (2015). *Fusion of multimedia and mobile technology in audioguides for museums and exhibitions*, volume 84 of *Intelligent Systems Reference Library*. Pages: 205.

[87] Gesch, D., Michael, O., Susan, G., Nelson, C., Steuck, M., and Tyler, D. (2002). The National Elevation Dataset. *ASPRS*.

[88] Girardeau-Montaut, D. (2011). Cloud Compare-Open Source Project; OpenSource Project: Grenoble, France.

[89] Gonizzi Barsanti, S., Caruso, G., Micoli, L., Covarrubias Rodriguez, M., and Guidi, G. (2015). 3D visualization of cultural heritage artefacts with virtual reality devices. volume 40, pages 165–172. Issue: 5W7.

[90] Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1):33–51.

[91] Grenzdörffer, G. J., Naumann, M., Niemeyer, F., and Frank, A. (2015). SYMBIOSIS OF UAS PHOTOGRAMMETRY AND TLS FOR SURVEYING AND 3D MODELING OF CULTURAL HERITAGE MONUMENTS - A CASE STUDY ABOUT THE CATHEDRAL OF ST. NICHOLAS IN THE CITY OF GREIFSWALD. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1-W4:91–96. Conference Name: ISPRS UAV-g2015 (Volume XL-1/W4) - 30 August–2 September 2015, Toronto, Canada Publisher: Copernicus GmbH.

[92] Gruen, A. W. (1985). ADAPTIVE LEAST SQUARES CORRELATION: A POWERFUL IMAGE MATCHING TECHNIQUE.

[93] Grussenmeyer, P. and Khalil, O. A. (2002). Solutions for Exterior Orientation in Photogrammetry: A Review. *The Photogrammetric Record*, 17(100):615–634. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1477-9730.2002.tb01907.x.

[94] Grussenmeyer, P., Landes, T., Doneus, M., and Lermat, J.-L. (2018). Basics of range-based modelling techniques in Cultural Heritage. In *3D Recording, Documentation and Management of Cultural Heritage*, pages 305–368. Whittles Publishing.

[95] Gîrbacia, F., Butnariu, S., Orman, A., and Postelnicu, C. (2013). Virtual restoration of deteriorated religious heritage objects using augmented reality technologies. *European Journal of Science and Theology*, 9(2):223–231.

[96] Han, P.-H., Chen, Y.-S., Liu, I.-S., Jang, Y.-P., Tsai, L., Chang, A., and Hung, Y.-P. (2020). A Compelling Virtual Tour of the Dunhuang Cave With an Immersive Head-Mounted Display. *IEEE Computer Graphics and Applications*, 40(1):40–55.

[97] Hanan, H., Suwardhi, D., Nurhasanah, T., and Bukit, E. S. (2015). Batak Toba Cultural Heritage and Close-range Photogrammetry. *Procedia - Social and Behavioral Sciences*, 184:187–195.

[98] Harris, C. and Stephens, M. (1988). A Combined Corner and Edge Detector. In *Procedings of the Alvey Vision Conference 1988*, pages 23.1–23.6, Manchester. Alvey Vision Club.

[99] Hassani, F. (2015). Documentation of cultural heritage; techniques, potentials, and constraints. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5-W7:207–214. Conference Name: 25th International CIPA Symposium 2015 (XL-5/W7) - 31 August–4 September 2015, Taipei, Taiwan Publisher: Copernicus GmbH.

[100] Haukaas, C. and Hodgetts, L. M. (2016). The Untapped Potential of Low-Cost Photogrammetry in Community-Based Archaeology: A Case Study From Banks Island, Arctic Canada. *Journal of Community Archaeology & Heritage*, 3(1):40–56.

[101] Haydar, M., Roussel, D., Maïdi, M., Otmane, S., and Mallem, M. (2011). Virtual and augmented reality for cultural computing and heritage: A case study of virtual exploration of underwater archaeological sites (preprint). *Virtual Reality*, 15(4):311–327.

[102] Heberle, H., Meirelles, G., da Silva, F., Telles, G., and Minghim, R. (2015). InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16(1).

[103] Heilig, M. L. (1962). Sensorama simulator.

[104] Herbig, U., Stampfer, L., Grandits, D., Mayer, I., Pöchtrager, M., Ikaputra, and Setyastuti, A. (2019). DEVELOPING A MONITORING WORKFLOW FOR THE TEMPLES OF JAVA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2-W15:555–562.

[Hernández et al.] Hernández, L. A., Taibo, J., and Seoane, A. Empty Museum. An Immersive, Walkable VR Framework for Multiuser Interaction and Telepresence.

[106] Hernández, L. A., Taibo, J., and Seoane, A. (2002). Empty Museum. An Immersive, Walkable VR Framework for Multiuser Interaction and Telepresence.

[107] Hirschmüller, H. (2011). Semi-global matching motivation, developments and applications. Pages: 5 –9 Publication Title: In 53. Photogrammetric Week.

[108] Hossain, M. R. I. and Little, J. J. (2018). Exploiting Temporal Information for 3D Human Pose Estimation. In Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., editors, *Computer Vision – ECCV 2018*, volume 11214, pages 69–86. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

[109] Hou, L., Wang, Y., Wang, X., Maynard, N., Cameron, I. T., Zhang, S., and Jiao, Y. (2014). Combining Photogrammetry and Augmented Reality Towards an Integrated Facility Management System for the Oil Industry. *Proceedings of the IEEE*, 102(2):204–220.

[110] Hsieh, C.-K., Liao, W.-C., Yu, M.-C., and Hung, Y.-P. (2014). Interacting with the past: Creating a time perception journey experience using kinect-based breath detection and deterioration and recovery simulation technologies. *Journal on Computing and Cultural Heritage*, 7(1).

[111] Ibrahim, N. and Ali, N. M. (2018). A Conceptual Framework for Designing Virtual Heritage Environment for Cultural Learning. *Journal on Computing and Cultural Heritage*, 11(2):1–27.

[112] Interrante, V., Ries, B., and Anderson, L. (2007). Seven League Boots: A New Metaphor for Augmented Locomotion through Moderately Large Scale Immersive Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*.

[113] Invitto, S., Spada, I., Turco, D., and Belmonte, G. (2014). *Easy perception lab: Evolution, brain and virtual and augmented reality in museum environment*, volume 8853 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Pages: 310.

[114] Ionescu, C., Papava, D., Olaru, V., and Sminchisescu, C. (2014). Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339.

[115] Jahangiri, E. and Yuille, A. L. (2017). Generating Multiple Diverse Hypotheses for Human 3D Pose Consistent with 2D Joint Detections. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 805–814, Venice. IEEE.

[116] Jones, C. A. and Church, E. (2020). Photogrammetry is for everyone: Structure-from-motion software user experiences in archaeology. *Journal of Archaeological Science: Reports*, 30:102261.

[117] Kadri, M., Khalloufi, H., and Azough, A. (2020). V-Museum: A Virtual Museum Based on Augmented and Virtual Realities for Cultural Heritage Mediation. In *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pages 1–5.

[118] Kang, J. (2013). AR teleport: Digital reconstruction of historical and cultural-heritage sites for mobile phones via movement-based interactions. *Wireless Personal Communications*, 70(4):1443–1462.

[119] Kateros, S., Georgiou, S., Papaefthymiou, M., Papagiannakis, G., and Tsioumas, M. (2015). A comparison of gamified, immersive VR curation methods for enhanced presence and human-computer interaction in digital humanities. *International Journal of Heritage in the Digital Era*, 4(2):221–233.

[120] Katsouri, I., Tzanavari, A., Herakleous, K., and Poullis, C. (2015). Visualizing and assessing hypotheses for marine archaeology in a VR CAVE environment. *Journal on Computing and Cultural Heritage*, 8(2).

[121] Kersten, T. P. and Lindstaedt, M. (2012). Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects. In Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., and Caffo, R., editors, *Progress in Cultural Heritage Preservation*, volume 7616, pages 1–10. Springer Berlin Heidelberg, Berlin, Heidelberg.

[122] Kim, K., Seo, B.-K., Han, J.-H., and Park, J.-I. (2009). Augmented reality tour system for immersive experience of cultural heritage. pages 323–324.

[123] Kishino, F. and Milgram, P. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems*, pages 1321–1329.

[124] Kohli, L., Whitton, M. C., and Brooks, F. P. (2012). Redirected touching: The effect of warping space on task performance. *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 105–112. Conference Name: 2012 IEEE Symposium on 3D User Interfaces (3DUI) ISBN: 9781467312059 9781467312042 Place: Costa Mesa, CA Publisher: IEEE.

[125] Kwoczynska, B., Sagan, W., and Dziura, K. (2016). Elaboration and Modeling of the Railway Infrastructure Using Data from Airborne and Mobile Laser Scanning. In *2016 Baltic Geodetic Congress (BGC Geomatics)*, pages 106–115.

[126] Lachat, E., Landes, T., and Grussenmeyer, P. (2016). COMBINATION OF TLS POINT CLOUDS AND 3D DATA FROM KINECT V2 SENSOR TO COMPLETE INDOOR MODELS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5:659–666. Conference Name: XXIII ISPRS Congress, Commission V (Volume XLI-B5) - 12–19 July 2016, Prague, Czech Republic Publisher: Copernicus GmbH.

[127] Lachat, E., Landes, T., and Grussenmeyer, P. (2017). FIRST EXPERIENCES WITH THE TRIMBLE SX10 SCANNING TOTAL STATIONFOR BUILDING FACADE SURVEY. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3:405–412.

[128] Lachat, E., Landes, T., and Grussenmeyer, P. (2018). COMPARISON OF POINT CLOUD REGISTRATION ALGORITHMS FOR BETTER RESULT ASSESSMENT – TOWARDS AN OPEN-SOURCE SOLUTION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2:551–558. Conference Name: ISPRS TC II Mid-term Symposium <q>Towards Photogrammetry 2020</q> (Volume XLII-2) - 4–7 June 2018, Riva del Garda, Italy Publisher: Copernicus GmbH.

[129] Lachat, E., Macher, H., Landes, T., and Grussenmeyer, P. (2015). Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling. *Remote Sensing*, 7(10):13070–13097.

[130] Lerma, J. L., Navarro, S., Cabrelles, M., and Villaverde, V. (2010). Terrestrial laser scanning and close range photogrammetry for 3D archaeological documentation: the Upper Palaeolithic Cave of Parpalló as a case study. *Journal of Archaeological Science*, 37(3):499–507.

[131] Lerones, P. M., Fernández, J. L., Gil, M., Gómez-García-Bermejo, J., and Casanova, E. Z. (2010). A practical approach to making accurate 3D layouts of interesting cultural heritage sites through digital models. *Journal of Cultural Heritage*, 11(1):1–9.

[132] Li, C. and Lee, G. H. (2019). Generating Multiple Hypotheses for 3D Human Pose Estimation with Mixture Density Network.

[133] Liestøl, G. (2014). *Along the appian way. Storytelling and memory across time and space in mobile augmented reality*, volume 8740 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Pages: 257.

[134] Lin, Z., Zhang, L., Tang, S., Song, Y., and Ye, X. (2021). Evaluating Cultural Landscape Remediation Design Based on VR Technology. *ISPRS International Journal of Geo-Information*, 10(6):423.

[135] Liritzis, I., Laskaris, N., Vafiadou, A., Karapanagiotis, I., Volonakis, P., Papageorgopoulou, C., and Bratitsi, M. (2020). ARCHAEOMETRY: AN OVERVIEW. *SCIENTIFIC CULTURE*, 6:49–98.

[136] Liu, R., Zhu, Y., Luo, Y., and Liu, X. (2009). Construction Urban Infrastructure Based on Core Techniques of Digital Photogrammetry and Remote Sensing. In *2009 International Forum on Information Technology and Applications*, pages 536–539. IEEE. Place: Chengdu, China.

[137] Liu, W., Bao, Q., Sun, Y., and Mei, T. (2021). Recent Advances in Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective.

[138] Liu, Z., Yan, S., Lu, Y., and Zhao, Y. (2022). Generating Embodied Storytelling and Interactive Experience of China Intangible Cultural Heritage "Hua'er" in Virtual Reality. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, pages 1–7, New York, NY, USA. Association for Computing Machinery.

[139] Loizides, F., El Kater, A., Terlikas, C., Lanitis, A., and Michael, D. (2014). Presenting Cypriot Cultural Heritage in Virtual Reality: A User Evaluation. In Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., and Quak, E., editors, *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, volume 8740, pages 572–579. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

[140] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.

[141] Luhmann, T., Robson, S., Kyle, S., and Boehm, J. (2019). Close-Range Photogrammetry and 3D Imaging. In *Close-Range Photogrammetry and 3D Imaging*. De Gruyter.

[142] Magnenat-Thalmann, N. and Papagiannakis, G. (2005). Virtual worlds and augmented reality in cultural heritage applications. *Recording, Modeling and Visualization of Cultural Heritage*, pages 419–430.

[143] Magnenat-Thalmann, N., Papagiannakis, G., Foni, A., Arevalo, M., and Cadi-Yazli, N. (2004). Simulating life in ancient sites using mixed reality technology. *CEIG04*.

[144] Mallik, A., Chaudhury, S., and Ghosh, H. (2011). Nrityakosha: Preserving the intangible heritage of Indian classical dance. *Journal on Computing and Cultural Heritage*, 4(3):1–25.

[145] Manis, K. T. and Choi, D. (2019). The virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware. *Journal of Business Research*, 100:503–513.

[146] Manovich, L. (2006). The poetics of augmented space. *Visual Communication*, 5(2):219–240.

[147] Martinez, J., Hossain, R., Romero, J., and Little, J. J. (2017). A Simple Yet Effective Baseline for 3d Human Pose Estimation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2659–2668, Venice. IEEE.

[148] Mase, K., Kadobayashi, R., and Nakatsu, R. (1996). VSMM'96 in Gifu Meta-Museum: A Supportive Augmented-Reality Environment for Knowledge Sharing.

[149] Matuk, C. (2016). The learning affordances of augmented reality for museum exhibits on human health. *Museums and Social Issues*, 11(1):73–87.

[150] McMahan, R. P. (2011). Exploring the Effects of Higher-Fidelity Display and Interaction for Virtual Reality Games.

[151] McMahan, R. P., Bowman, D. A., Zielinski, D. J., and Brady, R. B. (2012). Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):626–633. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[152] Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., and Theobalt, C. (2017). Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision.

[153] Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., and Jorge, J. (2019). A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum*, 38(1):21–45.

[154] Menna, F., Nocerino, E., Drap, P., Remondino, F., Murtiyoso, A., Grussenmeyer, P., and Börlin, N. (2018). IMPROVING UNDERWATER ACCURACY BY EMPIRICAL WEIGHTING OF IMAGE OBSERVATIONS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2:699–705. Conference Name: ISPRS TC II Mid-term Symposium <q>Towards Photogrammetry 2020</q> (Volume XLII-2) - 4–7 June 2018, Riva del Garda, Italy Publisher: Copernicus GmbH.

[155] Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., Gerl, O., Arnaudov, S., and Lieberknecht, S. (2008). An augmented reality museum guide. pages 103–106.

[156] Moe, D., Sampath, A., Christopherson, J., and Benson, M. (2010). SELF CALIBRATION OF SMALL AND MEDIUM FORMAT DIGITAL CAMERAS.

[157] Moesgaard, T. G., Witt, M., Fiss, J., Warming, C., Klubien, J., and Schønau-Fog, H. (2015). Implicit and Explicit Information Mediation in a Virtual Reality Museum Installation and its Effects on Retention and Learning Outcomes.

[158] Moon, G. and Lee, K. M. (2020). I2L-MeshNet: Image-to-Lixel Prediction Network for Accurate 3D Human Pose and Mesh Estimation from a Single RGB Image. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, volume 12352, pages 752–768. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

[159] Moreno-Noguer, F. (2016). 3D Human Pose Estimation from a Single Image via Distance Matrix Regression.

[160] Mulahusić, A., Tuno, N., Gajski, D., and Topoljak, J. (2020). Comparison and analysis of results of 3D modelling of complex cultural and historical objects using different types of terrestrial laser scanner. *Survey Review*, 52(371):107–114.

[161] Murtiyoso, A., Grussenmeyer, P., Suwardhi, D., and Awalludin, R. (2018). Multi-Scale and Multi-Sensor 3D Documentation of Heritage Complexes in Urban Areas. *ISPRS International Journal of Geo-Information*, 7(12):483.

[162] Murtiyoso, A., Koehl, M., Grussenmeyer, P., and Freville, T. (2017). ACQUISITION AND PROCESSING PROTOCOLS FOR UAV IMAGES:3D MODELING OF HISTORICAL BUILDINGS USING PHOTOGRAMMETRY. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W2:163–170.

[163] Muñumer, E. and Lerma, J. L. (2015). Fusion of 3D data from different image-based and range-based sources for efficient heritage recording. In *2015 Digital Heritage*, volume 1, pages 83–86.

[164] Münster, S., Kröber, C., Weller, H., and Prechtel, N. (2017). Virtual reconstruction of historical architecture as media for knowledge representation. *Mixed Reality and Gamification for Cultural Heritage*, pages 313–330.

[165] Nie, Q., Liu, Z., and Liu, Y. (2020). Unsupervised 3D Human Pose Representation with Viewpoint and Pose Disentanglement. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, volume 12364, pages 102–118. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

[166] Okura, F., Kanbara, M., and Yokoya, N. (2015). Mixed-reality world exploration using image-based rendering. *Journal on Computing and Cultural Heritage*, 8(2).

[167] Papaefthymiou, M., Plelis, K., Mavromatis, D., and Papagiannakis, G. (2015). Mobile Virtual Reality featuring a six degrees of freedom interaction paradigm in a virtual museum application.

[168] Pavlakos, G., Zhou, X., and Daniilidis, K. (2018). Ordinal Depth Supervision for 3D Human Pose Estimation. arXiv:1805.04095 [cs].

[169] Pavllo, D., Feichtenhofer, C., Grangier, D., and Auli, M. (2019). 3D human pose estimation in video with temporal convolutions and semi-supervised training.

[170] Pescarin, S., Pagano, A., Wallergård, M., Hupperetz, W., and Ray, C. (2012). Archeovirtual 2011: An evaluation approach to virtual museums. pages 25–32.

[171] Petridis, P., Dunwell, I., Liarokapis, F., Constantinou, G., Arnab, S., De Freitas, S., and Hendrix, M. (2013). The herbert virtual museum. *Journal of Electrical and Computer Engineering*.

[172] Peña-Villasenín, S., Gil-Docampo, M., and Ortiz-Sanz, J. (2017). 3-D Modeling of Historic Façades Using SFM Photogrammetry Metric Documentation of Different Building Types of a Historic Center. *International Journal of Architectural Heritage*, 11(6):871–890.

[173] Pierdicca, R., Frontoni, E., Malinverni, E., Colosi, F., and Orazi, R. (2016a). Virtual reconstruction of archaeological heritage using a combination of photogrammetric techniques: Huaca Arco Iris, Chan Chan, Peru. *Digital Applications in Archaeology and Cultural Heritage*, 3(3):80–90.

[174] Pierdicca, R., Frontoni, E., Zingaretti, P., Malinverni, E., Colosi, F., and Orazi, R. (2015). *Making visible the invisible. Augmented reality visualization for 3D reconstructions of archaeological sites*, volume 9254 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Pages: 37.

[175] Pierdicca, R., Malinverni, E., Frontoni, E., Colosi, F., and Orazi, R. (2016b). 3D visualization tools to explore ancient architectures in South America. *Virtual Archaeology Review*, 7(15):44–53.

[176] Pietroni, E., Pagano, A., and Rufa, C. (2013). The Etruscanning project: Gesture-based interaction and user experience in the virtual reconstruction of the Regolini-Galassi tomb. volume 2, pages 653–660.

[177] Plecher, D. A., Herber, F., Eichhorn, C., Pongratz, A., Tanson, G., and Klinker, G. (2020). HieroQuest - A Serious Game for Learning Egyptian Hieroglyphs. *Journal on Computing and Cultural Heritage*, 13(4):1–20.

[178] Portalés, C., Lerma, J., and Pérez, C. (2009). Photogrammetry and augmented reality for cultural heritage applications. *Photogrammetric Record*, 24(128):316–331.

[179] Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. (1996). The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pages 79–80, New York, NY, USA. Association for Computing Machinery.

[180] Prokop, A. (2008). Assessing the applicability of terrestrial laser scanning for spatial snow depth measurements. *Cold Regions Science and Technology*, 54(3):155–163.

[181] Putra, E. Y., Wahyudi, A. K., and Dumingan, C. (2016). A proposed combination of photogrammetry, Augmented Reality and Virtual Reality Headset for heritage visualisation. In *2016 International Conference on Informatics and Computing (ICIC)*, pages 43–48.

[182] Quattrini, R., Pierdicca, R., Frontoni, E., and Barcaglioni, R. (2016). Virtual reconstruction of lost architectures: From the TLS survey to AR visualization. volume 41, pages 383–390.

[183] Rahaman, H. and Champion, E. (2019). To 3D or Not 3D: Choosing a Photogrammetry Workflow for Cultural Heritage Groups. *Heritage*, 2(3):1835–1851. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[184] Ramos, M. M. and Remondino, F. (2015). Data fusion in Cultural Heritage – A Review. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5-W7:359–363.

[185] Remondino, F. and Campana, S. (2014). *3D Recording and Modelling in Archaeology and Cultural Heritage - Theory and Best Practices*.

[186] Remondino, F., Del Pizzo, S., Kersten, T. P., and Troisi, S. (2012). Low-Cost and Open-Source Solutions for Automated Image Orientation – A Critical Overview. In Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., and Caffo, R., editors,

Progress in Cultural Heritage Preservation, volume 7616, pages 40–54. Springer Berlin Heidelberg, Berlin, Heidelberg.

[187] Remondino, F., Spera, M. G., Nocerino, E., Menna, F., and Nex, F. (2014). State of the art in high density image matching. *Photogrammetric record*, 29(146):144–166. Publisher: Wiley-Blackwell.

[188] Reshetyuk, Y. (2009). *Terrestrial laser scanning : Error sources, self-calibration and direct georeferencing*.

[189] Reshetyuk, Y. (2019). Self-calibration and direct georeferencing in terrestrial laser scanning.

[190] Reunanen, M., Díaz, L., and Horttana, T. (2015). A holistic user-centered approach to immersive digital cultural heritage installations: Case Vrouw Maria. *Journal on Computing and Cultural Heritage*, 7(4).

[191] Richards-Rissetto, H., Robertsson, J., von Schwerin, J., Agugiaro, G., Remondino, F., and Girardi, G. (2014). Geospatial Virtual Heritage: A Gesture-Based 3D GIS to Engage the Public with Ancient Maya Archaeology. *Archaeology in the Digital Era*, pages 118–130.

[192] Ridel, B., Reuter, P., Laviole, J., Mellado, N., Couture, N., and Granier, X. (2014). The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing and Cultural Heritage*, 7(2).

[193] Rogez, G., Weinzaepfel, P., and Schmid, C. (2017). LCR-Net: Localization-Classification-Regression for Human Pose. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1216–1224, Honolulu, HI. IEEE.

[194] Rogez, G., Weinzaepfel, P., and Schmid, C. (2019). LCR-Net++: Multi-person 2D and 3D Pose Detection in Natural Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.

[195] Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *2011 IEEE International Conference on Robotics and Automation*, pages 1–4. ISSN: 1050-4729.

[196] Saggio, G. and Borra, D. (2011). Augmented reality for restoration/reconstruction of artefacts with artistic or historical value. *Augmented reality - Some emerging application areas. InTech.*, pages 59–86.

[197] Santos, P., Acri, D., Gierlinger, T., Schmedt, H., and Stork, A. (2010). Supporting outdoor mixed reality applications for architecture and cultural heritage.

[198] Schenk, T. (2005). Introduction to photogrammetry. Department of Civil and Environmental Engineering and Geodetic Engineering.

[199] Schnädelbach, H., Koleva, B., Flintham, M., Fraser, M., Izadi, S., Chandler, P., Foster, M., Benford, S., Greenhalgh, C., and Rodden, T. (2002). The augurscope: A mixed reality interface for outdoors. volume 4, pages 9–16. Issue: 1.

[200] Selmanović, E., Rizvic, S., Harvey, C., Boskovic, D., Hulusic, V., Chahin, M., and Sljivo, S. (2020). Improving Accessibility to Intangible Cultural Heritage Preservation Using Virtual Reality. *Journal on Computing and Cultural Heritage*, 13(2):1–19.

[201] Shanoer, M. M. and Abed, F. M. (2018). Evaluate 3D laser point clouds registration for cultural heritage documentation. *The Egyptian Journal of Remote Sensing and Space Science*, 21(3):295–304.

[202] Sharma, S., Varigonda, P. T., Bindal, P., Sharma, A., and Jain, A. (2019). Monocular 3D Human Pose Estimation by Generation and Ordinal Ranking. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2325–2334, Seoul, Korea (South). IEEE.

[203] Sigal, L., Balan, A. O., and Black, M. J. (2010). HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion. *International Journal of Computer Vision*, 87(1-2):4–27.

[204] Skarbez, R., Smith, M., and Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2. Publisher: Frontiers.

[205] Skovfoged, M. M., Viktor, M., Sokolov, M. K., Hansen, A., Nielsen, H. H., and Rodil, K. (2018). The tales of the Tokoloshe: safeguarding intangible cultural heritage using virtual reality. In *Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities*, AfriCHI '18, pages 1–4, New York, NY, USA. Association for Computing Machinery.

[206] Slater, M. (2003). A note on presence terminology.

[207] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557. Publisher: Royal Society.

[208] Slater, M., Steed, A., McCarthy, J., and Maringelli, F. (1998). The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors*, 40(3):469–477. Publisher: SAGE Publications Inc.

[209] Slater, M., Usoh, M., and Steed, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction*, 2(3):201–219.

[210] Soga, A. (2015). Virtual Show, Go In!: Walk-Through System and VR Goggles of a Temple for Museum Exhibits. In *2015 International Conference on Culture and Computing (Culture Computing)*, pages 199–200.

[211] Staiger, R. (2003). Terrestrial Laser Scanning Technology, Systems and Applications.

[212] Stanco, F., Tanasi, D., Gallo, G., Buffa, M., and Basile, B. (2012). Augmented perception of the past - The case of Hellenistic Syracuse. *Journal of Multimedia*, 7(2):211–216.

[213] Steinicke, F., Bruder, G., Jerald, J., Frenz, H., and Lappe, M. (2010). Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16(1):17–27. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[214] Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4):73–93. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.1992.tb00812.x.

[215] Stylianidis, E. and Remondino, F., editors (2016). *3D recording, documentation and management of cultural heritage*. Whittles Publishing, Caithness, Scotland, UK.

[216] Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA. Association for Computing Machinery.

[217] Sylaiou, S., Mania, K., Karoulis, A., and White, M. (2010). Exploring the relationship between presence and enjoyment in a virtual museum. *International Journal of Human Computer Studies*, 68(5):243–253.

[218] Szeliski, R. (2011). *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer, London.

[219] Tekin, B., Marquez-Neila, P., Salzmann, M., and Fua, P. (2017). Learning to Fuse 2D and 3D Image Cues for Monocular Body Pose Estimation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3961–3970, Venice. IEEE.

[220] Tompson, J., Jain, A., LeCun, Y., and Bregler, C. (2014). Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation.

[221] Tsiafaki, D. and Michailidou, N. (2015). Benefits And Problems Through The Application Of 3D Technologies In Archaeology: Recording, Visualisation, Representation And Reconstruction. *SCIENTIFIC CULTURE*, 1:37–45.

[222] Vanoni, D., Seracini, M., and Kuester, F. (2012). ARtifact: Tablet-based augmented reality for interactive analysis of cultural artifacts. pages 44–49.

[223] Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., and Schmid, C. (2017). Learning from Synthetic Humans. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4627–4635. arXiv:1701.01370 [cs].

[224] Venkatesh, V. and Davis, F. D. (1996). A Model of the Antecedents of Perceived Ease of Use: Development and Test*. *Decision Sciences*, 27(3):451–481. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.1996.tb00860.x.

[225] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3):425–478. Publisher: Management Information Systems Research Center, University of Minnesota.

[226] Verhoeven, G., Taelman, D., and Vermeulen, F. (2012). Computer Vision-Based Orthophoto Mapping of Complex Archaeological Sites: The Ancient Quarry of Pitaranha (portugal–Spain). *Archaeometry*, 54(6):1114–1129. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-4754.2012.00667.x.

[227] Verykokou, S., Ioannidis, C., and Kontogianni, G. (2014). *3D visualization via augmented reality: The case of the middle stoa in the ancient agora of Athens*, volume 8740 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Pages: 289.

[228] von Marcard, T., Henschel, R., Black, M. J., Rosenhahn, B., and Pons-Moll, G. (2018). Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. pages 601–617.

[229] Vosselman, G. and Maas, H.-G. (2010). *Airborne and terrestrial laser scanning*. Whittles Publishing Dunbeath, Dunbeath.

[230] Waagen, J. (2019). New technology and archaeological practice. Improving the primary archaeological recording process in excavation by means of UAS photogrammetry. *Journal of Archaeological Science*, 101:11–20.

[231] Wandt, B. and Rosenhahn, B. (2019). RepNet: Weakly Supervised Training of an Adversarial Reprojection Network for 3D Human Pose Estimation. arXiv:1902.09868 [cs].

[232] Wang, C., Li, J., Liu, W., Qian, C., and Lu, C. (2020). HMOR: Hierarchical Multi-person Ordinal Relations for Monocular Multi-person 3D Pose Estimation. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, volume 12348, pages 242–259. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

[233] Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., and Shao, L. (2021). Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*, 210:103225.

[234] Wang, M., Chen, X., Liu, W., Qian, C., Lin, L., and Ma, L. (2018). DRPose3D: Depth Ranking in 3D Human Pose Estimation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 978–984, Stockholm, Sweden. International Joint Conferences on Artificial Intelligence Organization.

[235] Way, D.-L. and Wei, Y.-H. (2023). Use of Cloud-Based Virtual Reality in Chinese Glove Puppetry to Preserve Intangible Cultural Heritage. *Applied Sciences*, 13(9):5699.

[236] Weiquan Lu, Linh-Chi Nguyen, Teong Leong Chuah, and Ellen Yi-Luen Do (2014). Effects of mobile AR-enabled interactions on retention and transfer for learning in art museum contexts. *2014 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design (IMSAR-MASH'D)*, pages 3–11. Conference Name: 2014 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design (ISMAR-MASH'D) ISBN: 9781479968879 Place: Germany Publisher: IEEE.

[237] Wiley, B. and Schulze, J. (2015). ArchAR - An archaeological augmented reality experience. volume 9392.

[238] Wingrave, C. A., Bowman, D. A., and Ramakrishnan, N. (2002). Towards preferences in virtual environment interfaces. In *Proceedings of the workshop on Virtual environments 2002*, EGVE '02, pages 63–72, Goslar, DEU. Eurographics Association.

[239] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg.

[240] Wojciechowski, R., Walczak, K., White, M., and Cellary, W. (2004). Building virtual and augmented reality museum exhibitions. pages 135–144.

[241] Xie, R. (2021). Intangible Cultural Heritage High-Definition Digital Mobile Display Technology Based on VR Virtual Visualization. *Mobile Information Systems*, 2021:1–11.

[242] Xu, X., Chen, H., Moreno-Noguer, F., Jeni, L. A., and De la Torre, F. (2020). 3D Human Shape and Pose from a Single Low-Resolution Image with Self-Supervised Learning.

[243] Yamazaki, M., Kasada, K., Hayashi, O., Narumi, T., Tanikawa, T., and Hirose, M. (2010). Wide FOV displays for Digital Museum. In *2010 16th International Conference on Virtual Systems and Multimedia*, pages 63–68, Seoul, Korea (South). IEEE.

[244] Yang, W., Ouyang, W., Wang, X., Ren, J., Li, H., and Wang, X. (2018). 3D Human Pose Estimation in the Wild by Adversarial Learning. arXiv:1803.09722 [cs].

[245] Zanfir, A., Marinoiu, E., and Sminchisescu, C. (2018). Monocular 3D Pose and Shape Estimation of Multiple People in Natural Scenes: The Importance of Multiple Scene Constraints. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2148–2157, Salt Lake City, UT, USA. IEEE.

[246] Zara, J. (2004). Virtual reality and cultural heritage on the web. *Proceedings of the 7th International Conference on Computer Graphics and Artificial Inteligence*, pages 101–112.

[247] Zhang, J. and Hu, Q.-W. (2018). A Visualization Progress Management Approach of Bridge Construction Based on Mixed Panoramic and Oblique Photogrammetry. In *2018 26th International Conference on Geoinformatics*, pages 1–6. IEEE. Place: Kunming.

[248] Zhang, L., Wang, Y., Tang, Z., Liu, X., and Zhang, M. (2023). A Virtual Experience System of Bamboo Weaving for Sustainable Research on Intangible Cultural Heritage Based on VR Technology. *Sustainability*, 15(4):3134.

[249] Zhang, T., Huang, B., and Wang, Y. (2020). Object-Occluded Human Shape and Pose Estimation From a Single Color Image. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7374–7383, Seattle, WA, USA. IEEE.

[250] Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., and Shah, M. (2022). Deep Learning-Based Human Pose Estimation: A Survey.

[251] Zhou, K., Han, X., Jiang, N., Jia, K., and Lu, J. (2019). HEMlets Pose: Learning Part-Centric Heatmap Triplets for Accurate 3D Human Pose Estimation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2344–2353, Seoul, Korea (South). IEEE.

[252] Zoellner, M., Keil, J., Drevensek, T., and Wuest, H. (2009). Cultural heritage layers: Integrating historic media in augmented reality. pages 193–196.